



# **Defect detection in infrared thermography by deep learning algorithms**

**Thèse**

**Qiang Fang**

**Doctorat en génie électrique**  
Philosophiæ doctor (Ph. D.)

Québec, Canada

# **DEFECT DETECTION IN INFRARED THERMOGRAPHY BY DEEP LEARNING ALGORITHMS**

**Thèse**

Qiang Fang

Sous la direction de :

Xavier P.V. Maldague, directeur de recherche

# Résumé

L'évaluation non destructive (END) est un domaine permettant d'identifier tous les types de dommages structurels dans un objet d'intérêt sans appliquer de dommages et de modifications permanents. Ce domaine fait l'objet de recherches intensives depuis de nombreuses années. La thermographie infrarouge (IR) est l'une des technologies d'évaluation non destructive qui permet d'inspecter, de caractériser et d'analyser les défauts sur la base d'images infrarouges (séquences) provenant de l'enregistrement de l'émission et de la réflexion de la lumière infrarouge afin d'évaluer les objets non autochauffants pour le contrôle de la qualité et l'assurance de la sécurité.

Ces dernières années, le domaine de l'apprentissage profond de l'intelligence artificielle a fait des progrès remarquables dans les applications de traitement d'images. Ce domaine a montré sa capacité à surmonter la plupart des inconvénients des autres approches existantes auparavant dans un grand nombre d'applications. Cependant, en raison de l'insuffisance des données d'entraînement, les algorithmes d'apprentissage profond restent encore inexplorés, et seules quelques publications font état de leur application à l'évaluation non destructive de la thermographie (TNDE).

Les algorithmes d'apprentissage profond intelligents et hautement automatisés pourraient être couplés à la thermographie infrarouge pour identifier les défauts (dommages) dans les composites, l'acier, etc. avec une confiance et une précision élevée. Parmi les sujets du domaine de recherche TNDE, les techniques d'apprentissage automatique supervisées et non supervisées sont les tâches les plus innovantes et les plus difficiles pour l'analyse de la détection des défauts.

Dans ce projet, nous construisons des cadres intégrés pour le traitement des données brutes de la thermographie infrarouge à l'aide d'algorithmes d'apprentissage profond et les points forts des méthodologies proposées sont les suivants:

1. Identification et segmentation automatique des défauts par des algorithmes d'apprentissage profond en thermographie infrarouge. Les réseaux neuronaux convolutifs (CNN) pré-entraînés sont introduits pour capturer les caractéristiques des défauts dans les images thermiques infrarouges afin de mettre en œuvre des modèles basés sur les CNN pour la détection des défauts structurels dans les échantillons composés de matériaux composites (diagnostic des défauts). Plusieurs alternatives de CNNs profonds pour la détection de défauts dans la thermographie infrarouge. Les comparaisons de performance de la détection et de la segmentation automatique des défauts dans la thermographie infrarouge en utilisant différentes méthodes de détection par apprentissage profond : (i) segmentation d'instance (Center-mask ; Mask-RCNN) ; (ii) détection d'objet (Yolo-v3 ; Faster-RCNN) ; (iii) segmentation sémantique (Unet ; Res-unet);

2. Technique d'augmentation des données par la génération de données synthétiques pour réduire le coût des dépenses élevées associées à la collecte de données infrarouges originales

dans les composites (composants d'aéronefs.) afin d'enrichir les données de formation pour l'apprentissage des caractéristiques dans TNDE;

3. Le réseau antagoniste génératif (GAN convolutif profond et GAN de Wasserstein) est introduit dans la thermographie infrarouge associée à la thermographie partielle des moindres carrés (PLST) (réseau PLS-GANs) pour l'extraction des caractéristiques visibles des défauts et l'amélioration de la visibilité des défauts pour éliminer le bruit dans la thermographie pulsée;

4. Estimation automatique de la profondeur des défauts (question de la caractérisation) à partir de données infrarouges simulées en utilisant un réseau neuronal récurrent simplifié : Gate Recurrent Unit (GRU) à travers l'apprentissage supervisé par régression.



# Abstract

Non-destructive evaluation (NDE) is a field to identify all types of structural damage in an object of interest without applying any permanent damage and modification. This field has been intensively investigated for many years. The infrared thermography (IR) is one of NDE technology through inspecting, characterize and analyzing defects based on the infrared images (sequences) from the recordation of infrared light emission and reflection to evaluate non-self-heating objects for quality control and safety assurance.

In recent years, the deep learning field of artificial intelligence has made remarkable progress in image processing applications. This field has shown its ability to overcome most of the disadvantages in other approaches existing previously in a great number of applications. Whereas due to the insufficient training data, deep learning algorithms still remain unexplored, and only few publications involving the application of it for thermography nondestructive evaluation (TNDE).

The intelligent and highly automated deep learning algorithms could be coupled with infrared thermography to identify the defect (damages) in composites, steel, etc. with high confidence and accuracy. Among the topics in the TNDE research field, the supervised and unsupervised machine learning techniques both are the most innovative and challenging tasks for defect detection analysis.

In this project, we construct integrated frameworks for processing raw data from infrared thermography using deep learning algorithms and highlight of the methodologies proposed include the following:

1. Automatic defect identification and segmentation by deep learning algorithms in infrared thermography. The pre-trained convolutional neural networks (CNNs) are introduced to capture defect feature in infrared thermal images to implement CNNs based models for the detection of structural defects in samples made of composite materials (fault diagnosis). Several alternatives of deep CNNs for the detection of defects in the Infrared thermography. The comparisons of performance of the automatic defect detection and segmentation in infrared thermography using different deep learning detection methods: (i) instance segmentation (Center-mask; Mask-RCNN); (ii) objective location (Yolo-v3; Faster-RCNN); (iii) semantic segmentation (Unet; Res-unet);
2. Data augmentation technique through synthetic data generation to reduce the cost of high expense associated with the collection of original infrared data in the composites (aircraft components.) to enrich training data for feature learning in TNDE;
3. The generative adversarial network (Deep convolutional GAN and Wasserstein GAN) is introduced to the infrared thermography associated with partial least square thermography (PLST) (PLS-GANs network) for visible feature extraction of defects and enhancement of the visibility of defects to remove noise in Pulsed thermography;

4. Automatic defect depth estimation (Characterization issue) from simulated infrared data using a simplified recurrent neural network: Gate Recurrent Unit (GRU) through the regression supervised learning.

# Contents

<b>Résumé</b> .....	iii
<b>Abstract</b> .....	v
<b>Contents</b> .....	vii
<b>List of Tables</b> .....	x
<b>List of Figures</b> .....	xi
<b>Abbreviations</b> .....	xvi
<b>Acknowledgments</b> .....	xix
<b>Foreword</b> .....	xx
<b>Introduction</b> .....	1
<b>Chapter 1 The state of art and literature review</b> .....	4
1.1 Infrared thermography in Non-destructive techniques .....	4
1.2 Approaches In infrared Thermography.....	4
1.3 Data processing in infrared thermography.....	5
1.4 Defect detection in infrared thermography NDT.....	2
1.5 Neural Network Application in defect detection for TNDE.....	2
1.6 Deep learning algorithm application for defect detection in TNDE.....	6
1.7 Basic theory of deep learning .....	7
1.8. Evaluation Metrics in Deep learning with TNDE .....	19
1.9 Modelling and stimulation .....	25
1.10 The recent research progress in literature with deep learning in IRT-NDE .....	26
1.11 The representative DL models in literature with deep learning in TNDE.....	28
1.12 The summary .....	34
References.....	35
<b>Chapter 2 Issues, Objectives</b> .....	41
2.1 Problem Statement.....	41
2.2 Objectives and Description of the problems addressed throughout this research .....	41
2.3 Methodology .....	42
References.....	45
<b>Chapter 3 Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data</b> .....	48

3.1 Résumé.....	48
3.2 Abstract.....	50
3.3 Introduction.....	51
3.4 Principles .....	54
3.5. Thermography Consideration- Optical pulsed thermography .....	55
3.6 Specimens and Experimental Setting up.....	56
3.7 Validation Datasets and features.....	61
3.8 Methodologies: Defect detection methods by deep learning algorithms.....	63
3.9 Experimental results and implementation details .....	73
3.10. Results analysis and discussion .....	88
3.11. Conclusions.....	90
References.....	91
<b>Chapter 4 Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data .....</b>	<b>96</b>
4.1 Résumé.....	96
4.2 Abstract.....	97
4.3 Introduction.....	97
4.4 Thermophysical Consideration .....	100
4.5 Automatic Defect Segmentation Strategy.....	101
4.6 Dataset and Features .....	106
4.7 Experimental Results and Implantation Details.....	107
4.8. Result Analysis and Discussion .....	118
4.9. Conclusions.....	120
References.....	121
<b>Chapter 5 Defect detection and visibility enhancement analysis using Partial Least Square-Generative Adversarial Networks (PLS-GANs) in Pulsed Thermography...</b>	<b>126</b>
5.1 Résumé.....	126
5.2 Abstract.....	128
5.3 Introduction.....	128
5.4 Preliminaries .....	131
5.5 Proposed Methodology .....	139

5.6 Results discussion and experimental analysis .....	167
5.7 Conclusions.....	169
References.....	169
<b>Chapter 6 Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning methods .....</b>	<b>176</b>
6.1 Résumé.....	176
6.2 Abstract.....	177
6.3 Introduction.....	177
6.4 Thermal Consideration and FEM Stimulation.....	178
6.5 Temperature and Thermal Contrast Curves.....	183
6.6 Proposed Strategy for Defect Depth Estimation.....	186
6.7. Experimental Validation and Results.....	189
6.8 Depth Estimation Results and Validation.....	190
6.9. Conclusions.....	192
References.....	193
<b>General conclusions &amp; perspectives.....</b>	<b>195</b>
General conclusion .....	195
Future Perspectives .....	197
<b>Appendix A List of publications.....</b>	<b>199</b>
<b>Appendix B Programming codes and Implementations .....</b>	<b>206</b>

## List of Tables

Table 3. 1 The description of each experimental sample.....	58
Table 3. 2 The training detailed from each DL model.....	73
Table 3. 3 Results with semantic segmentation and object localization algorithms.....	78
Table 3. 4 Segmentation results with instance segmented algorithms; Absolute Thermal Contrast (ATC); Raw .....	79
Table 3. 5 Different detection methods of results with individual sample .....	84
Table 3. 6 Total detection results with different deep learning segmentation algorithms ....	85
Table 4. 1 Class mark list—comprehensive results.....	115
Table 4. 2 Running time complexity comparison with state-of-the-art methods. ....	118
Table 5. 1 The geometrics and parameters distribution of three CFRP specimens .....	134
Table 5. 2 SNR values indication with different numbers of generated thermal images integrations .....	161
Table 5. 3 The comparison of SNR values from CFRP specimens with 9 selected representative defects .....	162
Table 6. 1 Physical properties and parameters of stimulation. ....	182
Table 6. 2 Defect characteristics of training samples. ....	182
Table 6. 3 Defect characteristics of testing samples.....	183
Table 6. 4 The results of depth estimation of defects located in the designated specimen.	191

# List of Figures

Figure 1. 1 Principle of the statistical method .....	7
Figure 1. 2. The conceptual description of PLS model (b) and its correspondingly classical linear regression algorithms (a) .....	1
Figure 1. 3 Two possible architectures for dual network processing .....	3
Figure 1. 4 An artificial neural network with one hidden layer and an output.....	5
Figure 1. 5 Multilayer neural network.....	8
Figure 1. 6 The chain rule of derivatives.....	8
Figure 1. 7 A forward pass in a neural net with two hidden layers and one output layer.....	9
Figure 1. 8 The equations used for computing the backward pass.....	10
Figure 1. 9 The architecture of a typical ConvNet .....	11
Figure 1. 10 A recurrent neural network and the evaluation in time of computation during the forward propagation .....	11
Figure 1. 11 Residual learning: a building block.....	14
Figure 1. 12 Example network architectures for ImageNet.....	15
Figure 1. 13 Region Proposal Network .....	17
Figure 1. 14 Confusion matrix.....	20
Figure 1. 15 The definition of mean average precision .....	22
Figure 1. 16 sample POD curve.....	24
Figure 1. 17 Stimulation data .....	25
Figure 1. 18 Thermal infrared machine vision(a)(b) .....	26
Figure 1. 19 Training error (left) and test error (right) on CIFAR-10 .....	29
Figure 1. 20 Basic Structure of an Auto-encoder (AE) model .....	29
Figure 1. 21 LSTM Block .....	30
Figure 1. 22 GAN Images generation Model .....	32
Figure 1. 23 Scheme of the neural algorithm for defect depth estimation .....	33
Figure 2. 1 Detection framework in this project.....	44
Figure 3. 1 Proposed detection strategy (Instance segmentation; Semantic segmentation; objective detection).....	55

Figure 3. 2 Pulsed thermographic testing using optical excitation.....	56
Figure 3. 3 Pulsed thermography experiment platform.....	57
Figure 3. 4 Processing of labeling .....	62
Figure 3. 5 Three types of Deep Learning methods (Objective detection; Instance segmentation; Semantic segmentation) .....	63
Figure 3. 6 The architecture of Residual units in Yolo-v3.....	64
Figure 3. 7. An example of method A (a) the original thermal image; (b) the detected image .....	65
Figure 3. 8 Faster-RCNN defect detection for infrared data .....	66
Figure 3. 9 An example of method B (a) the original thermal image; (b) the Faster-RCNN detected image .....	66
Figure 3. 10 U-net Model structure .....	67
Figure 3. 11 An example of method C (a) the original thermal image; (b) the segmented image .....	68
Figure 3. 12 Res-U-net model structure .....	69
Figure 3. 13 An example of method 4 (a) the original thermal image; (b) the segmented image .....	69
Figure 3. 14 Mask- RCNN Processing Architecture .....	70
Figure 3. 15 An example of method 5 (a) the original thermal image; (b) the detected image .....	71
Figure 3. 16 The structure of Center-Mask .....	72
Figure 3. 17 An example of method 5 (a) the original thermal image; (b) the detected image .....	72
Figure 3. 18 An instance of method 7 applied on the thermal image (a) the original thermal image; (b) the detected image.....	72
Figure 3. 19 The loss curves in each deep segmentation models (a) Mask-RCNN; (b)Center-Mask ;(c) YOLO-V3 (d) Res-U-net; (e) U-net (f) Faster-RCNN.....	75
Figure 3. 20 Different detection methods with CFRP sample.....	83
Figure 3. 21 The mean average precision curve from each deep segmentation model: (a) Mask-RCNN; (b)Center-Mask; (c) YOLO-V3; (d) Faster-RCNN.....	87
Figure 3. 22 Average frame per second for each deep learning model.....	88
Figure 4. 1 Pulsed thermography experimental setup optical excitation.....	101



Figure 4. 2 Proposed segmentation strategy .....	102
Figure 4. 3 Mask-RCNN processing architecture [6].....	103
Figure 4. 4 (a) Finite Element Modeling (FEM) 3D model; (b) Simulated thermogram at t = 106.5 s; (c) Real experimental data t = 106.5 s. ....	104
Figure 4. 5 Proposed workflow to train with a deep learning model based on the data generation by Finite Element Modeling. ....	105
Figure 4. 6 Scheme of preprocessing stage .....	106
Figure 4. 7 Labels for preprocessed sample image. ....	106
Figure 4. 8 The best obtained validation results of Mask-RCNN segmentation on different training databases. From left to right: original images, training on the preprocessed raw images database, training on the mixed database (preprocessed data from synthetic and raw images). From the first three rows to the last two rows: plexiglass (a–c), carbon fiber-reinforced polymer (CFRP) (d–e); .....	110
Figure 4. 9 The average learning loss for two types of specimens: plexiglass (a, b); CFRP (c, d).....	112
Figure 4. 10 Different detection results with four groups of datasets (two types of materials). ....	113
Figure 4. 11 Probability of distribution curve of different methods for processing on CFRP samples (a)/PLEXI samples (b) (confidence score = 0.75).....	114
Figure 4. 12 The total performance of accuracy with Mask-RCNN on CFRP and PLEXI samples with/without synthetic data.....	116
Figure 4. 13 Detection results on a comprehensive CFRP specimen provided by different objective detection algorithms or scenarios (a) Master-RCNN without synthetic data; (b) Master-RCNN with synthetic data; (c)YOLO-V3; (d) Faster-RCNN. ....	117
Figure 4. 14 Probability of distribution of different deep learning methods on CFRP databases (confidence score = 0.75). ....	118
Figure 5. 1 Pulsed thermographic testing using optical excitation system and its three-dimensional thermographic data structure.....	132
Figure 5. 2 Conceptual illustration of PLS and its comparison with classical linear regression methods.....	135
Figure 5. 3The implementation steps for GAN network (a)-(c).....	139
Figure 5. 4 The architecture of DCGAN .....	140
Figure 5. 5 WGAN generator architecture .....	142

Figure 5. 6 The training procedure flowchart of WGAN .....	142
Figure 5. 7 Defect enhancement strategy with GAN model.....	145
Figure 5. 8 Schematic representation of the transformation of the 3D thermal data into a 2D raster-like matrix.....	147
Figure 5. 9 Graphical representation and decomposition of pulsed thermography by PLST .....	148
Figure 5. 10 feature extractor of encoder and decoder format convolutional neural network .....	149
Figure 5. 11 Overall PLS-GANs framework.....	150
Figure 5. 12 The learning loss from Discriminator (a) and Generator (b) .....	152
Figure 5. 13 The generated images 64*64 from different epoch (a) epoch=0; (b)epoch=155;(c) epoch=500 .....	152
Figure 5. 14 The generated images at epoch=1000 (64*64) .....	153
Figure 5. 15 Raw CFRP-A (left side) and Generated CFRP-A (right side) sequence thermal images from DCGAN.....	154
Figure 5. 16 Raw CFRP-A (left side) and Generated CFRP-A (right side) sequence thermal images from WGAN.....	154
Figure 5. 17 Gradient-weighted Class Activation Mapping; without defect (a)(c); with defect (b)(d).....	155
Figure 5. 18 Descriptions of the analysed defect regions for visibiliy in the CFRP specimen geography(a); The labelling coordinate values of each defect region matrix (b).....	156
Figure 5. 19 Enhanced thermal images from Specimen 1 (376*376): 64 raw representative thermal images (a)-(b); PLST(c)-(d); PLST-DCGAN(e)-(f); PLST-WGAN (g)-(h) (with limited 64 raw representative thermal images merging) .....	158
Figure 5. 20 Enhanced thermal images from Specimen 2 (376*376): 64 representative thermal raw images (a)-(b); PLST(c)-(d); PLST-DCGAN(e)-(f); PLST-WGAN (g)-(h) (with limited 64 representative raw thermal images merging) .....	158
Figure 5. 21 Enhanced thermal images from Specimen 3 (376*376): 64 representative thermal raw images (a)-(b); PLST(c)-(d); PLST-DCGAN(e)-(f); PLST-WGAN (g)-(h) (with limited 64 representative raw thermal images merging) .....	159
Figure 5. 22 Gaussian denoising results from two specific GAN models (WGAN; DCGAN) from Specimen (a)-(c) .....	163
Figure 5. 23 Comparison of signal-to-noise ratio at maximum signal contrast (PSNR) for raw data; PLS; PLS-WGAN; PLS-DCGAN(three specimens(a)(b)(c) in total caculation) with 64 raw thermal images being processed .....	163

Figure 5. 24 t-SNE visualization of raw images and DCGAN/WGAN-generated images from Specimens (a)(b)(c) .....	165
Figure 6. 1 Pulsed thermographic testing using optical excitation.....	179
Figure 6. 2 Heat transfer during the thermal excitation and cooling process during FEM stimulation. ....	181
Figure 6. 3 Representative CFRP training sample 1 configuration. ....	183
Figure 6. 4 Temperature contrast evaluation for defects located at different depths with the same size (0.5 mm, 1.0 mm, 1.5 mm) in training samples (the framerate: 158 frame/second) .....	184
Figure 6. 5 Data distribution of temperature contrast evaluation for different depths in training samples (a) 0.5 mm; (b) 1.0 mm; (c) 1.5 mm (the framerate: 158 frame/second); (d) five defective point with corresponding reference point on a defect region .....	185
Figure 6. 6 (a) 3D printed defects geometrics are highlighted in training samples; (b) the corresponding synthetic colorful thermograms in $t = 3$ s. ....	186
Figure 6. 7 GRU model-based defects depth estimation strategy.....	187
Figure 6. 8 Gated Recurrent Unit. ....	188
Figure 6. 9 The process of Gated Recurrent Unit (GRU) depth estimation. ....	189
Figure 6. 10 The mean absolute error and the training loss with GRU before (a) and after (b) standard deviation normalization.....	192
Figure 6. 11 The data distribution before (a) and after (b) standard deviation normalization from all the selected locations for training (the framerate: 158frame/second).....	193

## **Abbreviations**

AE Auto Encoder

CCFIPCT Candid Covariance Free Incremental Principal Component Thermography

CFRP Carbon Fiber Reinforced Polymer

CT Contrast of Thermal Bridge

CM Condition Monitoring

CNN Convolutional Neural Network

DAC Differential Absolute Contrast

DCGAN Deep Convolutional Generative Adversarial Networks

DFT Discrete Fourier Transform

DL Deep Learning

EOFs Empirical Orthogonal Functions

Faster-RCNN Faster Region Based Convolutional neural network

FEA Finite Element Analysis

FFT Fast Fourier Transform

FPR False Positive Rate

FT Fourier Transform

GAN Generative Adversarial Networks

GFRP Glass Fiber Reinforced Polymer

IR Infrared

IRT Infrared Thermography

IRMV Infrared Imaging Machine Vision

INDE Infrared Non-Destructive Evaluation

Mask RCNN Mask Regional Convolutional Neural Network

NDE Non-Destructive Evaluation

NDT Non-Destructive Testing

PCA Principal Component Analysis

PCs Principal Components

PCT Principal Component Thermography

PLST Partial Least Square Thermography

PPT Pulsed Phase Thermography

PT Pulsed Thermography

ROC Receiver Operating Characteristic

ROIs Regions of Interests

SVD Singular Value Decomposition

TPR True Positive Rate

TSR Temperature Signal Reconstruction

TNDE Thermographic Nondestructive Evaluation

WGAN Wasserstein Generative Adversarial Networks

YOLO You Only Look Once

## Acknowledgments

I would like to express my sincere gratitude and thankful to my supervisor Prof. Xavier P. V. Maldague and deeply grateful for his continuous guidance, tremendous and invaluable supports, patience, and strong encouragement to me through the entire journey of research and study in my Ph.D. He has significantly guided me forward to my research and gave me a lot of inspiration for pursuing and achieving the highest goal of my research. Thank you so much for your support, effort and time to discipline my Ph.D. research efficiently.

I would like also to deeply thank and appreciate my supervisory committee members Professor Denis Laurendeau, Professor Moulay Akhlouf and Dr. Zhong Ouyang for your time, fully support and guidance during this journey.

Moreover, I am sincerely thankful to Dr. Clemente Ibarra Castanedo from Department of Electrical and Computer Engineering, Laval University for his continuous support and guidance which gave me lots of knowledge and valuable experience related to infrared thermography through the whole project.

I am also truly appreciating and tremendously grateful to Dr. Annette Schwerdtfeger, from the Department of Electrical and Computer Engineering at Laval University, for all of her efforts, constructive comments, and help.

I appreciate all the efforts and helps from the members of Computer Vision and System Laboratory (CVSL) and Department of Electrical and Computer Engineering, Laval University: the respected professors, the management office and staffs for their support throughout my research work.

I cherish all the moments of my research in Multipolar Infrared Vision - Vision Infrarouge Multipolaire: MIVIM Canada Research Chair-tier 1. Last but not least, I would like to thank my family and friends especially my lab-colleagues (Patrick Deschênes Labrie, Dong Pan, Qingjie Zeng, Jue Hu, Lei Lei, Bardia Yousefi, Julien Fleuret, Farima Abdolahi.Mamoudan, Samira Ebrahimi) for putting up with me and constant support, patience and continuous encouragements.

Moreover, this research was supported and conducted under the Tier 1 Canadian research chair in Multipolar Infrared Vision (MIVIM) in the Department of Electrical and Computer Engineering at Laval University. This research received funding support from NSERC DG program, the Canada Foundation for Innovation and the Canada Research Chair program. We want also to acknowledge the financial support and help of the Chinese Scholarship Council.

# Foreword

This thesis is submitted to the "Faculté des études supérieures et postdoctorales-Université Laval" to fulfill the requirement of the Philosophiae Doctor (Ph.D.) degree in Electrical Engineering.

The presented thesis is consisted of 6 chapters along with introduction; general conclusions& perspectives; appendices.

The introduction section mentioned after foreword and then the first chapter is literature review in second chapter where it over-viewed the infrared thermography with deep learning applications in relevant fields, including infrared thermography for defect detection approaches, defect characterizations methods, a very brief reviewing of infrared thermography with defect detection review literature, deep learning with infrared thermography review literature. Following this chapter, the second chapter highlighted the problems, hypothesis and objectives.

Chapter third are in the form of journal article, several approaches and experiments were conducted in infrared thermography for automatic defect detection from deep learning algorithm, then followed by other chapters (fourth to sixth) for different approaches and algorithms about: synthetic data and experimental data for infrared thermography defect segmentation by deep learning; PLS-GANs data processing techniques; defect depth estimation via deep learning in stimulated infrared thermography. All of contents includes their results and their discussions accordingly.

At the end, the results obtained in the current thesis are concluded. The last section presents a general conclusion and provided the perspective of the future of the performed studies.

The first article in the 3<sup>rd</sup> chapter is titled **Automatic detection and identification of defects by deep learning algorithms**, was submitted for Nondestructive Testing and Evaluation Journal. Authors: Qiang Fang, Clemente Ibarra Castanedo, Xavier Maldague. This chapter also extended from a conference proceeding article - **Automatic defect detection in infrared thermography by deep learning algorithm**, published in Thermosense: Thermal Infrared Applications XLII; 2020, Authors: Qiang Fang, Ba Diep Nguyen, Clemente Ibarra Castanedo, Yuxia Duan, Xavier Maldague

The second article in the 4<sup>th</sup> chapter is titled **Automatic defects segmentation and identification by deep learning algorithm with pulsed thermography: Synthetic and experimental data**, was published in Big Data and Cognitive Computing, 2021, 5(1): 9. Authors: Qiang Fang, Clemente Ibarra Castanedo, Xavier Maldague



The third article in 5<sup>th</sup> chapter which is titled **Defects enhancement and images noise reduction analysis using Partial Least Square-Generative Adversarial Networks (PLS-GANs) in Pulsed Thermography**, Authors: Qiang Fang, Clemente Ibarra-Castanedo, Duan Yuxia, Jorge Erazo-Aux, Iván Garrido, and Xavier Maldague which was accepted for the Journal of Nondestructive Evaluation in August, 2021.

Finally, the fourth article in 6<sup>th</sup> chapter which is titled **A Method of Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning** which was presented on 2020 Structural Health Monitoring & Nondestructive Testing conference (Quebec City) and published in Applied Science journal, 2020, Authors: Qiang Fang, and Xavier Maldague, A Method of Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning, MDPI Appl. Sci. 2020, 10, 6819; doi:10.3390/app10196819

Qiang Fang was in charge of designing and implementing the proposed algorithms and data collection for all experiments. He was also responsible for designing and analyzing the statistical calculations of the collected data, analysis and preparation of the manuscripts.

Professor. Xavier P.V. Maldague, research supervisor, oversaw all steps of this and provided the idea for studies for applications of infrared thermography and guided the process.

Dr. Clemente Ibarra Castanedo also provided the scientific support throughout the project.

Mr. Iván Garrido, Mr. Ba Diep Nguyen, Mr. Jorge Erazo Aux, Mrs Yuxia Duan were involved in the preparation and designing of our partially experiment worked for part of the data validating and performing the experiments. Mrs. Farima Abdolahi.Mamoudan has contributed to the partial COMSOL simulation, numerical implementation for experimental work in this project.

Furthermore, the candidate has presented this research in several different scientific conferences such as International Congress on Thermal Infrared Applications XXXVIII (Thermosense), April 2020, California, USA (remote, electronic venue); 15th Quantitative InfraRed Thermography Conference, September 2020, (Porto) Portugal (remote, electronic venue); 2020 Structural Health Monitoring & Nondestructive Testing conference, Quebec city (remote, electronic venue) and going to present the research at 20th World Conference on Non-Destructive Testing, Songdo Convensia, Incheon, South korean, 2022.

Interestingly, our manuscript was pre-selected for QIRT journal publication, article titled: Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data.

# Introduction

Infrared inspection techniques have been widely applied to evaluate subsurface defects and hidden structures etc. for the quality control of materials such as metals, composites and so on. Due to the analysis of thermal wave propagation and attenuation, each thermal non-homogeneity perturbs the thermal waves propagation on the specimen surface in comparison to the surrounding sound area and we can thus see the changes of the temperature variation. The infrared camera and corresponding equipment can record this thermal perturbation and the results can be analyzed to obtain further information. This approach also has some appreciated advantages such as safety, comparatively low cost, and non-invasiveness etc.

The main object of Infrared Non-Destructive Evaluation (INDE) is to detect and classify the Regions of Interest (ROIs) that could be represented as a defect or an anomaly by analyzing the sequence images or singular image. In addition, another goal is to detect such ROIs as accurately as possible, visible when the conditions are maintained invariantly. For defect identification and segmentation, the Pulsed Phase Thermography (PPT) [1], Principal Component Thermography (PCT) [2], Difference of Absolute Contrast (DAC) [3], Thermographic Signal Reconstruction (TSR) [4], as well as Candid Covariance Free Incremental Principal Component Thermography (CCFIPCT) [5] are powerful tools which have been documented to give noticeable results.

These mentioned methods show remarkable efficacy in improving defect visibility during data processing in infrared thermography. Whereas, all of these methods are traditional regular pattern-based unsupervised data processing which leads us to seek for innovative methods to further enhance the visibility of defect, analyze the inverse problem (defect characterization) in the industrial application, all of which are still crucial issues in Thermographic Nondestructive Evaluation (TNDE) literature.

Moreover, the Condition monitoring (CM) [6] is also a crucial topic in non-destructive evaluation to avoid the unnecessary expense of supervising and preserving the lifetime quality of a machine or aerospace materials from the systems in order to recognize abnormal behavior of the components from the machine. Regular inspection of machine and materials by a human can easily be hampered by the effects of fatigue. Therefore, an automated inspection system compatible with high inspection rates to avoid human inspector fatigue, to meet accuracy and quality demands for quantitative analysis, and maintenance costs is highly recommended in the non-destructive evaluation of infrared thermography in like manner.

Since the beginning of the twenty first century, the deep learning [7] (deep neural network)-one of the fields of artificial intelligence, has made remarkable progress. Initially developed in order to make neural networks more efficient, deep learning neural networks have shown their capability to outperform most of the other approaches existing previously in a great number of applications and this has contributed to make them very popular in many scientific

communities. However, one must consider that the computing expense caused by the massive structure and multilayers of deep learning in its process and requirement of training datasets to be labelled and used in network which still remains unexplored in TNDE projects. Only a few works in the literature dealing with defect detection in active infrared thermography coupling with deep learning algorithms due to training complexity and training data limitation.

As a result, there are some innovative points which could further improve the feasibility of deep Learning models with infrared thermography. In this project, we are going to exploit the proposed deep learning algorithms during the infrared non-destructive evaluation for achieving automatic defect detection in order to extract and separate defects efficiently and accurately in an industrial NDT system while evaluating these approaches comparatively to the other state-of-the-art methods and obtain defect inverse information for those infrared signals (including defect characterization issues: defect depth, etc.) as well as enhance the visibility of defects (including internal and less visible cracks, delamination, subsurface defects case, etc.) by deep learning algorithms. Measurements are going to be conducted using active infrared thermography.

Moreover, these innovative deep learning techniques will be validated and analyzed to enable them to be fully compatible to being applied with the infrared non-destructive evaluation cases. Several representative pretrained deep learning models will be introduced as a generic feature extractor and combined with an infrared thermography database to learn specific and detailed features during the learning process.

It is worth noting that there are several scientific challenges involved in the successful application of these techniques in the projects which includes:

1. The limited training data of thermography for deep learning. Due to issues such as the presence of noise, limited resolution, and probing of depths in TINDE, it is more difficult to precisely and fully extract defect information;
2. For the defect characterization issue, it is complex to combine the information of thermal contrast and pixels of images with the depth in each pixel of the defect to train the neural network;
3. Analysis with the cases of samples which contain noise and non-uniform heating; weak detection signal; limited IR data; etc.

Our most compatible deep learning models and methods with infrared imaging to overcome these challenging issues in defect detection, to obtain compelling results through the subsections are going to introduce and illustrate from chapter 4 – chapter 7 in this thesis as follows: a. Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data (chapter 4); b. Synthetic and experimntal data for defect segmentation in deep learning (chapter 5); c. Defects visibility enhancement and analysis by

deep learning in IRT (chapter 6); d. Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning (chapter 7).

## References

- [1] *Maldague X, Galmiche F, Ziadi A. Advances in pulsed phase thermography[J]. Infrared physics & technology, 2002, 43(3-5): 175-181.*
- [2] *Rajic N. Principal component thermography[R]. DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION VICTORIA (AUSTRALIA) AERONAUTICAL AND MARITIME RESEARCH LAB, 2002.*
- [3] *Pilla M, Klein M, Maldague X, et al. New absolute contrast for pulsed thermography[C]//Proc. QIRT. 2002, 5: 53-58.*
- [4] *Bison P, Cadelano G, Grinzato E. Thermographic Signal Reconstruction with periodic temperature variation applied to moisture classification[J]. Quantitative InfraRed Thermography Journal, 2011, 8(2): 221-238.*
- [5] *Weng J, Zhang Y, Hwang W S. Candid covariance-free incremental principal component analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8): 1034-1040.*
- [6] *Janssens O, Van de Walle R, Loccufier M, et al. Deep learning for infrared thermal image-based machine health monitoring[J]. IEEE/ASME Transactions on Mechatronics, 2017, 23(1): 151-159.*
- [7] *“Deep Learning[M]”. Goodfellow I, Bengio Y, Courville A. The MIT Press, 2016.*

# **Chapter 1 The state of art and literature review**

## **1.1 Infrared thermography in Non-destructive techniques**

Non-destructive evaluation (NDE) has been widely applied in science and industry fields to detect the different kinds of material properties without causing damage. Due to the fact that NDE does not permanently change and modify any sample and article being inspected, this has become a highly worth technique that can reduce the cost and save time for any product inspection and research.

In the non-destructive evaluation research field [1], traditional methodologies include eddy current testing (ECT), ultrasonic test, magnetic particle testing (MPT), and Infrared Thermography. Due to its rapid, contactless, quantitative analysis capabilities, and remote imaging infrared thermography is widely applied in the defect detection for various materials.

Therefore, the development of commercial infrared cameras, the main equipment in performing infrared thermography, has been improved in spatial resolution and sensibility. Experiments can thus be performed more rapidly and at a lower cost.

## **1.2 Approaches In infrared Thermography**

The traditional infrared thermography can be separated into [1] two different branches: the object or system has a distinguishable thermal contrast with respect to the ambient environment. We could use an infrared camera to display the whole process in the computer system directly.

In active thermography like lock-in thermography (LT), the periodic deposition at a sampled frequency is used in the stationary regime to monitor the thermal response when it has amplitude and/or phase delay. In pulsed thermography (PT), during a short period time, a thermal stimulation pulse is used to heat t sample from inspection. The surface temperature is measured under the theory that a different temperature cooling down period will take place in the defect area with respect to the surrounding sound area. Square pulse thermography (SPT); step heating (SH); point or line scan thermography (LST) and so on, are additional methods that can be used. These methods have been designed through specific configurations with specific applications [1].

## 1.3 Data processing in infrared thermography

### 1.3.1 Thermal Contrast-Based Techniques

Thermal contrast is the basic concept for the fundamental cooperation in Pulsed thermography. Many types of thermal contrast all share the common demand to define a sound region, which is a surrounding non-defective area with respect to the subsurface defect.

The classical definition equation (1.1) is used to compute the thermal contrast which is named the absolute thermal contrast [2]:

$$\Delta T(t) = T_{d(t)} - T_{Sa(t)} \quad (1.1)$$

where  $T_{(t)}$  is defined as the temperature at a given time  $t$ , the  $T_{d(t)}$  is the temperature of a pixel or the average value of a group pixels in the inspected defective location, and  $T_{Sa(t)}$  is the temperature at time  $t$  for the sound area  $S_a$ .

Under normal conditions, we identified a sound region in the image either automatically or by some operator. Then the thermal contrast can be computed in all the thermogram with corresponding time sequence. The advantages for computing  $\Delta T$  are a better visualization of the defect and the visible comparison with respect to the background.

### 1.3.2 Differential Absolute Contrast

Another definition is differential absolute contrast (DAC) [3], using a sound region, we use some initial images (at least one at time  $t'$  in particular) to compute the sound area temperature  $T_{Sa}$ . Suppose that at time  $t'$ , this local point temperature in a sound area  $S_a$  is the same as in an area containing a defect even though no defect is visible with:

$$T(0, t) = T_o + \frac{Q}{e\sqrt{\pi t}} \quad (1.2)$$

Hence this  $t'$  will be defined as a time value during the period that the pulse has been launched at  $t_0$  to the precise moment that the first defective spot appears on the thermogram sequence  $t_1$  [1].

Due to fact that at  $t'$  no existence defective region yet, therefore the sound region local temperature has the exact same value with the defective area [4]:

$$T_{Sa}(t') = T(t') = \frac{Q}{e\sqrt{\pi t'}} \rightarrow \frac{Q}{e} = \sqrt{\pi t'} \cdot T(t') \quad (1.3)$$

### 1.3.3 Principal Component Thermography

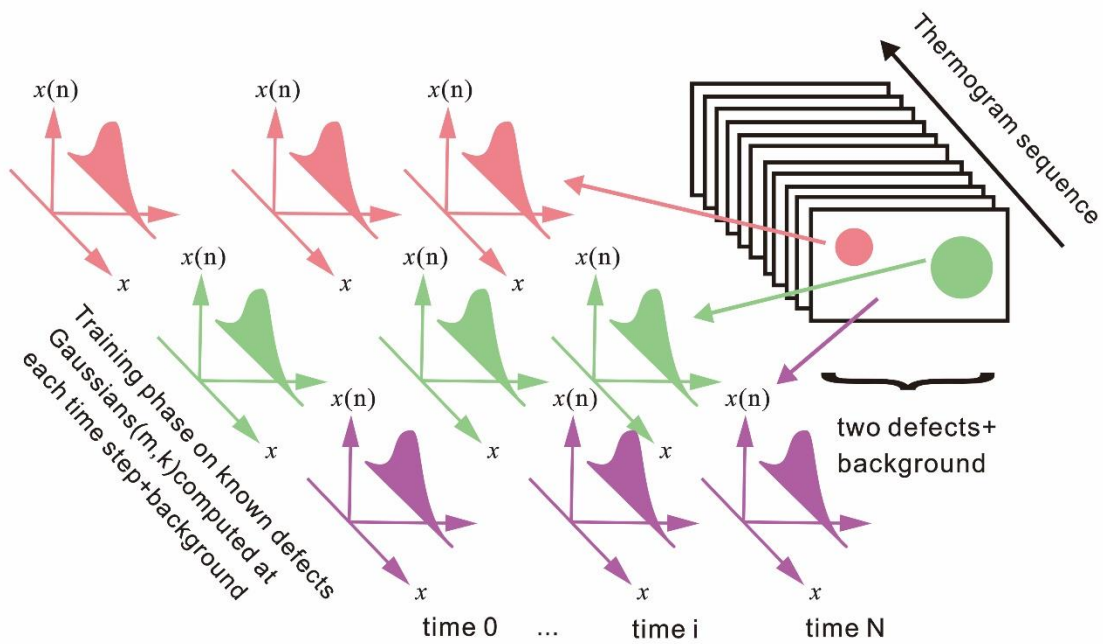
Principal Component Thermography (PCT) is the data extraction tool to improve the visibility of defects in thermal images through the dimension reduction. [5] However due to the computing expense for the covariance matrix, it still suffers from the complexity of the analysis and the time-consuming nature of the computation.

In this topic, PCT divided the data into a group of Empirical Orthogonal Functions (EOF) obtained by Singular Value Decomposition acquired through S. Supposing data can be illustrated as a  $W \times H$  matrix  $X$  ( $W > H$ ), then the Singular Value Decomposition can be described as follow [1]:

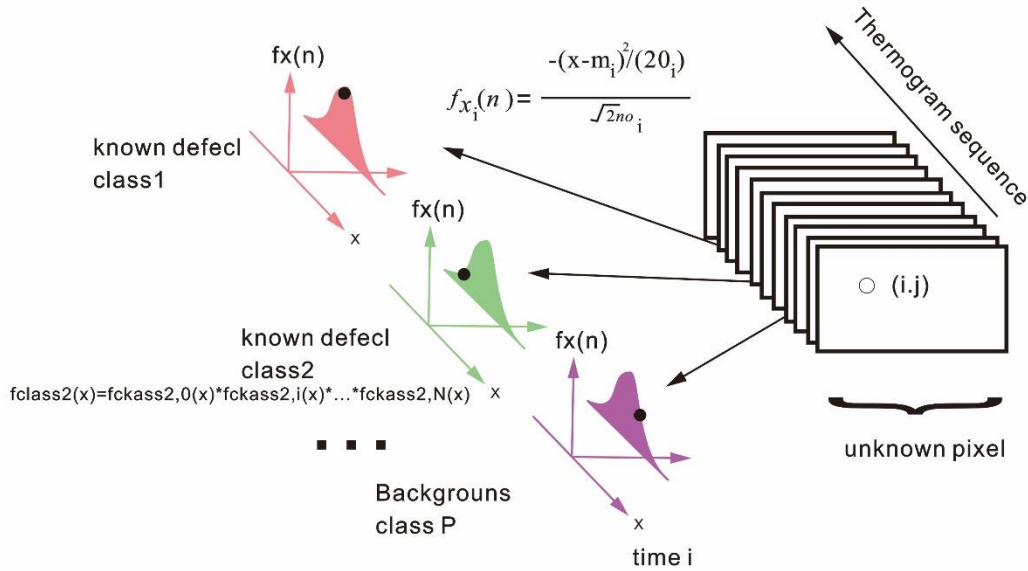
$$X = UBV \quad (1.4)$$

$U$  is a  $W \times K$  matrix;  $B$  is a diagonal  $K \times K$  matrix.  $V$  is a  $K \times H$  matrix.  $K$  columns of  $U$  involve a set of orthonormal eigenvectors corresponding to the  $K$  largest eigenvalues of  $XX^T$ . It is noteworthy that when we unfold the thermal data as a 2D matrix (columns with temporal; rows with spatial), then we know the  $K$  columns of  $U$  represent a set of EOFs.

### 1.3.4 Statistical Method



(a)



The pixel is assigned to the class with largest probability

(b)

Figure 1. 1 Principle of the statistical method

(a) Learning stage; (b) Analysis stage

(Ibarra-Castaneda C, Maldague X P V. et al: [1])

Statistical regions of interest including background and defects have been used in the application of depth classification in TNDE. The basic theory described in Figure 1.1, which is that thermographic (temperature), amplitude, phase can be modeled as a Gaussian random process [2]. In the statistical principle, two stages are included. First, in a “learning” stage, the local means,  $m$ , and standard deviations,  $s$ , are computed during each discrete time for all the classes including background and known defects, since temperature images with defects of known depths and background location are made available. Second, in the analysis step, for the image sequence, we analyzed all the pixels one by one. For the unknown pixels, computing with  $m$  and for each given pixel to be part of a given class. Suppose that individual probabilities at each time step are multiplied together to form the global probabilities. The most probable category corresponds to the largest probability value. Through this method, unknown pixels can be located at the more probable known flaw or sound region.



### 1.3.5 Partial Least Squares Thermography

Partial least squares method is a type of multivariate statistical data analysis method, which was first proposed by S. Wold and C. Albano et al. in 1983. It has been rapidly developed in terms of theory being applied in different fields recently. The partial least squares algorithm organically merged regression patterning (multiple linear regression), data structure decomposition (principal component analysis), and correlation analysis between two sets of variables (classical correlation analysis) to be implemented simultaneously under united algorithm. This is a significantly improvement in the analysis of multivariate database statistic. It is a regression modeling method of manifold dependent variables on multiple independent variables, which can furtherly solve the issues that were previously unsolvable with regular multiple regression.

The briefs theory of partial least square regression could be illustrated in the following in the below: PLSR decomposes the predictor  $X$  ( $n \times N$ ) and predicted  $Y$  ( $n \times M$ ) matrices into a series of latent variable (loadings, scores and residuals). The PLS model is illustrated as [6]:

$$X = TP^T + E \quad (1.5)$$

$$Y = TQ^T + F \quad (1.6)$$

In equation (1.5) and (1.6),  $T$  ( $n \times a$ ) is defined as the scores matrix and its elements is represented by  $t_a$  ( $a = 1, 2, \dots, A$ ). These scores can be deemed as a small amount of underlying or latent variables being responsible for the systematic variations from  $X$ . The matrices  $P$  ( $N \times a$ ) and  $Q$  ( $M \times a$ ) are named as the loadings (or coefficients) matrices which illustrated the relationship between the original data matrices  $X$  and  $Y$  and the variables in  $T$ . At the end, the matrices  $E$  ( $n \times N$ ) and  $F$  ( $n \times M$ ) are also named residuals matrices which means the noise or irrelevant information from  $X$  and  $Y$  separately.

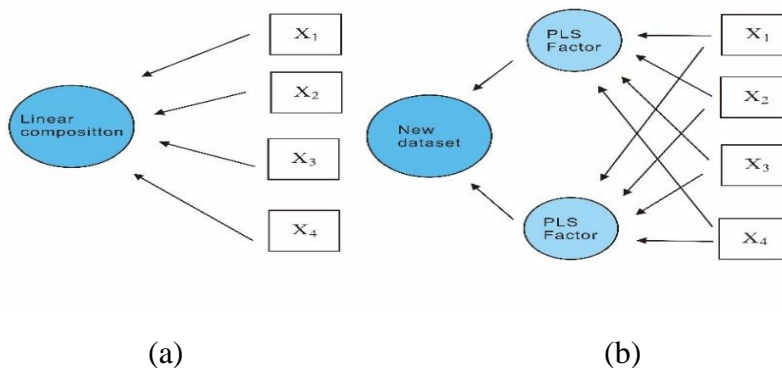


Figure 1. 2. The conceptual description of PLS model (b) and its correspondingly classical linear regression algorithms (a)

## 1.4 Defect detection in infrared thermography NDT

### 1.4.1 Defect Detection

The most important application in TNDT is to detect the existence of subsurface defects. The traditional methodology involves checking out dynamic thermal data on the screen by the operator who has a dynamic eye-brain entity filtering to reduce noise. Once trained well, we could be able to see the difference between the defect and the non-defect area based on this operator.

In order to further improve the inspection to get more obvious and reliable detection, a new innovative method has been used. This method set up a threshold  $T_{th}$ . For a thermal contrast image, then if pixel values  $>T_{th}$ ; they will be set to 1, otherwise, they will be 0. In processed images, defects are indicated by 1 value pixels. Besides, to solve the difficulty of selecting threshold, distance criterion has been proposed to achieve the automatic defect detection [2].

### 1.4.2 Defect Characterization

The original defect sizing approach is to compute the thermal contrast image's gradient when it arrives the maximum. Since defect visibility reaches its maximum at highest contrast, therefore this moment is easier for defect detection. However, the edge of the contrast curve is also blurred since if there the thermal front spreading tridimensional during maximum thermal contrast moment. The mean of an iterative technique was proposed: first recording the thermal data by pulse thermography during the elapse time, then for every single thermogram the defect size is recorded as the thermal contrast arriving at the half of the maximum amplitude, afterwards a plot of the defect size is defined by the function of square-root with time.

The interesting parameters for the quantitative characterization are temperature contrast values computed at specific times through thermal contrast-based methods including: time of maximum contrast  $t_{max}$ , half maximum contrast  $T_{1/2}$ , time of maximum slope  $T_{slope}$ , or time of the beginning of thermal contrast. During the traditional specific experiment, to extract these parameters from the recorded thermogram sequence could be instrumental to subsurface detection [1]. Regarding the evaluation of the defect characteristics for the depth  $Z_{def}$ , it can also find some similarity characterization with the thermal resistance  $R_{def}$  and the size which is essentially represented by defect diameter  $D_{def}$ .

## 1.5 Neural Network Application in defect detection for TNDE

A neural network is an efficient mathematical model to analyze diverse and non-linear reality issues, due to the potentially powerful ability to detect and classify targets under changing

conditions of signature or environment. Through the provided training data, networks can generate the learning information from these examples for the data, not contained in the training database.

The single neuron block model can be described below. We can also make an analogy between a neuron and a cell with several inputs  $X_1, X_2 \dots, X_n$ , and one scaler output. All the inputs from each neuron multiply with corresponding weight  $W_1, \dots, W_n$ , and then these products are accumulated linearly. The global scaler  $S = X_1 \cdot W_1 + X_2 \cdot W_2 + \dots + X_n \cdot W_n$ . The global scaler  $S$  is an input for another activation function:

$$Y = f(S) \tag{1.7}$$

This activation function  $f(.)$  is a threshold function corresponding to a different mathematical transformation. Based on this single neuron block, there are many possible diverse architectures which can involve a multitude of layers, longer feedforward or back-propagation procedure.

During the 1990s, the researchers started to use neural networks as a TNDE defect detector and classifier for analysis detection issues. In 1993 D. R. Prabhu et al. [8] applied two Neural networks to detect subsurface corrosion including a network to detect flaw and a corrosion estimation network. He tried to use these two networks through Parallel or Serial architecture connection to process data. Experiment results reveal that to use both these two networks can obtain greater results as shown in Figure 1.3 than only using one of network to train separately. Beside it reveals clearly that a huge training database also requires a longer period.

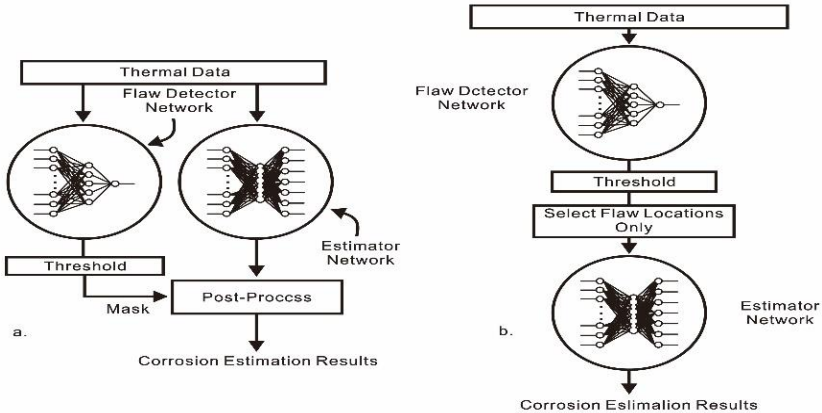


Figure 1. 3 Two possible architectures for dual network processing

- a. Parallel architecture; b. Serial architecture (Martin, R.E., Gyekenyesi, A.L. [3])

In 1997, H. Tretout, et al. [9] used the variation curves of temperature in time to represent each pixel of images. Then to proceed to spatial analysis, a neural network is used to train and distinguish the difference between the thermal curves of the defects from the non-defect regions curves. In order to choose the input of Artificial neural network, the authors selected three groups of curves respectively: raw curves; derivative of raw curves; contrast curves to make the comparison. The results prove that training on the curves of temperature contrast could improve the situation, but a precise theoretical base is also required to generate accurate temperature reference curves.

In 1998, X. Maldague et al. [10], presented the first study to acquire quantitative data basing on the analysis pulsed phase thermography (PPT) results. During the experiment of this paper, the neural network has been used to test on the aluminum and plastic specimens for detecting the depth of each subsurface defect. The results revealed that the quantitative analysis basing on the neural network is limited by the insufficient sampling frequency of the thermal diffusivity with respect to the interest materials.

In 2000, S. Vallerand et al. [11] proposed an innovative statistical processing method. This method was compared in terms of defect detection and characterization with the machine learning method, involving two different architectures of neuron networks (Perceptron (PNN) and Kohonen network (KNN)). The proposed novel method was also compared with the traditional statistical method: Gaussian random process. During the training, PNN is using supervised learning due to the training samples that are provided. However, unlike perceptron network, KNN involves unsupervised learning since this network determines how to classify the training samples into several groups. As a result, PNN implemented a greater performance than KNN. The reason for the poor performance in KNN is because the lack of quantity information. Therefore, the networks are not able to adjust the weights and bias. It also reveals that enough data is an Achilles heel during the neural network training.

More detailed neural network analysis with TNDE was undertaken in a new study entitled. The new study in defect detection and characterization of defect depth with neural network shown by A. Darabi, et al. [12]. In this paper, authors used neuron network to extract the defect feature with carbon fiber reinforced plastic (CFRP). Two types of networks with multilayer have been applied. The running-contrast data curves vector was regarded as the input-output pair during the defect detector network training. Inversely, defect depth estimation networks keep the contrast vectors as input, replacing pixel depth of contrast vectors as output. These networks train with simulated and experimental data. The experimental results further prove defects quantities information can be successfully characterized in TNDE.

In 2008, H. D. Benitez et al. [13] proposed a feature extraction of data basing on differential absolute contrast (DAC) with artificial neural networks to identify the defects in composite materials. They combined the DAC with thermal quadrupoles in order to reduce the non-uniform heating and the influence by the shape of plate in the inspection. This method enhanced the visibility of defective area from the sound area and eliminated the necessity of picking up a sound area.

In [14], H. D. Benitez et al. developed a new approach for defect characterization in TNDE with learning machines (neural networks, support vector machines, radial based function). In this study, these learning machines combined together to evaluate the performance of defects detection in composite materials. However, during this study, it can also show the disadvantage of a neuron network which is only useful for testing the similar type material when it trained. And it required a group of defects and samples to train the system.

In 2012, L. Junyan et al. [15] proposed a single hidden layer network to estimate the depth of subsurface defects in lock-in thermography. Experimental samples included composite and metal materials. In this study they use a feedback network architecture to obtain two neuron cells output layer respectively corresponding with identifying depth and measuring defect depth as shown in Figure.1.4.

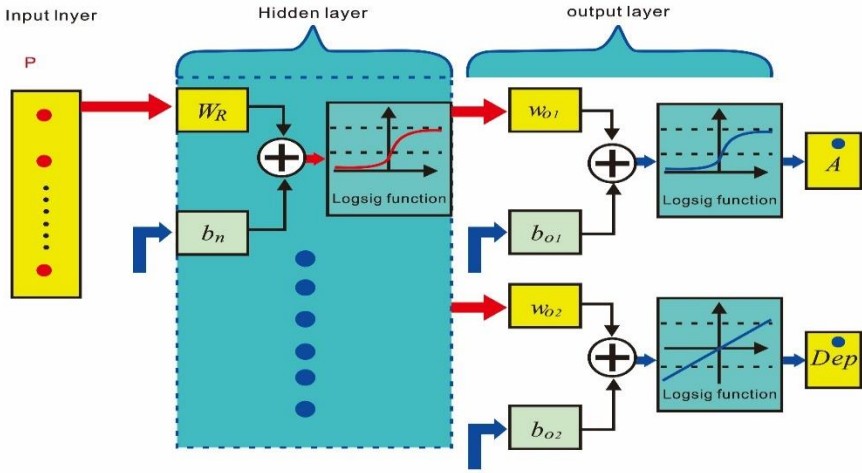


Figure 1. 4 An artificial neural network with one hidden layer and an output

(L. Junyan et al. [15])

During the last five years, neural networks were still actively applied for the subsurface defect’s detection with composites or metal materials in TNDE research fields by J H. Sheng et al. [16] 2013; R. Marani et al. [17] 2016; H. Halloua et al. [18] 2017. Artificial Neural

Networks were used to characterize the defect depth, shape, and location. Feature extraction through neural networks has shown its ability to predict the results basing on test samples as accurately as possible with the ground truth of data, and generate the learning information from those samples for the data are not contained in the training database. However, to improve the efficiency of common neural networks, fitting the data with the deeper learning structure is still an interesting topic which requires further investigating with this project.

In 2019, Yuxia Duan et al [19] design a neural network to classify the type of defects in the composite materials which reported how to adapt a neural network combine with thermal databases to classify the types of defects such as air, oil, and water which could be able to influence the aircraft material performance. The thermal databases from two different resource experiment and finite element method were used to train and extract feature. The original thermal data, and the coefficients of thermographic signal reconstruction were being used to train, test, evaluate with two layers neural network. Furtherly, the Quantitative comparisons and analysis showed that the data feature extracted from using coefficients as performed better than the one using original data which show more precise result and test repeatability and reveal more generalizable.

## **1.6 Deep learning algorithm application for defect detection in TNDE**

In comparison to the common neural network, deep learning is making a huge improvement in solving issues which were difficult to handle in the artificial intelligence research fields for many years. It has turned out to be impactful at discovering the meaningful feature from the highly and complex dimensional structures. As a result, it was applicable to different kinds of research fields. Besides, in regards to beating records in computer vision, it has surpassed other learning techniques in predicting and analyzing the activity basing on scientific data especially in image recognition and classification.

Why use a deep neural network?

As we can see, in the data trained model, detailed and enough parameters to train data can lead to a much better performance for the trained results to some extent. However, in real training issues, collecting a huge amount of suitable data is a computationally expensive task. During the deep neural network training, every neuron represents a basic classifier (every classifier for a different attribute) and the neuron of every layer is shared by the following layer's neurons as a module to build classifiers, and this modularization is learned from data. That is why we are using deep neuron networks to extract more effectively data used to train a learning system, in comparison to other artificial intelligence learning methods.

Through the investigation in the field of TNDE, the research literature on infrared thermography for defect detection using deep learning algorithms is a new and relatively

unexplored topic. Based on the relevant research which scholars have carried out on detection of defect in machine vision and neural networks during the last decades, to develop a precisely enough and faster learning feature deep learning algorithm will have a meaningful impact.

For this project, we propose to design the advanced deep learning algorithm to extract more information from IR thermography for Non-Destructive evaluation in order to achieve infrared defects visibility enhancement; automatic defect detection and segmentation; defect depth estimation; thermal data sequence model reconstruction etc. This would also lead to improvement in the detection of deeper subsurface defects for which IR signals are limited.

## **1.7 Basic theory of deep learning**

### **1.7.1 Supervised learning**

During the training in deep learning, regarding shallow or multi-layer of architecture, the supervised learning [20] has been the most common and widely used form of learning method. Before the training, first we collect a huge amount of data such as a set of images, each of data has been labelled with own corresponding kind. At output of this training, a series of score vectors have been shown on the images or other kinds of data. The target for this training is to have the highest score for the desired category in all data. However, this situation is less possible to occur during the initial stages of learning. We established a loss function to minimize the error between prediction scores and the objective scores. The neural network adjusts the internal variables with respect to each neuron of all the layers based on the error from loss function. We define these internal variables as weights which are parameters that could make an analogy with ‘knobs’ to define this machine. It is also noteworthy that in a complex construction learning system, a huge amount of adjustable and labelling data must be provided to train this machine.

### **1.7.2 Unsupervised learning**

The researcher introduced unsupervised learning [21] to design feature detectors layers without labelling data. Therefore, the objective for feature detectors of each layer learning could be possible to reconstruct the feature information from the previous layers. In practice, through the ‘pre-training’ procedure, complex highly dimensional feature detectors evolve from several layers using the objective from the reconstruction, and the weight can also be set up as sensitive values initially. In the end, in the output of the final layer, the reconstruction feature could be added at the top of the network and then the fine-tune training procedure start up at the entire deep network.

### 1.7.3 Multilayer neural networks and backpropagation

From Figure 1.5, the input space can be distorted by the multi-layers of the neural network. Neural networks also classify the data (such as the data on the red and blue lines) linearly separable. We also see the input feature space (in the left) and then hidden feature space transformed from input (in the middle). In addition, one can see 2 input neurons; 2 hidden neurons; one output neuron in this architecture. However, usually tens or hundreds of thousands of units can be contained in the networks used for object recognition or image segmentation.

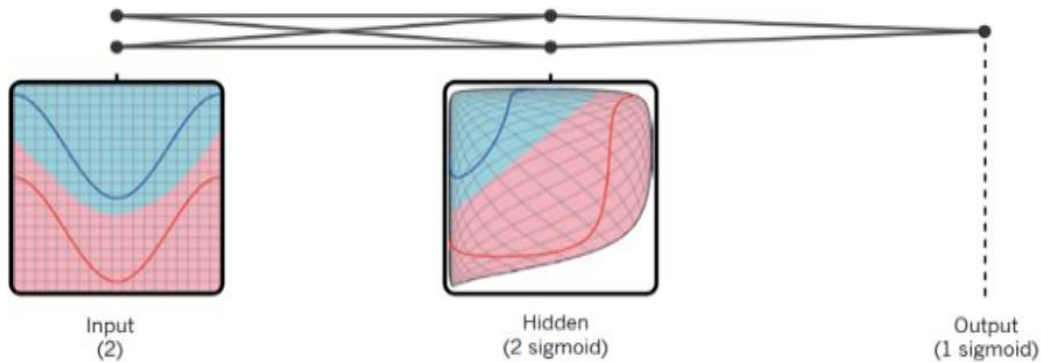


Figure 1. 5 Multilayer neural network

((Deep learning Review Yann LeCun et al [22])

### 1.7.4 The Chain Rule of derivatives

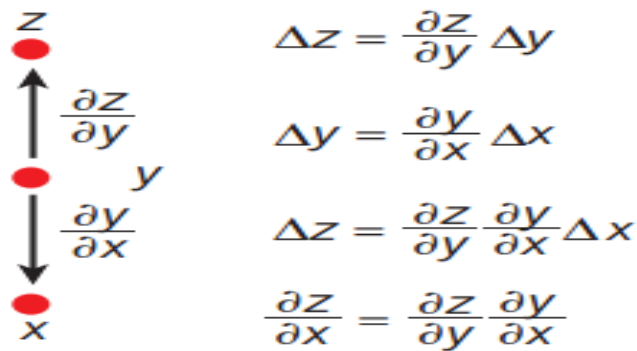


Figure 1. 6 The chain rule of derivatives

(Deep learning Review Yann LeCun et al [22])



The chain rule of derivatives as shown in Figure 1.6 indicated that the small effects from  $x$  finally end up at  $z$  and also shows the connection from input  $x$  to the final output  $y$ . When a small difference of  $x$  changed  $y$ , then  $\Delta y$  transformed the difference to the final output  $z$ . In the mathematical description, the variation of  $\Delta x$  multiplied by the partial derivative definition  $\partial y / \partial x$  to create a change of  $\Delta y$  in  $y$ . Similarly, the variation of  $\Delta y$  creates another change of  $\Delta z$  in  $z$ . The researcher got this inspiration from the equation substitution. Besides, this chain rule also works when all the format of data is in the Jacobian matrices).

**1.7.5 The forward pass in a neural net**

From the equations in Figure 1.7, the complete forward-propagation is shown with the hidden and output layer. Each neuron of layers constructed a learning block to compute the results of feed-propagation. In the first steps, the total input  $z$  is are computed from each unit, then this input is multiplied with the weights from its corresponding unit to construct a total weight sum to the output unit. In this case, we omitted bias terms for simplicity. There are several non-linear functions which are useful in neural networks such as the rectified linear unit (ReLU):  $f(z) = \max(0, z)$ , as well as the logistic function,  $f(z) = 1 / (1 + \exp(-z))$ .

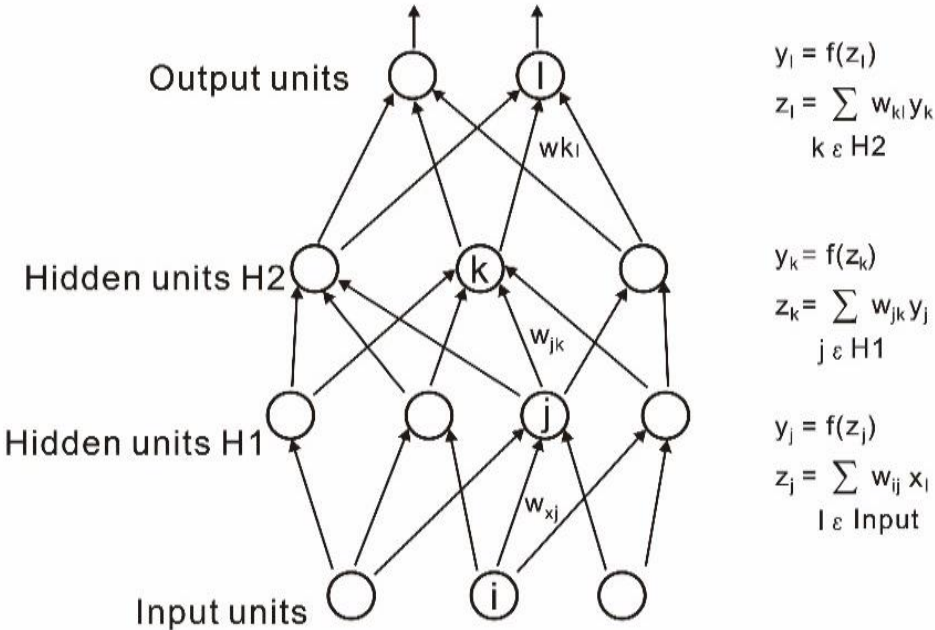


Figure 1. 7 A forward pass in a neural net with two hidden layers and one output layer  
 (Deep learning Review Yann LeCun et al [22])

### 1.7.6 Back propagation

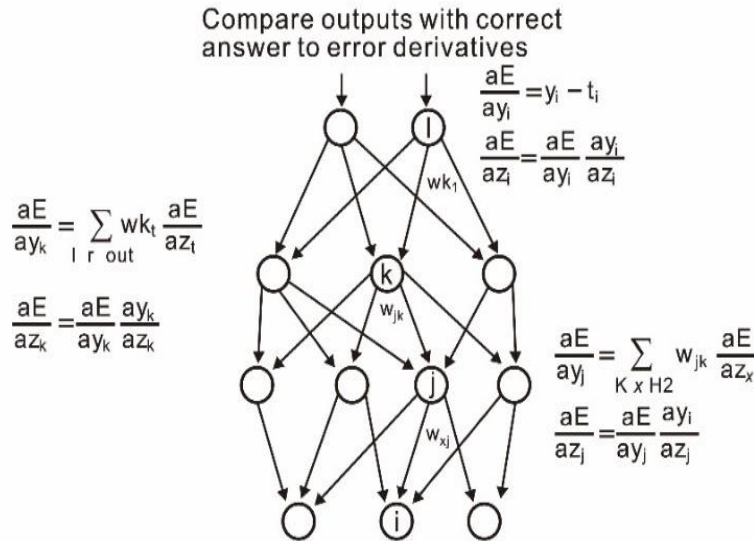


Figure 1. 8 The equations used for computing the backward pass

(Deep learning Review Yann LeCun et al [22])

Figure.1.8 shows the difference derivative of the output regarding the units in each layer represent by a weighted sum of difference derivative from the layer above totally. Afterward, these difference derivative regarding the output is converted into the inputs by multiplying the gradient from  $f(z)$ . At the output layer, the error derivative in regard to the output unit is computed by differentiating the cost function. For example, if the cost function for unit 1 is  $0.5(y_1 - t_1)^2$ , where  $t_1$  is the target value, the error derivative with respect to the output is  $y_1 - t_1$ . Besides, once the  $\partial E / \partial z_k$  is known, the error-derivative for the weight  $w_{jk}$  on the connection from unit  $j$  in the layer below is simply  $y_j \partial E / \partial z_k$ .

### 1.7.7 Convolutional neural network

Convolutional neural network is created in order to achieve feature extraction and learning through the multiple arrays data. For example, a color image is represented by a 3D matrix. This 3D matrix contains 2D spatial pixel and three dimensions of color category. At the same time a lot of data modalities could be regarded as multiple arrays, such as: 1D Sequences of signal: language; 2D images; 3D for the video or sequence images. Behind this Convolutional neural network, we can see four important ideas from multilayers; the information of signals; the sharing of weights; the connections of location of pixel.

As shown in Figure.1.9, the authors used a classical convolutional architecture network to detect a dog in the image. The outputs from different levels of layers are horizontally illustrated in the Fig. The authors regarded each rectangular image as the learned features output extracted from the raw images to detect the properties of the images and corresponding to each position in the images. The diverse feature information flows from bottom to top to represent the different level of feature extraction, such as the primary level features working as the rough edge detectors. Corresponding to each image class, a score is computed in output.

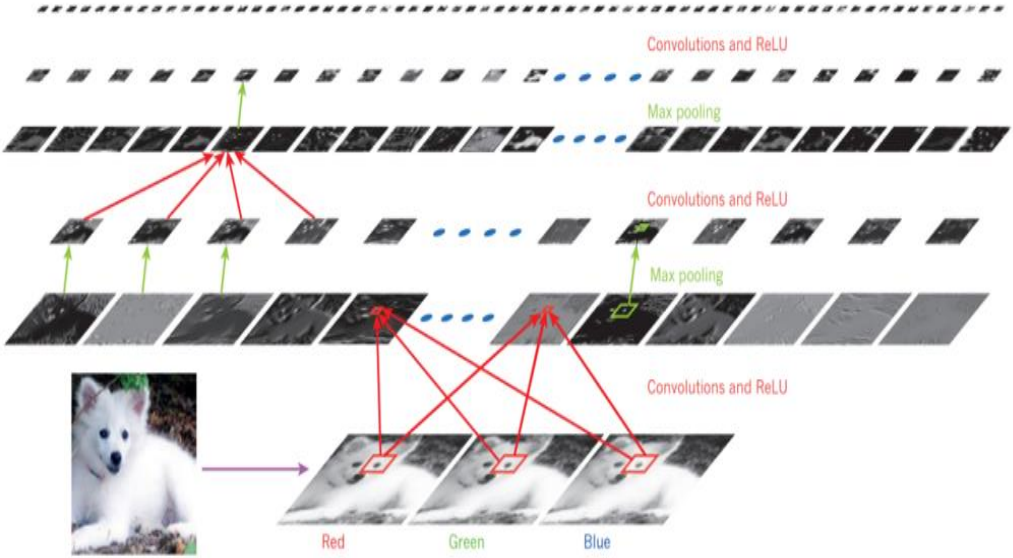


Figure 1. 9 The architecture of a typical ConvNet  
(Deep learning Review Yann LeCun et al [22])

**1.7.8 Recurrent neural networks**

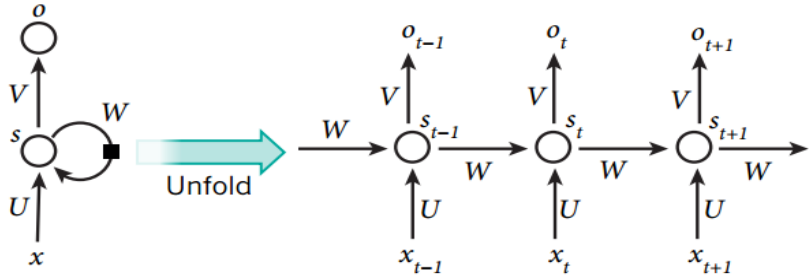


Figure 1. 10 A recurrent neural network and the evaluation in time of computation during the forward propagation

(Deep learning Review Yann LeCun et al [22])

For some projects that are relevant to sequential inputs, it is often a wise choice to apply the recurrent neural networks (RNNs) in it. As shown in Figure.1.10, RNNs process the input data each element at a time respectively. At the same time a 'state neuron' save all the memory from the past history for each single unit of the sequence. We can also make an analogy of the outputs from the hidden units in every single time like the outputs form the different neurons in one hidden layer of a convolutional neural network. For the training, it is also problematic since the shrink and explosion of gradients in each discrete time may result in vanish or explode of gradients finally.

RNNs, once having the evaluation through the discrete time, can be regarded as a rather long feed-forward network, where all the layers share the same weights. Although the main objective for the RNN is to extend terms of dependencies for learning, empirically, an amount of experiments shows that it is difficult to store information too long.

Besides an innovative idea is to enlarge the network with enough memory. The first inspiration for this idea arises from long short-term memory (LSTM) networks. This network owns one unique hidden unit to save the memory from the inputs for a long time. This unit acts like an accumulator, having a self-connection for the next time corresponding to a weight. Therefore, it saves its own real-value and accumulates the external source. This self-connection unit combined with another one by multiplication of the values to make decisions of whether or not to clean the memory.

Subsequently, LSTM networks have improved the efficacy of the memory of the network in comparison to RNNs, in particular for the case when it has multi-layers for every single step. Besides, nowadays LSTM or similar forms of gated units are useful for combining with the different types of encoder and decoder networks to improve the performance.

### 1.7.9 Transfer learning Strategy

Deep multilayer neural networks have been proven to outperform the state of the art for difference image processing databases. Especially CNNs illustrated an impressive performance in image segmentation. In classical deep neural networks, a certain amount of data is required to learn the features efficiently. However, limitation of the infrared thermography image dataset in this project is not the same case.

In practice, researchers rarely train an entire Convolutional network from scratch due to the lack of sufficient number of datasets. To extract meaningful deep features, usually we use a pre-trained ConvNet training on a huge dataset and then take the convolutional neural network as a generic feature extractor with fixed parameters initially for each particular task [23]. The main Transfer Learning scenarios can be described as follows:

- **Using ConvNets as a fixed feature extractor**

Pre-training the Convolutional neural network on ImageNet then extracting from the final fully-connected layer corresponding to a series of class score for each particular task and treating the remaining ConvNet as feature extraction for new datasets. Besides, different pre-trained network will output different dimension vector for the images. Those vectors are output from before the activation in hidden layers. For example, we used the ReLU activation to thresholded feature at zero. Finally, a softmax classifier or SVM could be trained on to classify the data.

- **Fine-tuning the ConvNets as Pre-trained Model**

Secondly, not simply replacing and retraining the classifier on the later layers in a network for a new database, but also fine-tuning the parameters through back-propagation to learn the feature. There are two options for fine-tuning: tuning all the layers for this ConvNet or only fine-tuning the higher-level portion due to the over-fitting issue. The reason is inspired by the observation that the general features (e.g. edge detectors) are always contained in the earlier layer in a ConvNet that could be useful in other tasks. However, we use the layers in the later of network to detect more specifically feature distribution of data.

In this project, there are three available options for pre-training CNNs: VGG, ResNet features which can help us overcome the lack of enough training data in infrared thermography for deep learning. Indeed, the different types of pre-trained architecture convolutional neural network have been widely applied in the computer vision research community.

- **VGG features**

VGG-16 is a deep convolutional network for object recognition developed and trained in Oxford's Visual Geometry Group. [24]. VGG is pre-trained on more than one million images. A Pre-trained network can be regarded as a generic feature extractor to achieve learning of

transfer [25]. Considered as a feature extractor, the vectors from the outputs from a pre-trained network can be fed into a new network like soft-max classifier. There are also two reasons for this strategy: First, the first several layers of CNNs always can extract the general and useful features for basic image segmentation. On the other hand, empirically, we take the pre-trained networks as an initial network and then a group of particular training samples with fine-tuning (thermal data for defect detection). Since all the weights and bias are tuned with particular tasks, the representative feature is adjusted to be more meaningful, the method can save an amount of data for learning than training in general, and it also trained much faster [25].

- **ResNet feature**

Residual Network was developed by K. He et al. [23] who were awarded the best paper at the ILSVRC 2015. In order to address a degradation problem, the authors introduced this learning framework which involves the special skip connections.

Rather than stacking a few layers directly to a target underlying projection, they force these layers to combine with a residual projection. First, defining this target underlying projection as  $H(x)$ , then they stacked plain layers to fit another projection:  $F(x) = H(x) - x$ . The original projection is recast to  $F(x) + x$  and this formulation of  $F(x) + x$  could be realized by a feed-forward neural network with special skip connections as shown in the Figure 1.11.

This building block could be defined as equation (1.8):

$$y = F(x, \{W_i\}) + x \tag{1.8}$$

Here  $x$  and  $y$  are the vectors from the input and output layers. The function  $F(x, \{W_i\})$  represents the residual projection to implement.

In order to match the dimensions, we add a linear projection  $W_c$  in the shortcut connection,

$$y = F(x, \{W_i\}) + W_c x \tag{1.9}$$

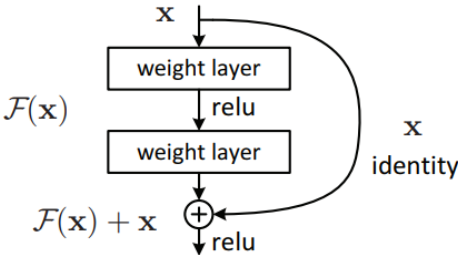


Figure 1. 11 Residual learning: a building block

(He K, Zhang X et.al [23])

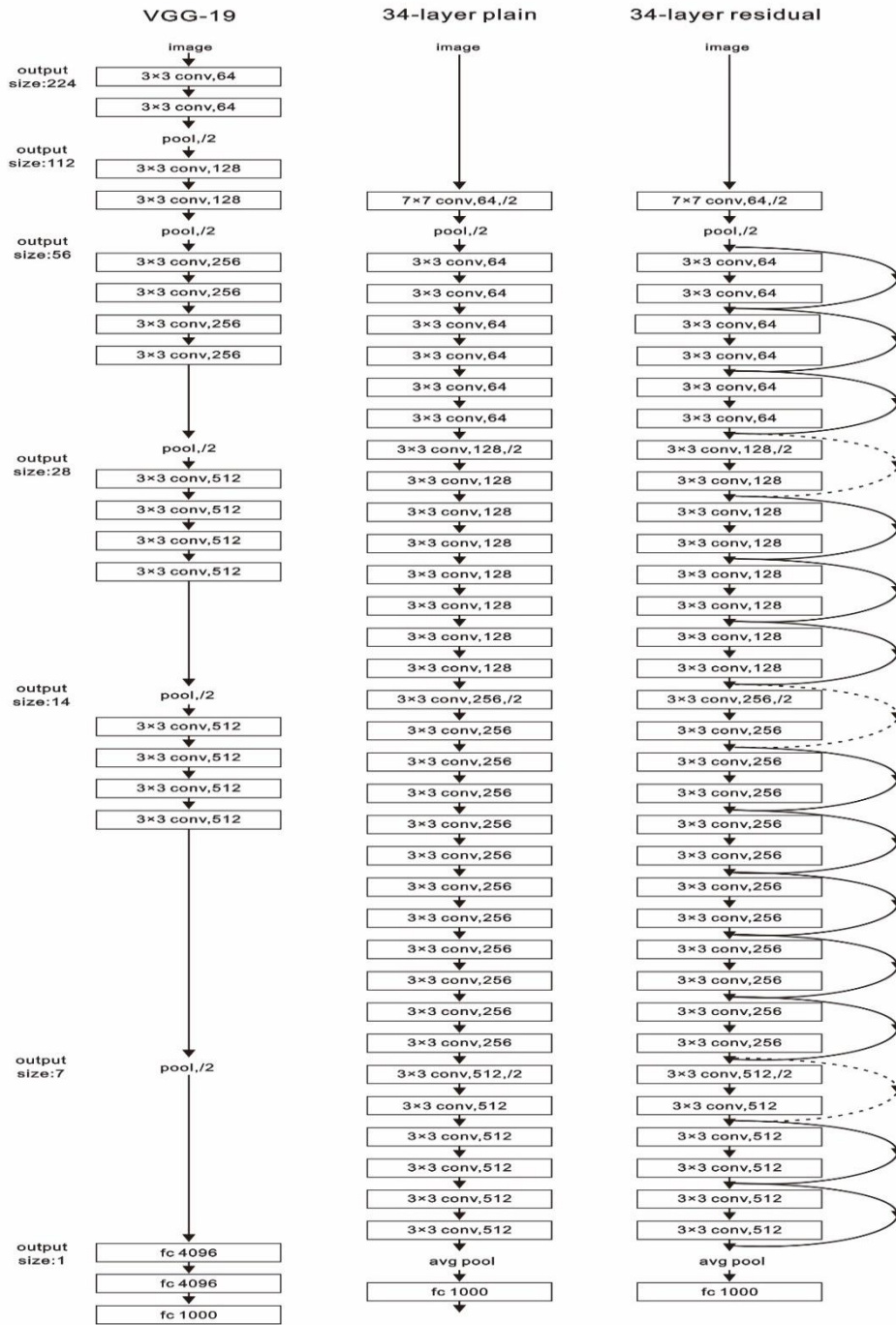


Figure 1. 12 Example network architectures for ImageNet.

(He K, Zhang X [23])

In article [23], the authors test different plain/residual nets to observe the consistent phenomena. As described in Figure.1.12, the plain baselines in the middle are inspired by the VGG nets [35] on the left. Each convolutional layer is structured with  $3 \times 3$  filters. This down-sampling procedure performed with convolutional layers with stride 2 and ends with an average pooling layer globally and a soft-max classifier with a certain amount of output. In Figure 1.12 at the right side, based on the plain network, the authors attached shortcut connects to reshape the network in the middle into residual version. The dotted lines are introduced in case of the increasing dimension. In Figure.1.12 below, two possible choices have been provided to solve this issue: (1) Using zero padding for increasing dimensions in shortcut performs identity mapping. (2) A shortcut path  $W_c$  in Equation. (1.9) is set up to be suitable with the dimensions. For these two choices, the convolutional net is performed with a stride of 2.

### 1.7.10 Region Proposal Networks (RPN)

During the process of the region proposal network (RPN) process, a group of rectangular object regions are outputted with corresponding score. A convolutional architecture network achieves this process, then mingles with the Fast R-CNN network.

In order to generate the feature proposals, a tiny network is over the entire final feature map output from the last convolutional layer in the convolutional architecture network to slide a spatial window, then generate a tiny feature map as input. All of the features have been extracted and are then put into another two other types of layers: a layer for regression (reg); a layer for classification (cls).

- **Anchors**

At each sliding window location, a set of  $K$ -object proposals is defined. We suppose that the number of maximum possible feature maps is denoted as  $k$  in every single location. As a result,  $4k$  outputs and  $2k$  scores encode as the coordinates and probability for each feature map respectively. Each feature map is name as an “anchor”. We use the scale and aspect ratio parameters to adjust the total number of “anchors”. Anchors improve the handling of objects of different sizes and aspect ratio. For example, for a feature map of size  $W \times H$ , there are  $WHk$  anchors as illustrated in Figure.1.13 below [26].



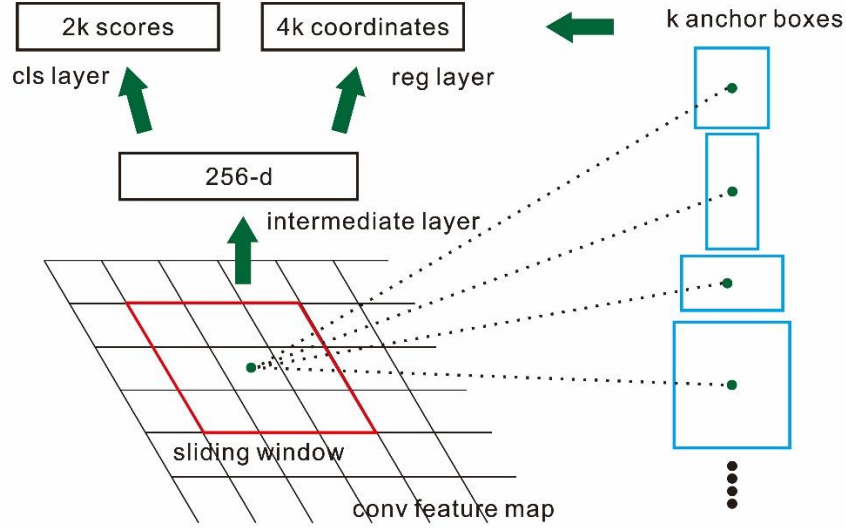


Figure 1. 13 Region Proposal Network

(Ren S, He K, Girshick R, et al: [26])

- **Training of RPN**

Object probability is with reference to anchors, eg:

1. Anchors is regarded as positive samples if  $IOU > 0.7$  (regular value) or  $IOU$  is max.
2. Anchors is regarded as negative samples if  $IOU < 0.3$  (regular value).

Lost function:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \gamma \frac{1}{N_{reg}} \sum_i P_i^* L_{reg}(t_i, t_i^*) \quad (1.10)$$

$$t_x = (x - x_a) / \omega_a, \quad t_y = (y - y_a) / h_a, \quad (1.11)$$

$$t_w = \log(\omega / \omega_a), \quad t_h = \log(h / h_a), \quad (1.12)$$

$$t_x^* = (x^* - x_a) / \omega_a, \quad t_y^* = (y^* - y_a) / h_a, \quad (1.13)$$

$$t_w^* = \log(\omega^* / \omega_a), \quad t_h^* = \log(h^* / h_a), \quad (1.14)$$

Here we illustrated those parameters in equation. (1.10) and equation. (1.11) – (1.14) from the original paper [37] as followed:

$i$ : index of an anchor in a min-batch.

$P_i$  : Probability of anchor being an object

$P_i^*$  : A ground truth label for 1 if it is positive or 0 if it is negative.

$t_i$ : A vector representing the coordinate's parameters of the predicted bounding box.

$t_x^*$ : The ground truth box associated with a positive anchor

$L_{cls}$ : The classification loss is the loss over two classes (object or not object)

$L_{reg}$ : The regression loss -robust loss.

$x, y, w, h$ : the box's center coordinates and its width and height.

$N_{cls}, N_{reg}$ : Normalization terms.

- **Fast R-CNN**

Fast R-CNN uses a technique such as ROI-Pool (Region of interest Pooling) to reduce the computation time spent to identify region proposals in each image as used by the R-CNN algorithm. Fast R-CNN takes as input a series of images and a list of R object proposals which for the purposes of this project, are the thermal defects contained in the subsurface of our samples. The second insight of this approach is to jointly train the CNN, classifier, the bounding box regression in a single model with a soft-max layer above the CNN to output the classification of the defect.

### 1.7.11 Generative adversarial network basic theory

Generative Adversarial Network (GAN) is proposed by (Goodfellow et al. [27]) to be used as a model to generate simulated data in unsupervised learning. To some extent, the generative data from GAM could be indistinguishable from the real data when the dataset is provided. The relationship between the main networks of GAN can make an analogy like a mini-max game is played between them. During the training first, we sample a vector  $z$  from the prior gaussian distribution  $P(z)$ . We design a generator network  $G$  to learn the feature from the distribution  $P(z)$  through the vector  $z$  randomly. This generate network projects a new output data  $G(z)$ . Another network called the discriminator  $D$  which has the responsibility to distinguish the true data (from training distribution) and fake input (from the generator)"  $D$ 's objective is  $D(x)=1$  for real  $x$  and  $D(x)=0$  for fake  $x$ . While the generator is optimized to create realistic data that fools the discriminator, the discriminator also evolved the ability to precisely distinguish true and fake.

Basic Generative Adversarial Network Algorithm:

Initialize Parameter  $\theta_d$  of Discriminator  $D(x)$  and  $\theta_g$  of Generator  $G(x)$

In each training iteration :

1. Sample  $m$  examples  $\{X^1, X^2, \dots, X^n\}$  from data Distribution  $P_{data}(x)$ ;
2. Sample  $m$  noise samples  $\{Z_1, Z_2, \dots, Z_n\}$  from the prior  $P_{prior}(x)$ ;
3. Obtaining generated data  $\{X'_1, X'_2, \dots, X'_n\}, G(X'_i) = G(Z_i)$ ;
4. Update discriminator parameter  $\theta_d$  to maximize

$$V' = \frac{1}{m} \sum_{i=1}^m \log D(x^i) + \frac{1}{m} \sum_{i=1}^m \log(1 - D(x'_i)) \quad (1.15)$$

$\theta_d \leftarrow \theta_d + \beta \nabla V'(\theta_d)$  (for this procedure to learn  $D(x)$ , we repeat  $K$  times)

Sample another  $m$  noise samples  $\{Z_1, Z_2, \dots, Z_m\}$  from the prior  $P_{prior}(z)$

5. Update generator parameter  $\theta_g$  to minimize

$$\frac{1}{m} \sum_{i=1}^m \log(D(G(z^i))) \quad (1.16)$$

$\theta_g \leftarrow \theta_g + \beta \nabla V'(\theta_g)$  (for this procedure to learn  $G(x)$ , we repeat only once)

## 1.8. Evaluation Metrics in Deep learning with TNDE

In this section, the main metrics to quantify and analysis the detection accuracy will be discuss in the following up subsections which include: 1. Confusion matrix, Accuracy, Precision, Recall, F-scores; 2. Receiver operating characteristic (ROC); 3. Probability of detection (POD); 4. mean Average Precision(mAP); 5. Signal to noise ratio (SNR & Peak signal to noise ratio (PSNR)

### 1.8.1 Confusion matrix Accuracy, Precision, Recall, F-scores

A confusion matrix (classifier or diagnosis) is a mapping of instances between certain classes/groups. Specifically, for a binary prediction problem, the results can be classified into two types as positive ( $p$ ) or negative ( $n$ ). Then four results could be obtained into in a binary classification. Therefore, given an instance, if the ground truth is positive and it is predicted as positive, it represents as a true positive ( $TP$ ); if it is predicted as negative, it represents as a false negative ( $FN$ ). If the ground truth is negative and it is predicted as negative, it represents as a true negative ( $TN$ ); if it is classified as positive, it represents as a false positive ( $FP$ ). When the certain number of samples and instance being calculated, the two-by-two confusion matrix could be illustrated a disposition from the samples as shown in the below:

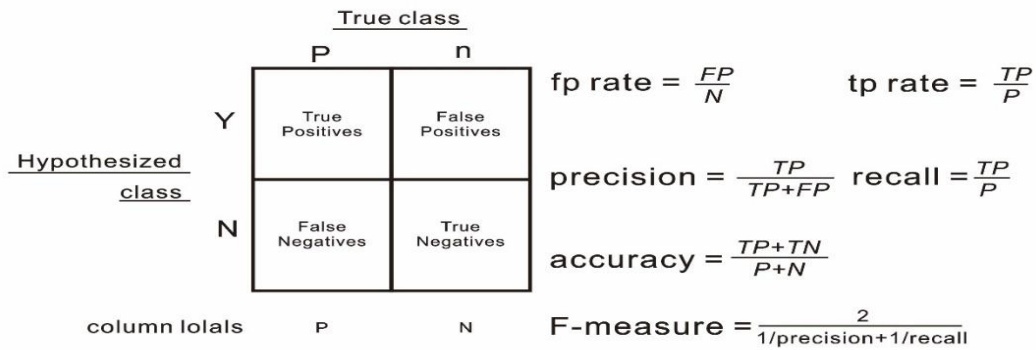


Figure 1. 14 Confusion matrix

The parameters' rates in Figure.1.14 are calculated as following:

The **true positive rate** (also called *recall*):

$$tp\ rate = \frac{\Sigma\ True\ positives}{\Sigma\ Total\ positives\ (or\ Condition\ positive)} \quad (1.17)$$

The **false positive rate** (also called *false alarm rate*):

$$fp\ rate = \frac{\Sigma\ False\ positives}{\Sigma\ Total\ negatives\ (or\ Condition\ positive)} \quad (1.18)$$

Additional rates associated are defined as:

The **true negative rate** (also called *specificity*):

$$tn\ rate = \frac{\Sigma\ True\ negatives}{\Sigma\ Total\ negatives\ (or\ Condition\ positive)} \quad (1.19)$$

The **false negative rate**:

$$fn\ rate = \frac{\Sigma\ False\ negatives}{\Sigma\ Total\ positives\ (or\ Condition\ positive)} \quad (1.20)$$

And: sensitivity = recall

Accuracy in the equation (1.21) is one of the evaluation metrics to analysis that the number of samples being correctly classified divided by the number of all samples. In the general, the higher the correct rate, the better model that being obtained. However, although accuracy

is a meaningful evaluation metric to represent good model, a high accuracy rate does not directly represent that an algorithm is compatible.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (1.21)$$

The precision rate equation (1.22) which indicates how many of the samples predicted to be positive are really positive samples for the predicted results. Furtherly there are two possibilities options to classify positive, one is to predict positive class as positive class ( $TP$ ), and the other is to predict negative class as positive class ( $FP$ ), which indicated equation (1.22),

$$Precision = \frac{TP}{TP+FP} \quad (1.22)$$

Recall indicates equation (1.23) how many positive cases in the sample were predicted correctly for the raw samples, correspondingly, there are also two possibilities for the recall metrics, one is to classify the original positive class as positive ( $TP$ ), and the other one is to classify the original positive class as negative ( $FN$ ).

$$Recall = \frac{TP}{TP+FN} \quad (1.23)$$

As a result, P(Precision) and R (Recall metrics has some contradiction to some extent). It is necessary to consider them together and to focus one of metric more than another in the projects. In this project, we focus more on the recall metrics and introduced the F-Measure (F-score), by calculating the F-value equation (1.24) to evaluate the results and analysis.

$$F \text{ score} = (\beta^2 + 1) \frac{Precision \times Recall}{(\beta^2 \times Precision) + Recall} \quad (1.24)$$

### 1.8.2 ROC concept

The Receiver operating characteristic (ROC) is a type of conceptions which helps to visualize, organize and compare classifiers derived from model performance. ROC metrics have been applied in different fields as signal detection (Swets et al. [28]), diagnostic systems (Swets [29]), medical decision (Zou [30]), and the earliest one in machine Learning (Spackman [31]). Although it has been widely chosen as a standard method in different engineering fields, ROC is still quite rarely applied in the thermographic field. (Bison et al. [32]).

The ROC curve, also called the subject operating characteristic curve, is based on the prediction results of machine learning to sort the samples into positive classes, and then

calculate the true case rate and false case rate. The ROC metric is drawn in the graph where the vertical axis defined as false positive rate and the horizontal axis defined as the true case rate. In practice, only a limited number of test data samples can be used to draw the ROC curve, so the ROC curve presents an unsmooth line, and the area under the curve is the AUC value, the larger the AUC value is, the better the classification effect is.

**1.8.3 Mean Average Precision (mAP)**

The precision and recall curve (P-R curve) are basically the way of visualizing the way of the deep learning models performing while the confidence threshold is decreasing that your model is making prediction that.

mAP (mean average precision) [33] is calculated as the mean value of average precisions. Mean average precision is a measure of recognition accuracy in target detection and an indicator to measure the detection accuracy in object detection. In multiple category target detection, a curve can be plotted for each category based on recall and precision. mAP is the area under the curve, and mAP means averaging the AP for each category again.

In the Figure 1.15 in the below, these are the drawing mAP curve which represent visual depiction but also has an illustration of the sense of the calculations of underneath for each average Precision. Each color of curve is corresponding to a specific class (red; green; blue; organs).

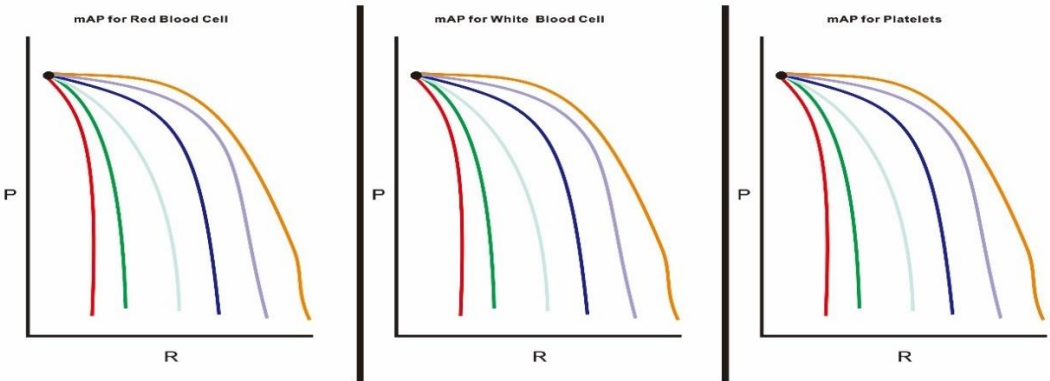


Figure 1. 15 The definition of mean average precision

**1.8.4 Probability of detection (POD)**

POD curves are the generally accepted way to quantify of nondestructive inspection reliability [34]. It is usually expressed as a function of the flaw size, but some time as a

function of other parameters metric (such as ratio = defect size/defect depth). POD curves are usually estimated from data taken from inspections.

Can we find a crack or defect of size a reliably?

We would like to find a defect at least 90% probability of detection with 95% confidence, which is a traditional-defined inspection approach to indicate a good trade-off what we can achieve in specific method. In NDT inspection, a confidence interval score needs to be defined as a statistical method to evaluate the POD metric for the practical testing. Therefore, the practical principle (90/95) in inspection industry represent the inspector could find the defect(flaws) with 90% chance and has 95% confidence to make sure about it. In modern inspector principle, 29/ 29 minimum rules had been introduced as the standard of the saturation of the 95% confidence score (29 flaws can be detected in 29 flaws totally) which is high burden to the inspector and need to assume the POD increases with the increasing flaw size.

In additional, the POD metrics are adapted to evaluate the minimum flaw size or ratio (defect size/defect depth) which being acceptable and reliably from NDT method. A good way is to illustrate the probability or percentage of detects that being detection against the minimum size of defects that being detected and introduced a threshold. In the ideal situation, all the defects that over than some key size that are going to be detected otherwise will not being detected (smaller than that). In Figure 1.16, it has introduced a POD curve for the description, which indicated that the defect that higher than 2.2 mm corresponding to 80% probability being detected, furtherly we have 90% confidence to make sure that the defects are not bigger than 3.3mm. In contrast, we can also state with 90% confidence that a defect at 2.2mm high could be detected over than a POD of 65%.

In regular, there is two way to analysis the inspection results. Based on the quantitative stimulus responses, the inspection results can be recorded as  $\hat{a}$ . In another way, the inspection results can be also classified as found/ non-found of the flaws (1 indicates a flaw was found; 0 indicates a flaw was no found), which is being called as binary data or hit/miss data. Therefore, the POD function could be formulated in two different formats either by quantitative response data ( $\hat{a}$  vs. a data) or the hit/miss data. The size of defect (or other variables such as the ratio (*size/depth*) need to be uniformly spaced on a suitable range on a Cartesian scale, which is a common practice in recommended [35]. It has to realized that if a large defect is always being found (or being saturating in the recording device) or a small defect is always being missed (or the noise obscuring the signal in the system), which means the only limited information from the POD function.

In this project, we will define a confidence threshold scores (CTS) [36] as a standard for measuring the confidence and reliability of detection as well in DL with POD function of the

NDT inspection, which adapted as the criterion is used to measure the correlation between ground true labeling and predicted detection results; the higher the correlation and accuracy, the higher the confidence threshold score value.

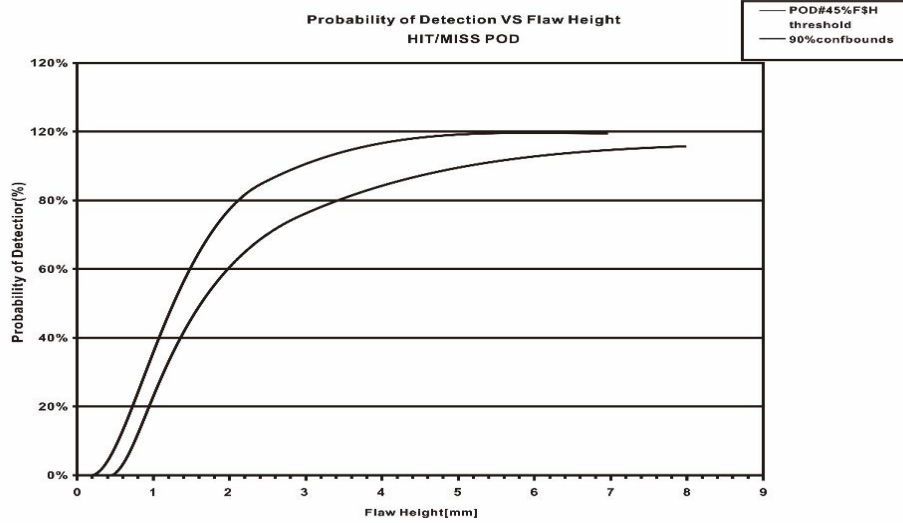


Figure 1. 16 sample POD curve

### 1.8.5 Signal-to-noise (SNR) and Peak signal-to-noise ratio analysis (PSNR)

The signal-to-noise ratio (SNR) parameter [37] could be adopted to evaluate the performance of the proposed processing DL models for defect visibility enhancement. It could quantify the relationship between the signal from the desired and the level of thermal noise when it at the certain signal contrast. It also effectively reflects the thermal contrast and defect feature based on the inequality between the defect regions and non-defect regions in infrared thermography. The higher SNR values indicate that the image contains more defect information and indicate a better implementation of the learning model for the defect visibility enhancement. In equation (1.25),  $M_{def}$  represents the average pixel of the defect' regions and the  $M_{in}$  represents the average pixel of the boundary non-defect regions.  $\sigma_{in}$  means the standard deviation of the pixel value in the non-defect regions.

$$SNR = \frac{M_{def} - M_{in}}{\sigma_{in}} \quad (1.25)$$

In order to discuss the SNR at maximum signal contrast, Peak Signal-to-Noise Ratio (PSNR) is also defined to objective measurement to assess picture quality, and the unit of PSNR is db. In this project, we define the results for the SNR at maximum signal contrast as Peak signal to noise ratio. In order to discuss the SNR at maximum signal contrast, the PSNR is



also defined through the mean square error (MSE) (equation (1.26)), assuming that the image has the size of  $(m \times n)$  from the denoising image  $I$  and the noise image  $k$  (the coordinate value in each image is noted by  $i, j$  in these images).

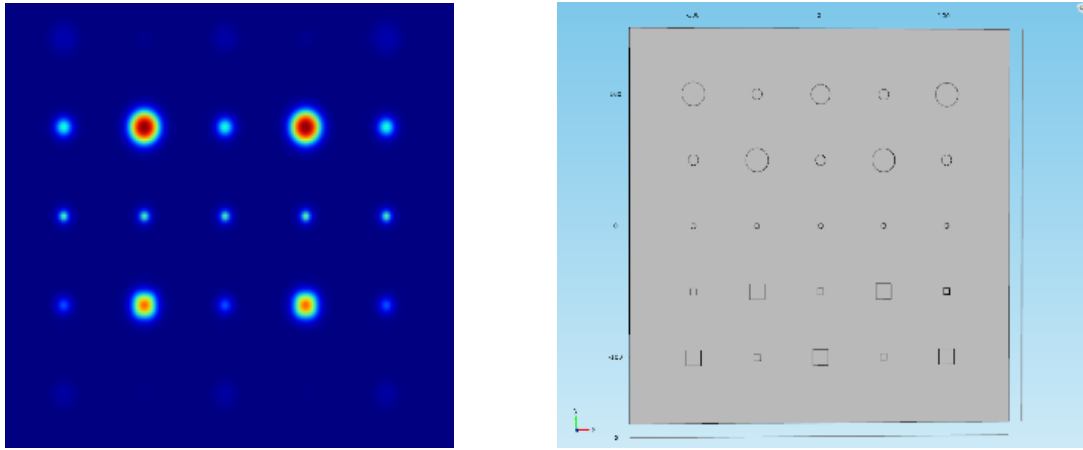
$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - k(i, j)]^2 ;$$

The PSNR (in dB) is defined as:

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) \quad (1.26);$$

The  $MAX_I$  is the maximum possible pixel value of the image (this is 255 when pixels using 8 bits represent the sample).

### 1.9 Modelling and stimulation



(a) Simulated thermogram at  $t = 106.5$  s;

(b) corresponding geometric distribution

Figure 1. 17 Stimulation data

In this thesis, all modeling and simulation work are undertaken in two different kind of stimulation platform: Finite element model (FEM) (COMSOL Multiphysics®) and Thermal ®, which might be helpful when comparing with the experimental results.

COMSOL Multiphysics is a stimulation software platform, based on advanced numerical methods, for modeling and analysis physics-based problems, which represent the finite element analysis, simulation software package applied in Nondestructive evaluation applications. In the Figure 1.17, it is an example that

The advantages of using COMSOL for DL defect detection research analysis modeling (assembles finite-element method matrices):

- clean and automated supervision data;

- rather than other physical software is highly non-linear and has uncertainty issue, it can be abundantly simplified due to the thermal principle;
- specific generate syntenic data in order to featuring and extracting less visible defect regions for training
- has an environment that integrated with modeling

One example of application of COMSOL simulation in infrared thermography for NDT & E can be found in (Cannas et al. [38]). Where a CFRP samples embedded 25 regular defects was heated by two halogen lamps to detect the defects position. A 3D model by finite element method under COMSOL platform has also been implemented to simulate the heating process. Experimental and numerical data have well matched, which allows parametric studies free from the experimental tests. In the Figure 1.18, they are the (a) colorful simulated thermograph at  $t=106.5s$ ; (b) its correspondents the geometric distribution of the sample.

### 1.10 The recent research progress in literature with deep learning in IRT-NDE

Similarly, with other non-destructive evaluation methods, the main objective of active infrared thermography being employed to analysis and evaluate the discontinuity or abnormally regions in the different industrialized fields or regions such as the materials in the composites from aerospace industry. Currently, the application and literature of DL in active thermography is fewer than that applied in passive thermography. Since the active thermography NDT method have several main application fields, we divided Deep learning for defect detections projects in active thermography NDT into three categories (Quality control; Structural health monitoring; Materials) based on the industrial application. In the Figure 1.18, two representative situation that could be applied with IRT and deep learning techniques (a) Transportation inspection; (b) aerospace materials inspection [39].

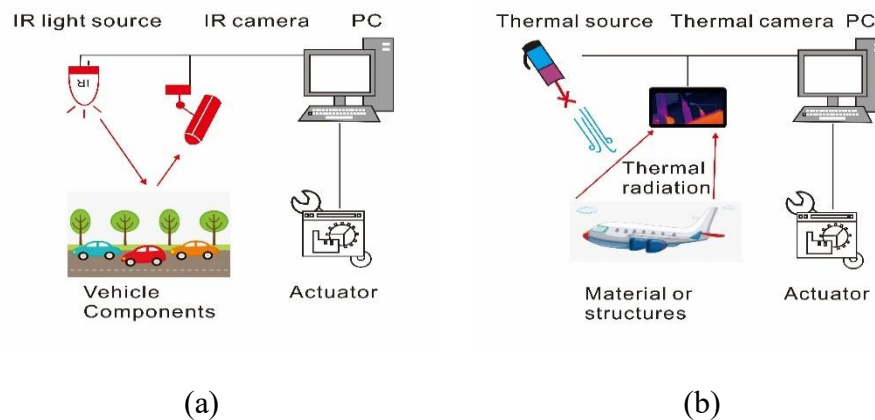


Figure 1. 18 Thermal infrared machine vision(a)(b)

- **Quality control**

In article [40], the merge of Infrared thermography with DL model was employed for the quality inspection of the adulterations from honey. During the procedure of cooling, the pure honey samples have been mixed with the rice syrup at the different level of concentrations which could be possible to being inspected from Infrared camera, then by analyzed by the CNN modeling. The whole procedure was trained and validated by the samples from the thermograms that captured from the experiments. The trained model has the capability to detect and identify the original floral honey with the adulterer honey with highly accuracy 95% and 93%, respectively. Similarly, when the method being employed to the of extra virgin olive oil with adulteration, it could reach the accuracy of 97% as well. The same situation is being applying for detect and classify of bruises of pears from the proposed DL model which research the accuracy of 99.25% [41].

- **Structural health monitoring**

In article [42], a modified deep neural network being proposed to detect damage from subsurface from steel samples of truss bridge through IRT. With a FLIR camera (LWIR T650SC), 35 thermal sequences were collected and converted to be 2000 images around with the resolution of  $299 \times 299$  pixels (200 thermal images with higher resolution of  $640 \times 480$  pixels). In order to save the computation expense, the state of art of deep inception neural network (DINN) is being modified in order to achieve the transfer learning strategy. The damage from the coating and steel surface are being detected, then visualized and localized with the bounding boxes through the proposed modified DI model. 200 tested thermal images are being evaluated and obtain the highly robustness and accuracy 96%. A combination from IRT and CNN model being adapted to the diagnosis of rotating machinery diagnosis fault in article [43]. A CNN (LeNet-like) is being proposed to identify the faults in the system and classify the health condition. Two different databases from experiment indicated that proposed algorithm has more advanced performance of detection for various faults of the components from rotor and bearings in comparison with other the state of deep learning models and regular vibration-based method.

- **Materials**

Inverse issue is a challenging topic to find the solution although it has ubiquitous heat conduction. In Ref [44], a data-driven semantic segmentation-U-net for heterogeneous composites for is proposed which could be adapted to distinguish the thermal images into defect and non-defect regions based on the pixel level wisely. The 1200 synthetic thermal databases for the temperature regions have been generated through simulation. The temperature distribution of fillers has been successfully predicted at the accuracy of 0.979

from the trained model when the true temperature from each pixel being given which indicated the capability of the semantic segmentation from deep learning model-U-net to undertake the inverse issues with a dimension transformation method.

### 1.11 The representative DL models in literature with deep learning in TNDE

Several classic DL models (Res-Net; Autoencoder; Long Short-Term Memory (LSTM); Generative Adversarial Networks (GAN)) have the meaningful potential to apply for the Active infrared thermography for the analysis and defects diagnosis, which will be discussed and illustrated in the several sessions in the following:

- **Applying CNN in INDE**

In paper [45], a proposed approach by B. Yousefi, et al. for these issues in infrared thermography is to use a Pre-trained CNN (ImageNet-vgg-f-13) to be used as a generic feature extractor and combined with the spectral analysis to learn specific features of the defects through the vector from the output. This methodology can be described as follows:

An infrared image 3D matrix is used in different time series as the raw input and is compared with the spectral analysis. Suppose that  $s$  is one frame of infrared image 3D matrix such that  $s \in R^{n \times m \times k}$ , where  $k$  represents the time sequences.  $S \in \{ \mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_K \}$ , and  $n$  and  $m$  are the infrared image's spatial resolution, and  $w$  is a thermal reflectance from  $s$ ,  $w \in \{ \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K \}$ ,  $\varphi$  is a re-scale squared 2D matrix reshaped by  $w$  as the input to a deep pre-trained convolutional network. To fully utilize this image  $\varphi$ , a color image  $\varphi_{rbc}$  concatenated by  $\varphi$  in three times is used to complete the application in CNN [45].

Using a pre-trained network could efficiently handle the learning issue of insufficient training data. During the forward propagation of training in a neural network, an input- output procedure could be illustrated as a function as follow:

$$f(\varphi) = f_n(\dots f_2(f_1(\varphi; \omega_1); \omega_2), \omega_n) \quad f: R^{A \times B \times C} \rightarrow R^{A' \times B' \times C'} \quad \varphi \rightarrow \gamma \quad (1.27)$$

$\varphi_i$  is a three-dimensional matrix  $\mathbf{A} \times \mathbf{B} \times \mathbf{C}$  where  $\mathbf{A} \times \mathbf{B}$  represents the size of the images and  $\mathbf{C}$  represent time frame. In fact,  $f_n$  is defined as a set of functions having convolutional structure applying to the input  $\varphi$ . Due to these convolutional structures involving a series of filters inner product, the output follows the input dimensional property also.

In Spectral similarity measurement, the Spectral Angle Mapper (SAM) is an algorithm having a vector in  $n$ -dimensional geometrical space. It calculates the un- similarity between the spectrum type vector  $FC7$  and the reference spectrum vector  $Ref$  using the equation below [45]:

$$\alpha = \cos^{-1} \frac{\sum_{i=1}^n FC_{34i} Ref}{(\sum_{i=1}^n FC_{34i}^2)^{\frac{1}{2}} (\sum_{i=1}^n Ref_i^2)^{\frac{1}{2}}} \quad (1.28)$$

where  $\mathbf{n}$  is a high dimension feature size vector (depending on the neural network architecture) and represents the defectiveness score from each infrared image as an input to the designed system.

However, due to the fact that deeper neural networks are more difficult to train, a degradation problem could be possible as shown in Figure 1.19 when VGG-net are able to start converging. With the layers of network increasing, accuracy becomes saturated and declines quickly. Unexpectedly, this degradation problem is not due to the overfitting, but is caused by the problem of gradients vanishing and exploding, as seen in Figure.1.19 [23].

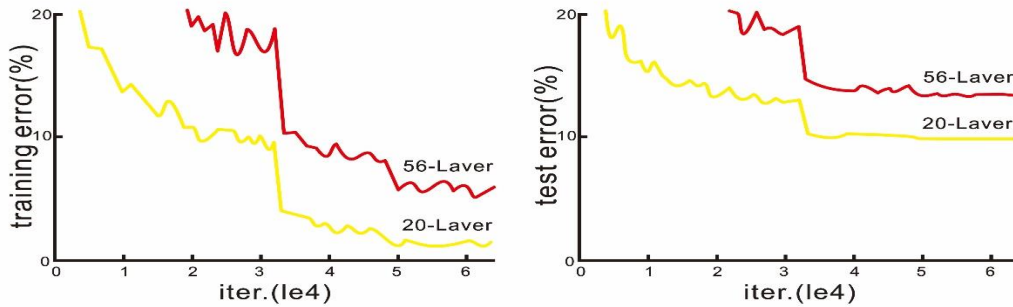


Figure 1. 19 Training error (left) and test error (right) on CIFAR-10

With 20-layer and 56-layer “plain” networks. The deeper network has higher training error, and thus test error. (He K, Zhang X, Ren S et.al [23])

- **Autoencoder for feature extraction**

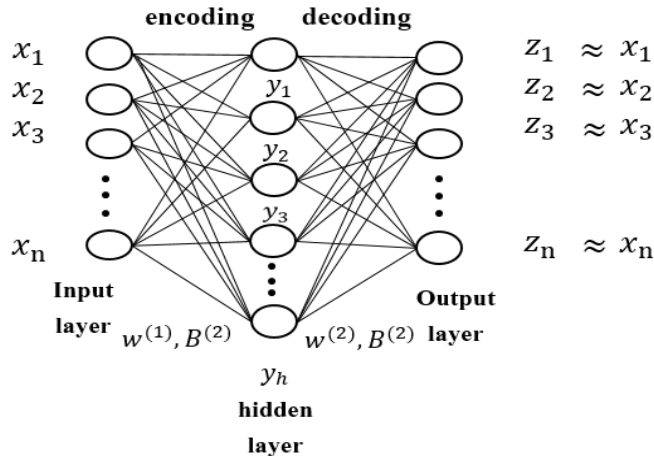


Figure 1. 20 Basic Structure of an Auto-encoder (AE) model

Auto-encoder (AE), is a feed-forward neural network which reproduces its input with the output layer in order to automatically extract meaningful features from the dataset and leverage the availability of unlabeled data.

The simplest architecture of the AE model is elaborated in Figure.1.20 which includes three layers, corresponding to the input, the hidden, and the output layer. There are two work periods called encoding and decoding during training. During the encoding period, the input data  $x$  is projected to the hidden layer vector  $y$ . Then, the decoding period, the hidden layer vector  $y$  is projected to a reconstruction data  $z$  in the output lay. During training, the CAE model forces  $z$  to be as similar as possible with the input data.

During the recent years, due to its wide application, the auto-encoder has been a powerful research tool as a building block in this research field. During previous study, an innovative approach termed as Auto-encoder thermography (AE) has been proposed to improve Pulse thermography for defect visibility enhancement in article [46]. However, the autoencoder model can only reflect variations observed in the training set which is difficult to keep the good feature and information to be invariant to the other variations.

- **The LSTM network with infrared data**

The infrared data is a type of temporal sequence data. As the 3D matrix, infrared data could reveal that temperature evolution through the time basing on a whole 2D spatial image. The recurrent neural networks have the ability to repeatedly learn and remember information through the time. This networks with the loops of recurrent units [48] could allow the information persist. In order to solve the “gradient vanishing” issue, a long-term memory (LSTM) recurrent unit has been designed [49,50].

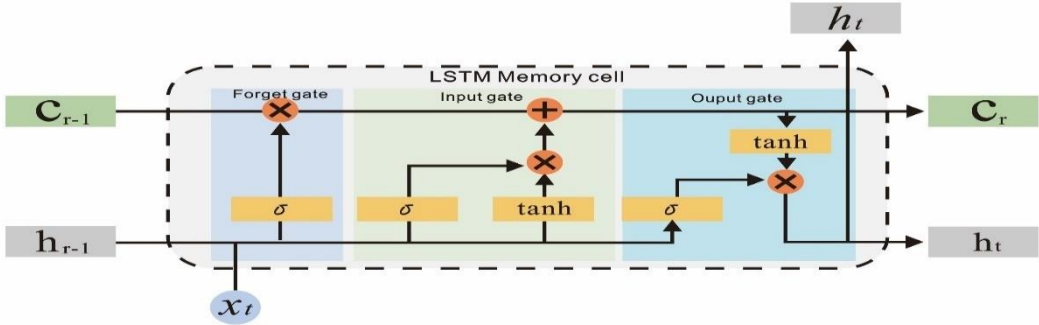


Figure 1. 21 LSTM Block

(Gao, Yang, et al: [47])

In article [50], the author introduced the spatial input from infrared data to be compatible explicitly with the temporal characteristic model multiple LSTM to segment the thermal images for defect analysis based on temporal behavior from the thermal sequences. The LSTM model is employed to handling the prediction of the transient characteristics. Unfortunately, based on the results from the article [50], the regular LSTM model still quite sensitive to the noises from the thermal sequences and its particular irregular shape of defects while partially defects cannot fully and automatic segmented. The LSTM formula is shown in Equation (1.29) - (1.33) and Figure.1.21. Notice that  $\Psi$  is a convolutional operator in Equation. (1.29) - (1.33), and the input data  $x_t$  be given at the current time point  $t$  in Equation. (1.29) - (1.33). Hidden state  $h_{t-1}$  and cell state  $C_{t-1}$  are input state at previous time point  $t - 1$  respectively.  $i_t, f_t$ , and  $o_t$  are the input, forget, output gates,  $c_t$  and  $h_t$  are the new cell and hidden states at time  $t$ .  $\sigma(x) = \frac{1}{1+e^{-x}}$ ,  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  and  $\circ$  are element-wise multiplication.

$$i_t = \delta(\Psi_{xi}(x_t) + \Psi_{hi}(h_{t-1}) + b_i) \quad (1.29)$$

$$f_t = \delta(\Psi_{xf}(x_t) + \Psi_{hf}(h_{t-1}) + b_f) \quad (1.30)$$

$$o_t = \delta(\Psi_{xo}(x_t) + \Psi_{ho}(h_{t-1}) + b_o) \quad (1.31)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(\Psi_{xc}(x_t) + \Psi_{hc}(h_{t-1}) + b_c) \quad (1.32)$$

$$h_t = o_t \circ \tanh(c_t) \quad (1.33)$$

- **Generative Adversarial Networks (GAN) with TNDE**

As shown in Figure.1.22, the Generative Adversarial Networks (GAN) are an innovative developed generative model that can use two types of networks (generative and discriminative). Particularly, for this project, we decomposed the thermal sequence data into  $N$  frame images as the real data. Using  $N$  samples random Gaussian noise was fed through the generator to generate  $N$  false images. Utilizing these  $2N$  images jointly as the input of discriminative network and train the discriminator to classify them as real or generated (As output, we labeling the thermal data as 1 and labelled the generated images as 0. In the second step, we fixed the parameters of the discriminator, the errors are back-propagated through both of two networks simultaneously. Therefore, the generator initially changes the parameters of network from learning to fool the discriminator and to generate more realistic images. In this way, these two networks have been optimized independently so that the generator improve the ability to generate approximately real images and discriminator is more sensible to identify the real images from fake. This way, the discriminator forces the generator to learn the feature further and to improve the generative data further until the

discriminator was fooled by considering this new data as real. The final results of these generated image can be used as training data for defect detection through a deep learning algorithm as the next step.

In article [51], it has introduced a modified GAN network as a joint loss network to achieve the semantic segmentation on the inner deboned defects from the carbon fiber reinforced polymer which enables a stable training performance and improved detection rate. However, the noise interference and non-uniform heating still being a crucial factor that influences the GANs network to segment the irregular shape of defect from the samples.

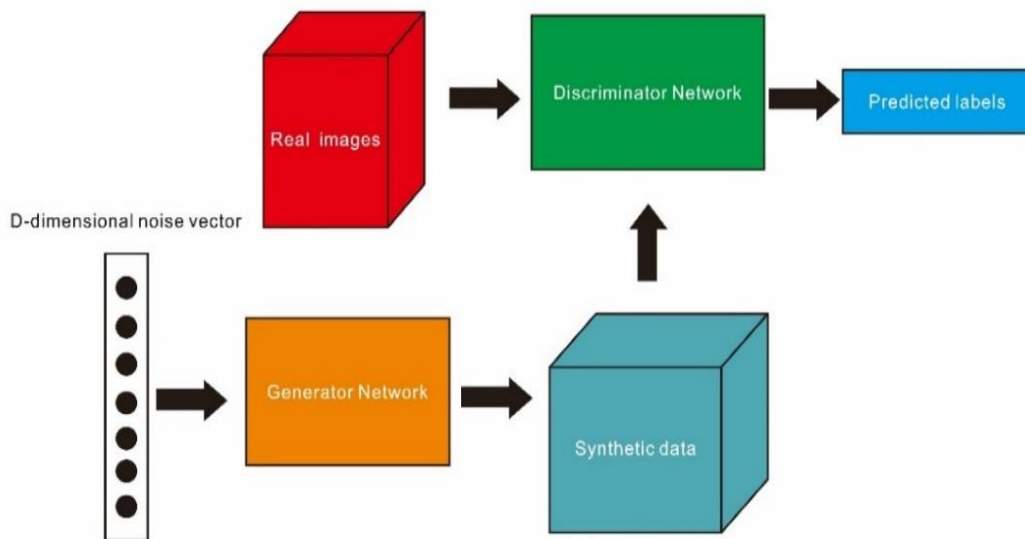


Figure 1. 22 GAN Images generation Model

- **Neural network for defect depth estimation (Defect characterization)**

In infrared thermography, it has been defined that the feature extraction like the depth of subsurface depth, an inverse problem. This problem also has been studied for many years in TNDE. The procedure for quantitative analysis of relevant parameters in defect characterization is also a complex heat transfer issue [52].

In article [53], it used a single neural network to detect defect depth and the scheme of this algorithm is shown in Fig.18. First, it needs to record a complete temporal thermal file from the investigated specimens. This sequence of thermograms is transformed into training feature maps through the process referred to unfolding. Next, after these feature maps have



formed, we labelled a group of temperature values with corresponding pixel positions in feature maps during every discrete time and these values as well as the values of all pixel's position in the map also corresponding to  $N$  columns totally (see: Figure.1.23). It is noteworthy that these temperature values are not directly fed into the regression Neural network. They are processed by Principal Component Analysis (PCA) in Figure 1.23. PCA is a data preprocessing tool to extract the meaningful feature and to reduce the dimension of raw data.

In this study, the regression network own  $M$  inputs neurons which is equal to the number of neurons in the middle layer (i.e. The dimension of the code vector in the middle layer) in PCA. inputs neurons also depend on the number of time frames  $N$  in the raw data we acquired. The quantity of vectors in a regression neural network is equal to the number pixels of sequence of thermograms in each frame. This regression neural network has a group of hidden layers and a single output. During the training, we labelled the known depth value with the given pixel. During the testing, we got the prediction values of pixel depth in the output of neural network This multi-layer regression neural network trained with the backpropagation to compute the difference with target values (the known depth and prediction) [54].

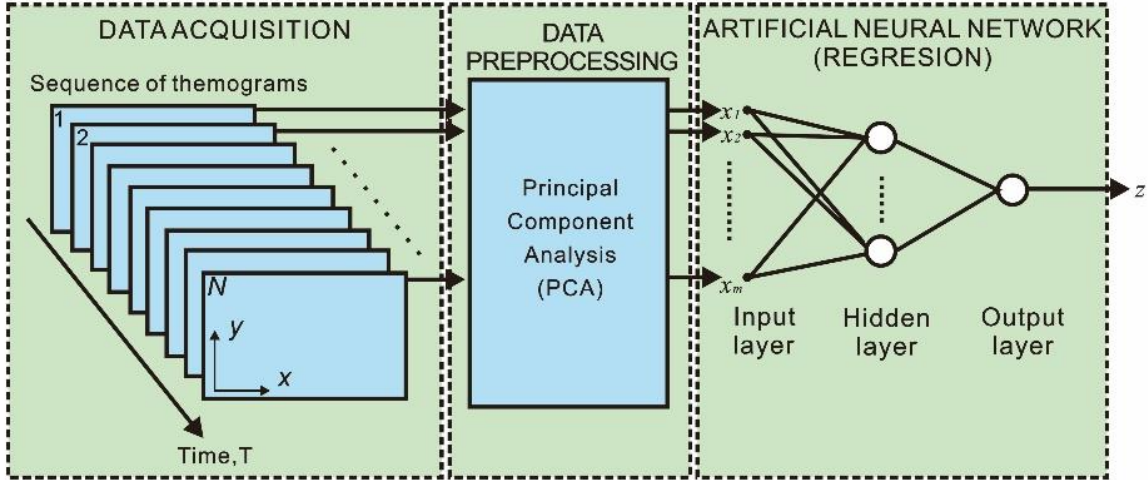


Figure 1. 23 Scheme of the neural algorithm for defect depth estimation

(Dudzik, et al [53])

During the training, all images are divided into ten subsets equally [53]. Then, ten training sessions were carried out. During one training procedure, it randomly sampled nine subsets

for training, the tenth subset for testing using the network as trained. Finally, a field of defects depth estimation has formed. The mean error of depth estimation in the defected area, can be described as the sum of the squared error loss function. (1.34)

$$L(x, y) = \frac{\sum_{i=1}^m \|y^{(i)} - x^{(i)}\|^2}{2m} + \gamma \frac{\sum w^2}{2} \quad (1.34)$$

where  $x(i)$  represents the depth of the defect predicted from the output of regression neural network.  $y(i)$  is a known depth of defects to participate training procedure as a target value. (corrected value).  $m$  is the quantity of training samples in databases the second term is the sum of the squares of all the weights scaled in order to prevent the overfitting problem.  $\gamma$  is the weight decay parameter.

However, in this article, the biggest issue from the neural network being applied on defect dept estimation (defect characterization) is the dimensionality correspondent. Especially when the materials that being evaluated is from the low thermal diffusivity materials with stepping heating method. The input number from neural network which is highly depending on the sample instants from the sequence. Although principal components analysis could help it overcome the issue through dimension reduction procedure, the depth training from the neural network caused high time complexity and instability.

### 1.12 The summary

In view of the above-mentioned research studies, the variety of deep learning applications of infrared thermography has been applied or combined with simple and various deep learning models in the form of defect detection and identification which play a significant role to improve the safety and quality of non-destructive evaluation in the previous literature.

Several research studies have adapted the infrared thermography and particularly temporal or spatial deep neural network (long-short term memory (LSTM); singer neural network; autoencoder; GANs) to extract the feature and analysis through feature classification and identification from deep learning based on the thermal sequences that captured from the infrared camera and experimental pipeline (Ruan L et al., 2020; [19] Duan Y et al., 2019; Zeng X et al., 2020; Yongbo L I et al., 2020; Wu H, 2020 et al.; Ali R et al., 2019; Marani R et al., 2016; Darabi A et al., 2002).

However, all the proposed applications of this research have some limitation and disadvantages that could still to be improved. Therefore, the furtherly points of the identification and classification of defects in the infrared thermography could be described in the following:

1. introduce the synthetic data from Finite element modeling or the GANs network training which could be not only provide a great benefit for improving the accuracy but also saving the highly expense from real experimental samples (CFRP; GFRP);
2. a combination of data processing method with deep learning model for feature extraction could be an efficient way for defect enhancement and identification to improve the visibility of defects in pulse thermography;
3. a modified deep learning neural network to achieved the automatic identification of defects from the training could be innovative method to replace the human inspectors to avoid the human fatigue and also represent an efficiently way of computer-aided identification which can be used to assist non-destructive evaluation and infrared experts in the field or laboratory conditions.

Therefore, in this project, we are going to propose and illustrate the several deep learning modeling's and methods to achieve the objectives and improved the feasibility of itself in order to compensate and disadvantage that previous academic literature that have.

## References

- [1] *Ibarra-Castanedo C, Maldague X P V. Infrared Thermography[M]// Handbook of Technical Diagnostics. Springer Berlin Heidelberg, 2013:826-826.*
- [2] *Maldague X P V. Theory and Practice of Infrared Technology for Nondestructive Testing[J]. Chapter, 2001, 4(3):307.*
- [3] *Martin, R.E., Gyekenyesi, A.L., Shepard, S.M.: Interpreting the results of pulsed thermography data. Mater. Eval. 61(5), 611–616 (2003)*
- [4] *Mariacristina Pilla, Matthieu Klein, Xavier Maldague, and Antonio Salerno. New absolute contrast for pulsed thermography. In Proc. QIRT, volume 5, 2002.*
- [5] *Bardia Yousefi, Stefano Sfarra, Clemente Ibarra Castanedo, and Xavier PV Maldague. Thermal NDT applying Candid Covariance-Free Incremental Principal Component Thermography (CCIPCT). In SPIE Commercial Scientific Sensing and Imaging, pages 102141I–102141I. International Society for Optics and Photonics, 2017.*
- [6] *Lopez F, Ibarra-Castanedo C, de Paulo Nicolau V, et al. Optimization of pulsed thermography inspection by partial least-squares regression[J]. Ndt & E International, 2014, 66: 128-138.*

- [7] Parker, W.J., Jenkins, R.J., Butler, C.P., Abbott, G.L.: *Flash method of determining thermal diffusivity, heat capacity, and thermal conductivity*. *J Appl. Phys.* 32(9), 1679–1684 (1961)
- [8] Prabhu D R, Winfree W P. *Neural Network Based Processing of Thermal NDE Data for Corrosion Detection*[M]// *Review of Progress in Quantitative Nondestructive Evaluation*. Springer US, 1993:775-782.
- [9] Tretout H, David D, Marin J Y, et al. *An evaluation of artificial neural networks applied to infrared thermography inspection of composite aerospace structures*[C]// *Eview of Progress in Quantitative Nondestructive Evaluation*. 1997:330-330(1).
- [10] Maldague X, Largouët Y, Couturier J P. *A study of defect depth using neural networks in pulsed phase thermography: modelling, noise, experiments*[J]. *Revue Générale De Thermique*, 1998, 37(8):704-717.
- [11] Vallerand S, Maldague X. *Defect characterization in pulsed thermography: a statistical method compared with Kohonen and Perceptron neural networks*[J]. *Ndt & E International*, 2000, 33(5):307-315.
- [12] Darabi A, Maldague X. *Neural network-based defect detection and depth estimation in TNDE*[J]. *Ndt & E International*, 2002, 35(3):165-175.
- [13] Benítez H D, Ibarra-Castanedo C, Bendada A H, et al. *Definition of a new thermal contrast and pulse correction for defect quantification in pulsed thermography*[J]. *Infrared Physics & Technology*, 2008, 51(3):160-167.
- [14] Benítez H D, Loaiza H, Caicedo E, et al. *Defect characterization in infrared non-destructive testing with learning machines*[J]. *Ndt & E International*, 2009, 42(7):630-643.
- [15] Liu J, Tang Q, Xun L, et al. *Research on the quantitative analysis of subsurface defects for non-destructive testing by lock-in thermography*[J]. *Advanced Materials Research*, 2012, 301-303(1):635-640.
- [16] Sheng J H. *Non-visible defect detection in glass using infrared thermography and artificial neural networks*[C]// *Thermosense: Thermal Infrared Applications XXXV*.

*International Society for Optics and Photonics, 2013:87050R.*

- [17] Marani R, Palumbo D, Galietti U, et al. Automatic detection of subsurface defects in composite materials using thermography and unsupervised machine learning[C]// IEEE, International Conference on Intelligent Systems. IEEE, 2016:516-521.
- [18] Halloua H, Elhassnaoui A, Saifi A, et al. Neural networks and genetic algorithms for the evaluation of coatings thicknesses in thermal barriers by infrared thermography data[J]. 2017, 5:997-1004.
- [19] Duan Y, Liu S, Hu C, et al. Automated defect classification in infrared thermography based on a neural network[J]. NDT & E International, 2019, 107: 102147.
- [20] Cunningham P, Cord M, Delany S J. Supervised learning[M]//Machine learning techniques for multimedia. Springer, Berlin, Heidelberg, 2008: 21-49.
- [21] Barlow H B. Unsupervised learning[J]. Neural computation, 1989, 1(3): 295-311.
- [22] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436.
- [23] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.
- [24] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015.
- [25] Lopez M B, Del-Blanco C R, Garcia N. Detecting exercise-induced fatigue using thermal imaging and deep learning[C]// International Conference on Image Processing Theory. 2017:1-6.
- [26] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]// International Conference on Neural Information Processing Systems. MIT Press, 2015:91-99.
- [27] “Deep Learning[M]”. Goodfellow I, Bengio Y, Courville A. The MIT Press, 2016.

- [28] John A Swets, Robyn M Dawes, and John Monahan. *Better decisions through science*. *Scientific American*, 283(4):82–87, 2000.
- [29] John A Swets. *Measuring the accuracy of diagnostic systems*. *Science*, 240(4857):1285–1293, 1988.45
- [30] Kelly H Zou. *Receiver operating characteristic (roc) literature research*. On-line bibliography available from: < <http://splweb.bwh.harvard.edu>, 8000, 2002.
- [31] Kent A Spackman. *Signal detection theory: Valuable tools for evaluating inductive learning*. In *Proceedings of the sixth international workshop on Machine learning*, pages 160–163. Elsevier, 1989.
- [32] P Bison, A Bortolin, G Cadelano, G Ferrarini, L Finesso, F Lopez, X Maldague, Corso Stati Uniti ITC-CNR, and Corso Stati Uniti ISIB-CNR. *Evaluation of frescoes detachments by partial least square thermography*. In *12th International Conference on Quantitative Infrared Thermography, Bordeaux (France)*, pages 07–11, 2014.
- [33] Robertson S. *A new interpretation of average precision[C]*//*Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. 2008: 689-690.
- [34] Pavlović M, Takahashi K, Müller C. *Probability of detection as a function of multiple influencing parameters[J]*. *Insight-Non-Destructive Testing and Condition Monitoring*, 2012, 54(11): 606-611.
- [35] US Air Force Aeronautical Systems Center, *Military Handbook 2009: Non-Destructive Evaluation System Reliability Assessment (Department of Defence Handbook)*.
- [36] Rezatofighi H, Tsoi N, Gwak JY, et al. *Generalized intersection over union: A metric and a loss for bounding box regression[C]*//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019: 658-666.
- [37] Johnson D H. *Signal-to-noise ratio[J]*. *Scholarpedia*, 2006, 1(12): 2088.
- [38] Barbara Cannas, Sara Carcangiu, Giovanna Concu, and Nicoletta Trulli. *Modeling of active infrared thermography for defect detection in concrete structures*. In *Proceedings of the 2012 COMSOL Conference*, 2012.

- [39] He Y, Deng B, Wang H, et al. *Infrared machine vision and infrared thermography with deep learning: a review*[J]. *Infrared Physics & Technology*, 2021: 103754.
- [40] Izquierdo M, Lastra-Mejías M, González-Flores E, et al. *Convolutional decoding of thermographic images to locate and quantify honey adulterations*[J]. *Talanta*, 2020, 209: 120500.
- [41] Zeng X, Miao Y, Ubaid S, et al. *Detection and classification of bruises of pears based on thermal images*[J]. *Postharvest Biology and Technology*, 2020, 161: 111090.
- [42] Ali R, Cha Y J. *Subsurface damage detection of a steel bridge using deep learning and uncooled micro-bolometer*[J]. *Construction and Building Materials*, 2019, 226: 376-387.
- [43] Yongbo L I, Xiaoqiang D U, Fangyi W A N, et al. *Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging*[J]. *Chinese Journal of Aeronautics*, 2020, 33(2): 427-438.
- [44] Wu H, Zhang H, Hu G, et al. *Deep learning-based reconstruction of the structure of heterogeneous composites from their temperature fields*[J]. *AIP Advances*, 2020, 10(4): 045037.
- [45] Yousefi, Bardia, et al. "Application of Deep Learning in Infrared Non-Destructive Testing."
- [46] Rifai S, Vincent P, Muller X, et al. *Contractive Auto-Encoders: Explicit Invariance During Feature Extraction*[C]// ICML. 2011.
- [47] Gao, Yang, et al. "Fully convolutional structured LSTM networks for joint 4D medical image segmentation." *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. IEEE*, 2018.
- [48] Bengio Y, Simard P, Frasconi P. *Learning long-term dependencies with gradient descent is difficult*. [J]. *IEEE Transactions on Neural Networks*, 1994, 5(2):157-166.
- [49] Stollenga M F, Byeon W, Liwicki M, et al. *Parallel Multi-Dimensional LSTM, With Application to Fast Biomedical Volumetric Image Segmentation*[J]. *Computer Science*, 2015.

- [50] Patraucean, Viorica, Handa, Ankur, Cipolla, Roberto. *Spatio-temporal video autoencoder with differentiable memory*[J]. *Computer Science*, 2016, 58(11):2415 - 2422.
- [51] Ruan L, Gao B, Wu S, et al. *DefectNet: Joint loss structured deep adversarial network for thermography defect detecting system*[J]. *Neurocomputing*, 2020, 417: 441-457.
- [52] Darabi A, Maldague X. *Neural network-based defect detection and depth estimation in TNDE*[J]. *Ndt & E International*, 2002, 35(3):165-175.
- [53] Dudzik. *Analysis of the accuracy of a neural algorithm for defect depth;estimation using PCA processing from active thermography data*[J]. *Infrared Physics & Technology*, 2013, 56(36):1-7.
- [54] Sheng J H. *Non-visible defect detection in glass using infrared thermography and artificial neural networks*[C]// *Thermosense: Thermal Infrared Applications XXXV. International Society for Optics and Photonics*, 2013:87050R.



# Chapter 2 Issues, Objectives

## 2.1 Problem Statement

For defect identification and detection, several state-of-the-art regular approaches in infrared thermography especially pulsed thermography are powerful tools which are known to give noticeable results, such as the Pulsed Phase Thermography (PPT) [1], Principal Component Thermography (PCT) [2], Difference of Absolute Contrast (DAC) [3], Thermographic Signal Reconstruction [4], as well as Candid Covariance Free Incremental Principal Component Thermography [5]. These mentioned methods show remarkable efficacy in improving defect visibility during INDE and also have been demonstrated to be powerful tools that are well-documented and provide noticeable results in Infrared Thermography.

However, these methods are pattern-based unsupervised methods [6] which means extracting more distinctly defect information from time and frequency domains, and seeking for new methods to further improve defect visibility are still crucial issues in TNDE literature. Convolutional Neural networks (CNN) have been widely applied by scientists in image processing, and have shown the ability to perform well in data analysis and feature extraction.

Due to the particular characteristics from the spatial deep learning model, the implementation of supervised learning algorithms in infrared non-destructive evaluation (pulsed thermography) is a potential tendency in future research. There are few available works in the literature dealing with automatic defect detection in infrared thermography by using deep learning algorithms [7] [8] [9]. Certain articles have introduced a few single algorithms from DL for defects detection analysis in infrared thermography. A systematic investigation and comparison of deep learning algorithms and innovative deep learning algorithms to further improve its efficacy for the defect detection and analysis in Pulsed thermography is still in demands.

## 2.2 Objectives and Description of the problems addressed throughout this research

The main purpose of this proposal is to design appropriate deep learning frameworks in thermal non-destructive evaluation for defect detection to extract and separate defects (including internal and invisible cracks and delamination etc.) efficiently and accurately. Furthermore, we are going to exploit the proposed deep learning algorithms to enhance the visibility of defects (subsurface defects case) and obtain defect inverse information for those infrared signals (including defect characterization issues: defect depth etc.), as well as achieve automatic defect detection by deep learning algorithms. Measurements are going to be conducted using active infrared thermography. For the knowledge of the author, there are very few works available in the literature dealing with defect detection in infrared

thermography using deep learning algorithms. The goal of this project also is to validate and develop those known techniques to the infrared non-destructive evaluation case.

The main scientific challenge lying in the successful application of these techniques is the limited training data of thermography for deep learning. Due to the issue such as the presence of noise, limited resolution, and probing of depths in TINDE, it is more difficult to precisely and fully extract defect information. For the defect characterization issue, it is complex to combine the information of thermal contrast and pixels of images with the depth in each pixel of defects to train the neural network.

A list of the objectives of this project are provide here:

1. Develop novel deep learning algorithms and frameworks to process inspected thermal data to enhance subsurface defect visibility in infrared thermography.
2. Building a novel automatic defect detection system by deep learning algorithms through inspected thermal data (including the case of less detectable subsurface defects in the limited infrared signals) and achieve the automatic condition monitoring.
3. Develop novel deep learning algorithms and frameworks to solve the defect inverse issues in infrared thermography using inspected data.
4. Build the numerical and experimental models to measure specific specimens in active infrared thermography and pre-process the experimental inspect data for deep learning feature extraction.

## **2.3 Methodology**

### **2.3.1 Infrared NDE**

In this project, different infrared NDE processing techniques will be used to inspect specimens. These techniques include Pulsed Thermography, Lock-in Thermography, and Inductive Thermography, etc. Infrared NDE techniques will be the fundamental techniques used for data acquisition of deep learning frameworks and pre-processing defects of ROI.

### **2.3.2 Deep algorithms in INDE Thermal data in Infrared NDE for defect detection.**

In this project, the innovative deep learning frameworks as shown below will be used to analyze thermal data in Infrared NDE for defect detection.

1) To improve the visibility of defects in the pulsed thermography, we introduce generative adversarial network (GAN). Proposing the GAN methods will allow the reconstruction of conspicuous images through the underlying features extraction in temperature evolutions of Pulse Thermography [10] [11]. We will use WGANs network for input dimension reduction

and visible feature extraction of defects in Pulse thermography by adding an explicit term in the loss that penalizes that solution which is an optimized approach to avoid uninteresting solutions to overcome model collapse and overfitting issues.

2) To perform defect segmentation in infrared thermography (PT, LT, etc.), we will introduce several types of detection and segmentation methods in deep learning, such as (i) objective detection: YOLO-V3 [12], Faster-RCNN [13]; (ii) semantic segmentation: U-net [14], Res-U-net [15]; and (iii) an instance segmentation model: Mask-RCNN, Center-Mask [16][17]. These neural network architectures could achieve detection and segmentation of defects automatically. A data augmentation is also introduced effectively to enlarge training data.

3) To achieve the defect depth detection in thermal sequence data, we will use a recurrent neural network (Gated Recurrent Units: GRUs) [18] to estimate the depth of defects in testing samples from the active thermography data. We will explicitly use the architecture of the simplified GRUs temporal model for spatial and temporal thermography data to be trained end to end. Due to the feature extraction, the deep auto-encoder architecture may be able to compress the meaningful feature and enhance the accuracy of defect depth estimation. A cross-validation routine will be introduced to improve the efficacy of the algorithm in the case of over-fit issues in deep learning.

4) We will try to use the Generative Adversarial Network and synthetic data from finite element modeling (COMSOL) to generate a series of simulated thermal data to offset limited thermal data and obtain enough training data to meet the standard for identifying the feature of the defect. Then a pre-trained Convolutional Neural Network (Mask-RCNN) could be used as a supervised feature extractor to analyze defects in specimens and fine-tune with specific thermal sequences to predict the vectorized features along with segmentation maps with detection probability.

### **2.3.3 The implementation steps and relationships of the novel deep learning defect detection frameworks**

In this project, we are going to discuss the subsections below as shown in each section which has certain connections as illustrated in Figure 3.1 including:

For a defect detection to adapt the deep learning algorithms and models, one could :1. Enhance the defect visibility and enhance the defect when we acquired the original thermal data from the thermal experiments to remove original noises; 2. Due to the limitation of the experimental data and the high expense of composites for the training, a data augmentation strategy will be introduced based on the synthetic data generation pipeline from software or augmented and generated data from unsupervised deep learning models, such as a generative adversarial network; 3. Automatic defect detection and segmentation without human

involvement and inspection with high accuracy results will avoid the issues associated with human fatigue.

For a quantitative analysis of the defect characterization issue, a defect depth estimation method using a recurrent neural network (Gated Recurrent Units (GRUs)) will be used to train with the stimulated data from the COMSOL software to evaluate the accuracy and performance of synthetic CFRP data from FEM for defect depth prediction.

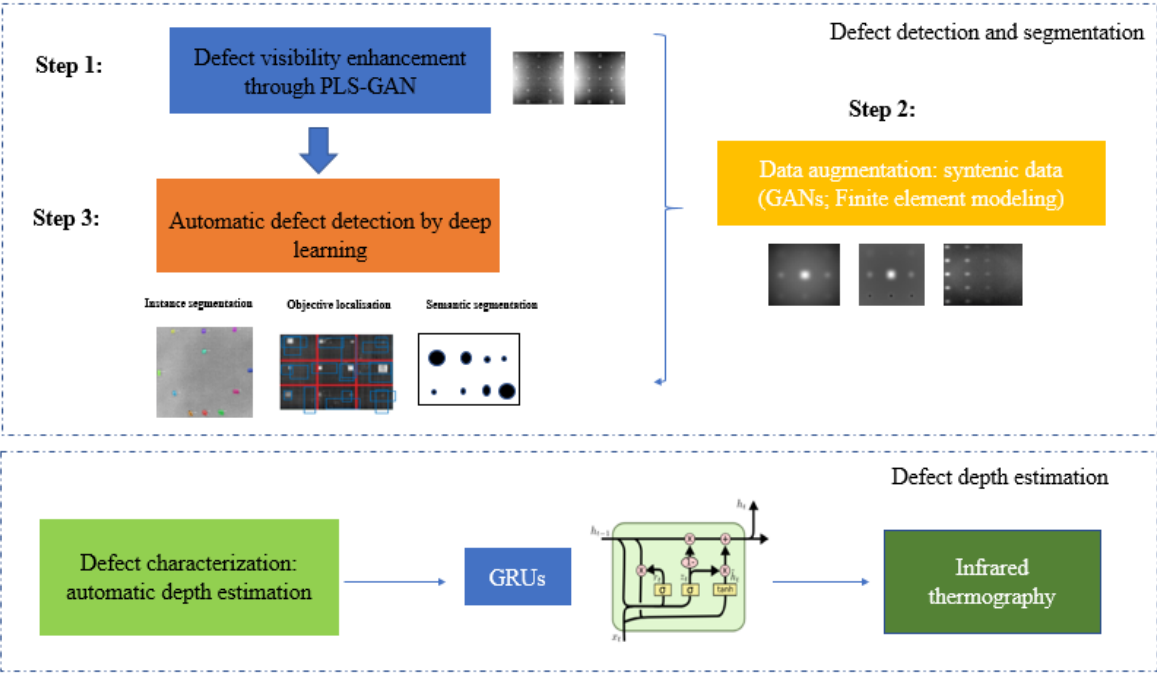


Figure 2. 1 Detection framework in this project

**2.3.4 Deep learning opensource software support**

Given that deep learning is the key to executing tasks of a higher level of sophistication, building and deploying them successfully proves to be quite the Herculean challenge for data scientists and data engineers across the globe. Deep learning frameworks offer building blocks for designing, training and validating deep neural networks, through a high-level programming interface. There is a myriad of frameworks that allows us to develop tools that can offer a better level of abstraction along with the simplification of difficult programming challenges. In this project, we use a flexible framework TensorFlow that allows users to build all relevant deep learning models of our designs through tensor arrays.

## References

- [1] X. Maldague, F. Galmiche, A. Ziadi, *Advances in pulsed phase thermography, Infrared Phys. Technol.* 43 (2002) 175–181.
- [2] N. Rajic, *Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures, Compos. Struct.* 58 521–528.
- [3] Mariacristina Pilla, Matthieu Klein, Xavier Maldague, and Antonio Salerno. *New absolute contrast for pulsed thermography. In Proc. QIRT, volume 5, 2002.*
- [4] Bardia Yousefi, Stefano Sfarra, Clemente Ibarra Castanedo, and Xavier PV Maldague. *Thermal NDT applying Candid Covariance-Free Incremental Principal Component*
- [5] *Thermography (CCIPCT). In SPIE Commercial+ Scientific Sensing and Imaging, pages 102141I–102141I. International Society for Optics and Photonics, 2017.*
- [6] Barlow H B. *Unsupervised learning[J]. Neural computation, 1989, 1(3): 295-311.*
- [7] Lecun Y, Bengio Y, Hinton G. *Deep learning[J]. Nature, 2015, 521(7553):436.*
- [8] Yousefi, Bardia, et al. "Application of Deep Learning in Infrared Non-Destructive Testing."
- [9] Dalmia, A., Kakileti, S.T. and Manjunath, G., *Exploring Deep Learning Networks for Tumour Segmentation in Infrared Images.2018.*
- [10] Xavier P.V. Maldague. *Theory and practice of infrared technology for nondestructive testing. John Wiley & Sons, Inc.*
- [11] Xie J, Xu C, Chen G, et al. *Improving Visibility of Rear Surface Cracks During Inductive Thermography of Metal Plates Using Autoencoder[J]. Infrared Physics & Technology, 2018.*
- [12] Redmon J, Divvala S, Girshick R, et al. *You Only Look Once: Unified, Real-Time Object Detection[J]. 2015:779-788*
- [13] Ren S, He K, Girshick R, Sun J. *Faster r-cnn: Towards real-time object detection with*

*region proposal networks. In Advances in neural information processing systems 2015 (pp. 91-99).*

- [14] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015:234-241.
- [15] Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 162: 94-114.
- [16] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
- [17] Lee Y, Park J. Centermask: Real-time anchor-free instance segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 13906-13915.
- [18] Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* 1997, 9, 1735–1780.

## **Part I. Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data**

The following chapter will present an article concerning the application of infrared thermography for Non-Destructive Testing & Evaluation applied in procedure. The whole section of this chapter was submitted to Nondestructive Testing and Evaluation Journal, 2021. The partial result of this study was firstly presented at an online oral session of SPIE Thermosense, then published in conference proceedings Volume 10214, Paper 11409-35, Thermal Infrared Applications XXXIX in 2020 (California, United States, remote electronic venue).

### **General explanation:**

During the industrial processing procedure, the inspection services involving manual inspection during the stages of quality control can be hampered by the fatigue of human inspectors. In this case, automatic quality control and defect detection become more vital so as to improve the inspection rates and achieve cost-effective condition monitoring.

In this study, we mainly focused on the proposed deep learning algorithms to achieve automatic defect detection and precise localization (subsurface defects case) from different thermal image sequences. To evaluate the efficiency and robustness of the proposed methodology, specimens containing artificial defects were selected for experimental configuration.

It systematically investigation and comparison of the different types of deep learning techniques for defect detection and analysis in Pulsed thermography. The innovative instance segmentation method is introduced for defects segmentation and identification for each object of defects with different specimens to predict each irregular shape of defects instance in the thermal images at the pixel level. It also introduced the numerical and experimental modeling and analysis for post-processing of inspected experimental thermal sequences from deep learning feature extraction.

## **Chapter 3 Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data**

*The results of this study were firstly presented at an oral session of SPIE Commercial+ Scientific Sensing and Imaging. International Society for Optics and Photonics, 2020, then it was published in Proceedings Volume 10214, Thermosense: Thermal Infrared Applications XXXIX; 102140T (2020), cited 15 times up to now.*

### **3.1 Résumé**

La thermographie infrarouge (IRT), est l'une des techniques les plus intéressantes pour identifier différents types de défauts tels que la délamination et les dommages existants pour la gestion de la qualité des matériaux. Les algorithmes de détection et de segmentation des objectifs en apprentissage profond ont été largement appliqués dans le traitement d'images, mais très rarement dans le domaine de l'IRT. Dans cet article, les méthodes de traitement d'image par apprentissage profond spatial pour la détection et l'identification des défauts ont été discutées et étudiées. Les algorithmes d'apprentissage profond ont été appliqués pour mettre en œuvre des modèles basés sur les réseaux de neurones convolutionnels (CNN) pour la détection de défauts structurels dans des échantillons en matériaux composites, puis ont été comparés à une méthode de combinaison de traitement d'image ordinaire. L'objectif de ce travail est d'intégrer de tels modèles d'apprentissage profond (DL) pour permettre l'interprétation automatique d'images thermiques pour la gestion de la qualité (QM). Cela nécessite d'atteindre une précision suffisamment élevée pour chaque méthode d'apprentissage profond afin qu'ils puissent être utilisés pour aider les inspecteurs humains en fonction de la formation. Il existe plusieurs alternatives de réseaux neuronaux convolutifs profonds pour la détection des images qui ont été employées dans ce travail, notamment : 1. les méthodes de segmentation d'instance Mask-RCNN (Mask Region based Convolutional Neural Networks) et Center-Mask ; 2. les méthodes de segmentation sémantique indépendante : U-net ; Resnet-U-net ; 3. les méthodes de localisation objective : You Only Look Once (YOLO-v3) ; Faster Region based Convolutional Neural Networks (Faster-RCNN) ; en outre, une méthode de traitement combinée de segmentation d'images infrarouges ordinaires (Absolute thermal contrast (ATC) and global threshold) a été introduite pour comparaison. Une série d'échantillons académiques composés de différents matériaux et contenant des défauts artificiels de différentes formes et natures (trous à fond plat, inserts en téflon) ont été évalués et tous les résultats ont été étudiés afin d'évaluer l'efficacité et la performance des algorithmes proposés. Les principales mesures d'évaluation ont été analysées sur la base des résultats de détection des modèles DL : Probabilité de détection (POD) ; précision moyenne (mAP) ;



complexité de l'accélération du temps. Les résultats montrent que les modèles DL ont la fiabilité requise pour l'évaluation des défauts par détection automatique. En comparant les résultats expérimentaux de ces méthodes, le masque central est le plus prometteur en termes de précision et YOLO-V3 est le plus rapide en termes de temps, d'après l'évaluation de la base de données de tests infrarouges.

**Contributing authors:**

**Qiang Fang (Ph.D. Candidate):** a part of the experiment planning, data collection, data analysis, designing and implementing the algorithm. Moreover, testing their accuracy and robustness throughout the process and writing the manuscript.

**Ba Diep Nguyen:** assisting the experiment planning, data collection, preparation.

**Yuxia. Duan:** assisting the experiment planning preparation.

**Clemente. Ibarra. Castanedo:** the experiment planning, revision and correction of the manuscript.

**Xavier Maldague (The research director):** supervision, revision and correction of the manuscript.

**Other contributors:**

**Annette Schwerdtfeger:** (research officer): manuscript preparation.

# Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data

Qiang Fang 1\*, Ba Diep Nguyen1, Yuxia. Duan1, Clemente Ibarra-Castanedo1 and Xavier Maldague 1\*

1 Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Université Laval, 1065, av. de la Médecine, Québec, QC G1V 0A6, Canada;

\* Correspondance: qiang.fang.1@ulaval.ca (Q.F.); Xavier.Maldague@gel.ulaval.ca (X.M.)

## 3.2 Abstract

Infrared thermography (IRT), is one of the most interesting techniques to identify different kinds of defects such as delamination and damage existing for quality management of material. Objective detection and segmentation algorithms in deep learning have been widely applied in image processing, although very rarely in the IRT field. In this paper, spatial deep learning image processing methods for defect detection and identification have been discussed and investigated. The deep learning algorithms were applied to implement Convolutional neural networks (CNNs) based models for detecting structural defects in samples made of composite materials, then were compared with a regular image processing combination method. The aim in this work is to integrate such deep learning (DL) models to enable interpretation of thermal images automatically for Quality management (QM). That requires achieving a high enough accuracy for each deep learning method so that they can be used to assist human inspectors based on the training. There are several alternatives of deep Convolutional Neural Networks for detecting the images that were employed in this work which included: 1. the instance segmentation methods Mask-RCNN (Mask Region based Convolutional Neural Networks) and Center-Mask; 2. the independent semantic segmentation methods: U-net; Res-net-U-net; 3. the objective localization methods: You Only Look Once (YOLO-v3); Faster Region based Convolutional Neural Networks (Faster-RCNN); in additionally a regular infrared image segmentation processing combination method (Absolute thermal contrast (ATC) and global threshold) has been introduced for comparison. A series of academic samples composed of different materials and containing artificial defects of different shapes and nature (flat-bottom holes, Teflon inserts) have been evaluated and all results were studied in order to evaluate the efficacy and performance of the proposed algorithms. The key evaluation metrics were analyzed based on the detection results from DL models: Probability of Detection (POD); Mean-average Precision(mAP); Time speeding complexity. The results show that DL models have the reliability required for the assessment of defects by automatic detection. Comparing the experimental results among these methods, the Center-Mask is the most promising one for accuracy and YOLO-V3 has the faster time frame speed based on the infrared testing database evaluation.

**Keywords:** Deep learning Non-destructive Evaluation (NDE); Automatic defect identification and segmentation; Infrared Thermography; Pulsed thermography; Infrared image processing; Convolutional neural network;

### 3.3 Introduction

Quality management (QM) [1] is playing a crucial role in modern industrial production fields. A qualified quality management and controlling system can be a significant technology advancement for industry and manufacturing applications such as in the aerospace field. In order to preserve the health structural monitoring, the application of visual inspection systems becomes more and more essential in the production lines. However, the inspection services involving manual inspection during the stages of quality control can be hampered by the fatigue of inspectors. In this case, automatic quality control and defect detection become more vital so as to improve the inspection rates and achieve cost-effective [2] condition monitoring.

Non-destructive Evaluation (NDE) [3] is a group of techniques used in industry for analyzing and evaluating the properties of a material without causing damage. NDE methods rely upon use of electromagnetic radiation, sound and other signal conversions to examine a wide variety of articles. Due to the analysis of thermal front propagation, each thermal non-homogeneity perturbs the thermal waves propagation on the surface of specimen in comparison to the surrounding sound region. We can then see the changes of the temperature variation. The infrared camera and corresponding equipment can record this thermal perturbation, and the results can be analyzed to obtain further information. Infrared inspection techniques have been applied frequently to evaluate subsurface defects and hidden structures etc. for the quality control of materials such as metals, composites and so on.

Infrared Thermography (IR) is one of many NDT techniques used to “see the unseen” is a non-destructive technique of measuring and mapping surface temperature of the materials. The main objective of Infrared Non-Destructive Evaluation (INDE) is to detect and classify the Regions of Interest (ROIs) that could be represented as a defect or an anomaly by analyzing the sequence images or singular image. In addition, another goal is to detect such ROIs as accurately as possible, visible when the conditions are maintained invariantly. Pulsed Thermography (PT) is one of the types of infrared thermography that is based on the heat flux diffusion (radiation) and energy absorption from the instant impulse, then to evaluate and visualize the defects based on the temperature difference which is reflected from the surface of Specimens.

For defect identification and detection, several state-of-the-art regular methods in infrared thermography especially on pulsed thermography are powerful tools which have been

documented to give noticeable results, such as the Pulsed Phase Thermography (PPT) [4], Principal Component Thermography (PCT) [5], Difference of Absolute Contrast (DAC) [6], Thermographic Signal Reconstruction [7], as well as Candid Covariance Free Incremental Principal Component Thermography [8]. These mentioned methods show remarkable efficacy in improving defect visibility during INDE and also have demonstrated to be powerful tools that have been well-documented and provide noticeable results in Infrared Thermography. However, although these methods are pattern-based unsupervised methods [9] which means that they can extract defect information distinctly from time and spatial domains. The search for new methods to further improve defect visibility and achieve automatic detection is still being crucial issues in TNDE literature.

Moreover, in a general way, every contribution working for defect detection with thermography in Nondestructive Evaluation (NDE) for Quality management (QM) has to deal with image processing issues. Some methods and networks have already introduced the researcher's efforts applied to other fields and they also have the potential capability to provide good results for Pulsed thermography (PT) in NDT, such as: (1) Artificial Neural Networks(ANNs; the machine learning model )which is discussed in [10] to achieve fiber orientation assessment from composite materials on a learning process based on interconnected elements (neurons); (2) A SLFNs (Single hidden Layer Feed forward neural net-works) which provide an unlimited number of neurons in one hidden layer for defect classification [11]; (3) A SVM(Support Vector Machine) automatic classification model has been proposed for breast cancer detection through images from Thermal pattern[12]; (4) the k-means clustering method that was applied on automatic defect detection in fruits for the quality classification[13]; (5) a self-learning softmax with a 9-layer Convolutional Neural Network (CNN) model was constructed in [14] to identify the near nighttime pedestrians which has potential competitive accuracy to classify(background; pedestrians; the vehicles) in real-time recognition.(6) In [15], a Fully Convolutional Network(FCN) has been introduced to estimate the surface damages from deck areas of bridges.

On the other hand, initially developed in order to make neural networks more efficient, deep learning methods have already shown their ability to outdate most of the other approaches existing previously in a great number of applications, which has contributed to making them very popular in many scientific communities. It also has been proven that Convolutional Neural Network (CNN) from Deep learning (DL) can perform well in Infrared Non-Destructive Testing, due to the fact that the vectors extracted from convolutional neural networks can be utilized as features for defect detections via Pulsed thermography.

To the best of our knowledge, there are only a few works available in the literature dealing with automatic defect detection in Pulsed thermography by using deep learning algorithms.

Although some articles have introduced a few single algorithms from DL for defects detection analysis, there is no systematically investigation and comparison of DL algorithms for defect detection and analysis in Pulsed thermography. Due to the particular characteristics from the spatial deep learning model, the implementation of supervised learning algorithms in infrared non-destructive evaluation (pulsed thermography) [16] is a potential tendency in the future research.

The main purpose of this research is to introduce the appropriate deep learning frameworks with thermal non-destructive evaluation of pulsed thermography for automatic defect detection so as to extract and separate defects (including internal and less visible cracks and delamination structure cases, etc.) efficiently and accurately. The pro-posed methods can effectively achieve the defect identification and segmentation with the cases of data limitation.

Therefore, a small amount of pulsed thermography experimental data will be used to train the deep spatial characteristic models for identification or segmentation of defects in this research. The convolutional neural network will be adapted to extract the specific feature for each visible region, then force the learning system to learn how to distinguish and detect less visible defects from limited infrared signal images based on the position and extraction of specific features. Three types for defect detection methods will be introduced and investigated in detail for automatic defect detection and identification of Pulsed (infrared) thermography in an infrared system respectively. In addition, a regular infrared segmentation method (Absolute Thermal Contrast (ATC) and global threshold) [17] has been introduced as well for the purpose of comparison with DL models. A systematic discussion and comparison of performance will be pro-vided in the further sections.

Three types of deep learning algorithms which have achieved potential results of defect identification or segmentation in PT thermal data will be discussed in this article, including:

a) Objective localization [18]

The objective detection algorithms can detect multiple objects with their bounding boxes in the images. It can identify or locate the defects in the real time thermal sequence and images.

b) Semantic segmentation [19]

The semantic segmentation associates each pixel of an image with a categorical label as a single entity. The idea of semantic segmentation is recognizing and understanding the objects (defects) at the pixel level.

c) Instance segmentation [20]

The idea of instance segmentation is that given an input image and to predict an image, the location and identities of objects in that image is carried out similar to object detection. But

rather than just predict a bounding box for each of those objects, instead it can predict a whole segmentation mask for each of those objects and predict which pixels in the input image correspond to each object instance.

The contribution of this work can be illustrated in the following:

- 1) systematic investigation and comparison of the three types of classical deep learning methods for the defect detection accuracy and efficiency analysis in Pulsed thermography.
- 2) an innovative instance segmentation method is being introduced for defect segmentation and identification for each defect with different specimens to predict each irregular shape of the defect instance in the thermal images at the pixel level.
- 3) experimental modeling and analysis for post-processing of inspected data is introduced, based on deep learning feature extraction.

The remainder of this paper is structured as follows: Section 2 provides the main principles and methods; Section 3 provides the introduction of the pulsed thermography (PT); Section 4 includes the related experimental set-up indication, the data and defect features and samples description; Section 5 illustrates the methodologies from the spatial deep learning models: (a)YOLO-V3[21]; (b) Faster-RCNN[22]; c) U-net[23]; (d) Resnet-U-net[24]; (e) Mask-RCNN[25]; (f) Center-Mask[26]. A full experimental results and training procedure from each investigated algorithm is provided in Section 6. Section 7 will present the discussion on the experiment results. Section 8 concludes the research and outlines the future work.

### **3.4 Principles**

In this section, a detection system trained with pulsed thermography data has been proposed to segment and identify defects in thermal images. The spatial characteristic deep learning model is introduced separately and comparatively in this strategy as shown in Figure 3.1. The design of this defect detection system is based on the three types of detection frameworks. The implementation steps can be illustrated as follows:

1. First, the infrared thermal sequences are acquired by the pulsed thermography (PT) system;
2. Secondly, the raw thermal sequences are preprocessed and decomposed by augmentation methods: 1. Principal Component thermography (PCT): sequences being decomposed into several orthogonal functions (Empirical Orthogonal Functions: EOF); 2. Flip; 3. Random crop; 4. Shift; 5. rotation etc.;
3. In the final step, the defect regions are recognized via deep neural networks which visualized the defects with the bounding boxes. All defects have to be labeled with the locations, then trained with the deep region neural network.

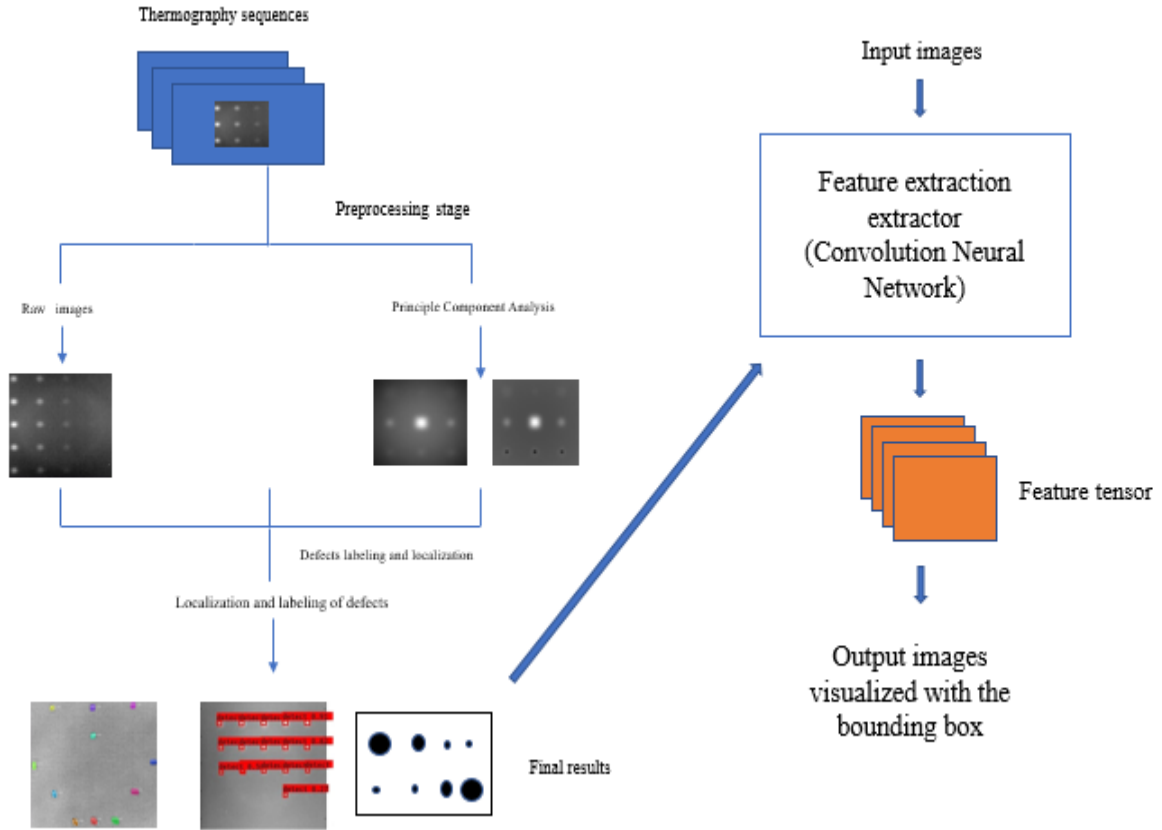


Figure 3. 1 Proposed detection strategy (Instance segmentation; Semantic segmentation; objective detection)

### 3.5. Thermography Consideration- Optical pulsed thermography

In PT [27], a high-power thermal pulse is applied to the surface of the specimen through heat radiation. Due to the heat conduction of the thermal front absorbed by the specimen's surface, the thermal front travels from the surface and propagates through the materials. As the time elapses, the surface temperature will decrease uniformly for a zone with-out defect. Conversely, if there is an internal defect beneath the surface (e.g. delamination, disbands, damage, etc.), this defect can become a resistance to heat flow that produces higher temperature patterns at the surface with a decay of temperature, which can be inspected by an infrared (IR) camera. Figure 3.2 indicates the fundamental principle of pulsed thermography. In a solid of semi-infinite isotropic conduction, a 1D solution of the propagation of the pulse of a Dirac heat pulse is indicated in Equation. (3.1) as a Fourier mathematical equation [28],

$$T(z, t) = T_0 + \frac{Q}{\sqrt{kpct}} \exp\left(-\frac{z^2}{4at}\right) \quad (3.1)$$

where the energy absorbed by the surface is  $Q [J/m^2]$  and  $T_0[K]$  is the temperature of initialization. The surface temperature progression at  $T(0, t)$  can be written as follows:

$$T(0, t) = T_0 + \frac{Q}{e\sqrt{\pi t}} \quad (3.2)$$

from Equation. (3.2), where  $e=kpc$  is the effusivity. The temperature of surface evolution following a Dirac heat pulse will decay as a monotonous decrease as  $t^{-1/2}$  without defects, while areas with defects will diverge more or less from this behavior based on the actual thermo-physical properties of the region.

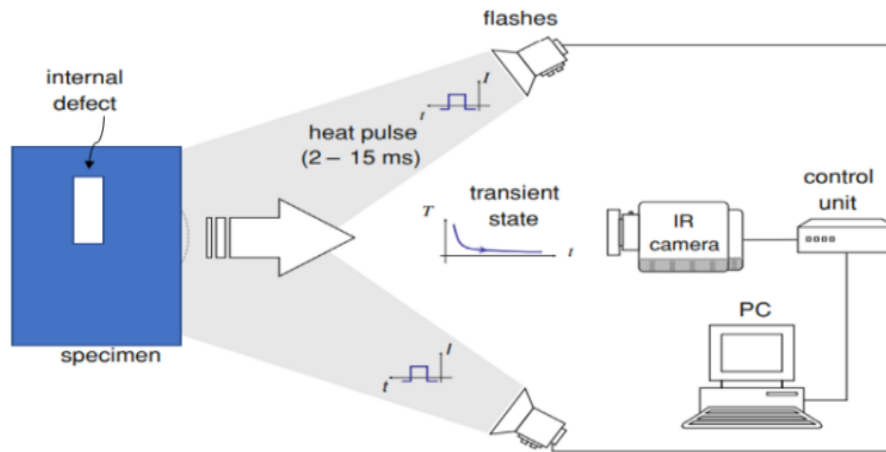


Figure 3. 2 Pulsed thermographic testing using optical excitation

### 3.6 Specimens and Experimental Setting up

#### 3.6.1 Experiment Setup

Infrared measurement and the inspected system are the essential parts of collecting infrared data from pulsed thermography. In order to evaluate the robustness of the proposed algorithms, a certain number of samples was tested. In general, the inspected system used in this experiment consists of: Two photographic flash lamps (Balcar FX 60,5 ms thermal pulse) 6.4 kj/flash, an infrared thermal camera and a personal computer (PC)- Ubuntu 14.04 as shown in Figure 3.3. To be more detailed, the sampling rate was 157 Hz, a total of three types (steel; CFRP; plexiglass) of 8 pieces of specimens were inspected.

The analysis of the thermography process has been conducted with the PC (Intel(R) Core (TM) i7-2600 CPU, 3.40 GHz, RAM 16.0 GB, 64-bit, Operating System) and the processing of the thermal data has been conducted using the MATLAB computer program R2019a and a Tensor-flow deep learning open source library. A Mid-wave infrared (MWIR) camera with a special mid-infrared lens (to filter the MWIR spectrum) and two normal lamps was utilized



for collecting the infrared data. The normal lamp (containing the entire visible spectrum) was used as an illumination source to illuminate the specimen during the inspection performed inside the laboratory.

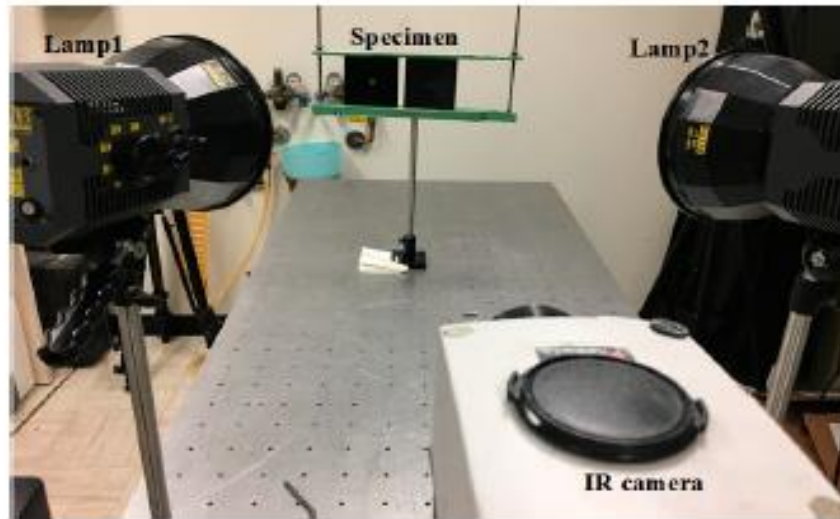


Figure 3. 3 Pulsed thermography experiment platform

### 3.6.2 Validation Samples Preparation

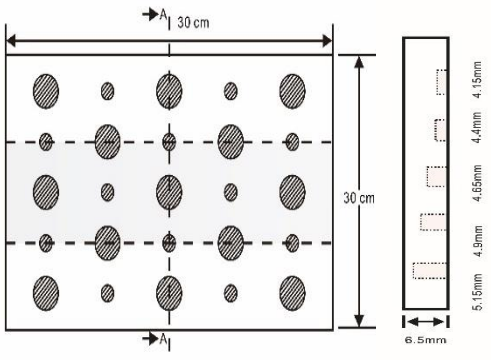
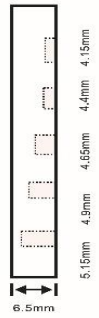
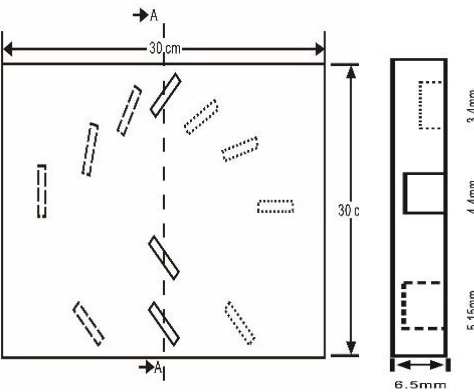

In order to evaluate the performance of the proposed method, academic samples were collected independently from three types of materials: Plexiglas (Plexi), Carbon Fiber-Reinforced Polymer (CFRP), Steel. All of the experiments with DL models were conducted under the databases collected from these samples.

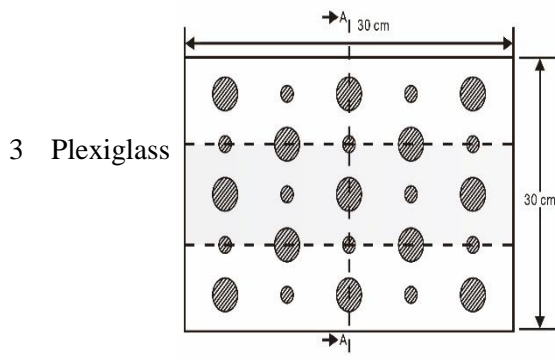
As shown in Table 4.1, the description of eight validation samples in this work are explained. The aspect ratio (*size/depth*) for all trained and validated samples is designed at  $[0,60]$  in order to reveal if the detection model has a flexible performance to detect defects. Among the eight validated specimens, the detailed description can be illustrated as follows:

1. The 1st sample (a) is from plexiglass material with 25 sub-surface circle defects of different diameter and depth;
2. The 2nd sample (b) has 8 multiple angle defects which are embedded on the surface of the plexiglass specimen;
3. The 3rd sample (c) is from plexiglass material with 25 sub-surface circle defects of same diameter but different depths, increasing from the left to right column (deeper);
4. The 4th sample (d) Plexiglass has 25 circle and quadrilateral defects of various depth and size;

5. The 5th sample (e) is a steel sample that has three different diameters of circle defects; the depth being shallower from top to bottom
6. The 6th sample (f) CFRP has 25 triangle defects embedded in the specimen in the form of a folding plane;
7. The 7th sample (g) CFRP has 25 triangle defects embedded in the specimen in the form of a flat plane;
8. The 8th sample (h) CFRP has 25 triangle defects embedded in the specimen in the form of a curved plane.

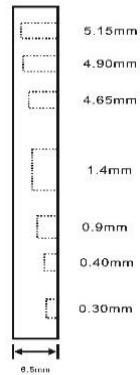
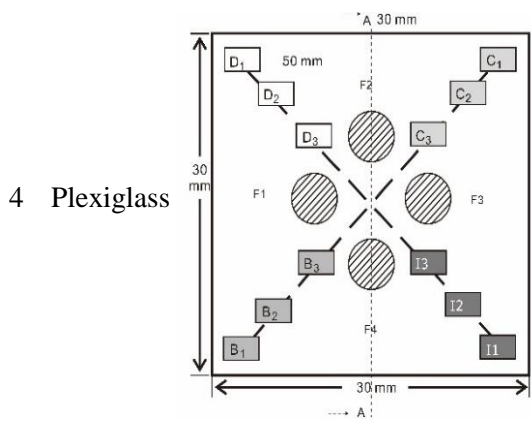
Table 3. 1 The description of each experimental sample

Number	Type of Materials	Geometrics Specimen	Cross Section	Dimension	Defect Diameters (mm)
1	Plexiglass			30*30cm	Depth: 5.15 mm, 4.9 mm, 4.65 mm, 4.4 mm, 4.15 mm; Diameter: 9 mm, 18 mm
2	Plexiglass			30*30cm	Different angle cracks (0°, 15°, 30°, 45°, 60°, 75°, 90°); Size: 15 mm × 3 mm; Depth: 3.4 mm; 4.4 mm; 5.15 mm;



Depth: 3.4 mm; 4.4 mm; 5.15 mm;  
Diameter: 10mm

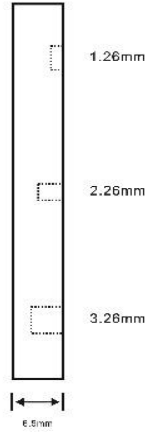
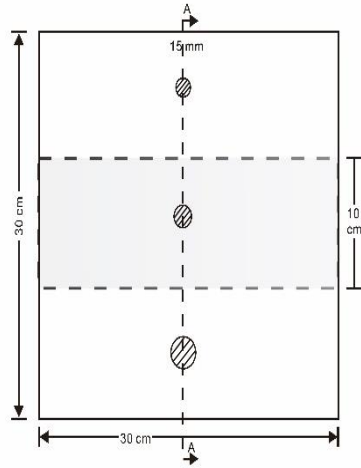
30\*  
30c  
m



Depth:  
0.2, 0.4, 0.6, 0.8, 1.0;  
Diameter or Size:  
3.4, 5.6, 7.9, 11.3, 16.9

D1: 15 mm x 6 mm; D2: 8 mm x 5 mm  
D3: 5 mm x 4 mm; C1: 14 mm x 6 mm  
C2: 9 mm x 5 mm; C3: 6 mm x 4 mm  
B1: 16 mm x 6 mm; B2: 15 mm x 6 mm  
B3: 8 mm x 4 mm; I1: 18 mm x 6 mm  
I2: 11 mm x 6 mm; I3: 7 mm x 4 mm

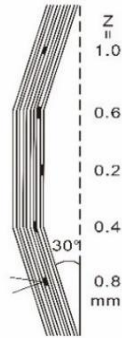
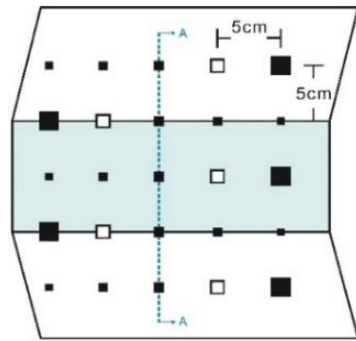
5 Steel



30\*  
30c  
m

Depth:  
Size:  
A = 5 x 5.0 mm<sup>2</sup>  
B = 2.5 x 10 mm<sup>2</sup>  
C = 1 x 25 mm<sup>2</sup>

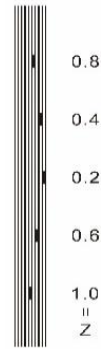
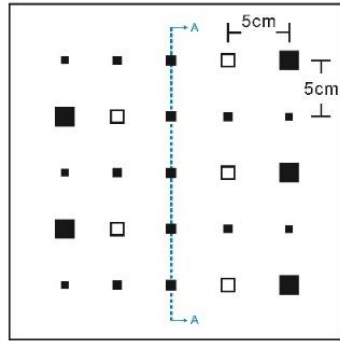
6 CFRP



30\*  
30c  
m

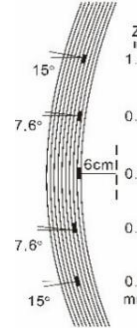
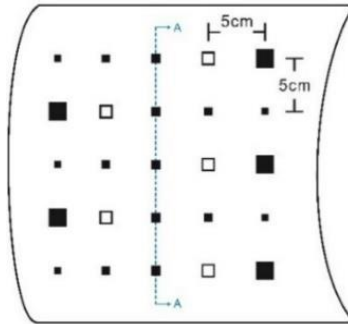
Five equivalent diameters  
(3.4mm, 5.6mm, 7.9mm,  
11.3mm,  
16.9mm)  
with five different depth of  
defects (1.0mm; 0.6mm;  
0.2mm; 0.4mm;  
0.8mm)  
The plate has two time folding  
and at 30 degrees to the  
horizontal level

7 CFRP



Five different lateral size of defects:(3mm, 5mm, 7mm, 10mm, 15mm) with five different depth (0.2mm; 0.4 mm; 0.6mm; 0.8mm; 1.0mm)

8 CFRP



Different angle of defects ( $0^\circ$ ;  $7.6^\circ$ ;  $15^\circ$ ) with different equivalent diameter (3.4mm, 5.6mm, 7.9mm, 11.3mm, 16.9mm) and corresponding depth ( $(1.0\text{mm}, 15^\circ)$ ,  $(0.6\text{mm}, 7.6^\circ)$ ,  $(0.2\text{mm}, 0^\circ)$ ,  $(0.4\text{mm}, 7.6^\circ)$ ,  $(0.8\text{mm}, 15^\circ)$ )

### 3.7 Validation Datasets and features

#### a) Acquisition of the training database

In order to maximize the probability of detection, we independently sampled 4000 thermal images in total from the pulsed thermography experiment in three types of materials (Plexiglas, Carbon fiber reinforced polymer (CFRP), Steel) respectively to build a training and testing database from pulsed thermography data. As the images used for training should be the same size, the database was split into 512x640 pixels.

#### b) Calibration of the data

The marking process was conducted with the Two labelling software based on the model type: Colabeler toolkit (YOLO-V3; Faster-RCNN); LabelMe toolkit (Mask-RCNN; Center-Mask; U-net; Res-Unet).

In the Colabeler toolkit, only one label (square shape label) was used for all of the different kinds of marks in Figure 3.4 (a). The bounding boxes were prepared by hand for each of the images, then exported to a .xml file by Colabeler. Each bounded defect was used as training for the algorithm. The process has to be repeated for all images used for training. Each representative image file from the four types of samples was extracted from the sfmov.format sequence files or matrix raw files. These samples created multiple shapes of defects in the database such as squares and rectangles.

In the Labelme toolkit, a different labeling curve in Figure 4.4(b)(c) from the procedure will be provided regardless of the shape of the defects for segmentation, a labeling curve on each object in the images is then exported to a json. file by labelme to transform into a large-scale object segmentation database (COCO).

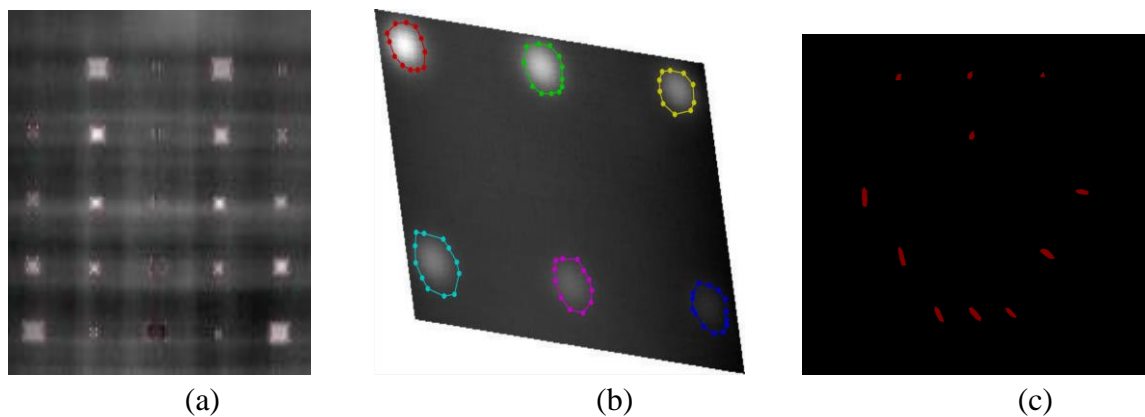


Figure 3. 4 Processing of labeling

#### c) Preprocessing and data augmentation

In the case of the overfitting issue during the training, data augmentation plays a significant role.

We encourage this model to learn the invariant and transformations by using rotation and flipping for the raw images. Since the defects in these materials remain in permanent positions and shapes, it leads to a requirement of capturing images in diverse conditions. As known, the defect is not clear because of the shaping process and/or the specifications of materials that lead to captured images on cluttered background. Those reasons lead to the augmentation of the captured images before entering them into a deep learning network which is important. Partial images for the training are undertaken in a preprocessing stage.

We adapted the preprocessed sequence images from feature extraction methods: Principal Component Thermography (PCT) which extracts meaningful features by dimension

reduction and reflects the intuitions of the data. For example, when the data arise from the high dimensional form (sparse and unstable estimation), the PCT can give more redundancy to our classifier to enable them to make a better decision.

### 3.8 Methodologies: Defect detection methods by deep learning algorithms

As shown in Figure 3.5 below, three main deep learning feature extraction methods and their implementation steps have been introduced: A. Objective localization algorithms: Method 1. Single stage real time algorithm-You Only Look Once (YOLO-V3), and Method 2. Two stages real time algorithm- Faster Region based Convolutional Neural Networks- Faster-RCNN. B. Semantic segmentations: Method 3. U-net, and Method 4. Res-U-net. C. Instance segmentation: Method 5. Mask-RCNN, and Method 6. Center-Mask. D. Regular thermal segmentation: Method 7. The Absolute thermal contrast with global threshold.

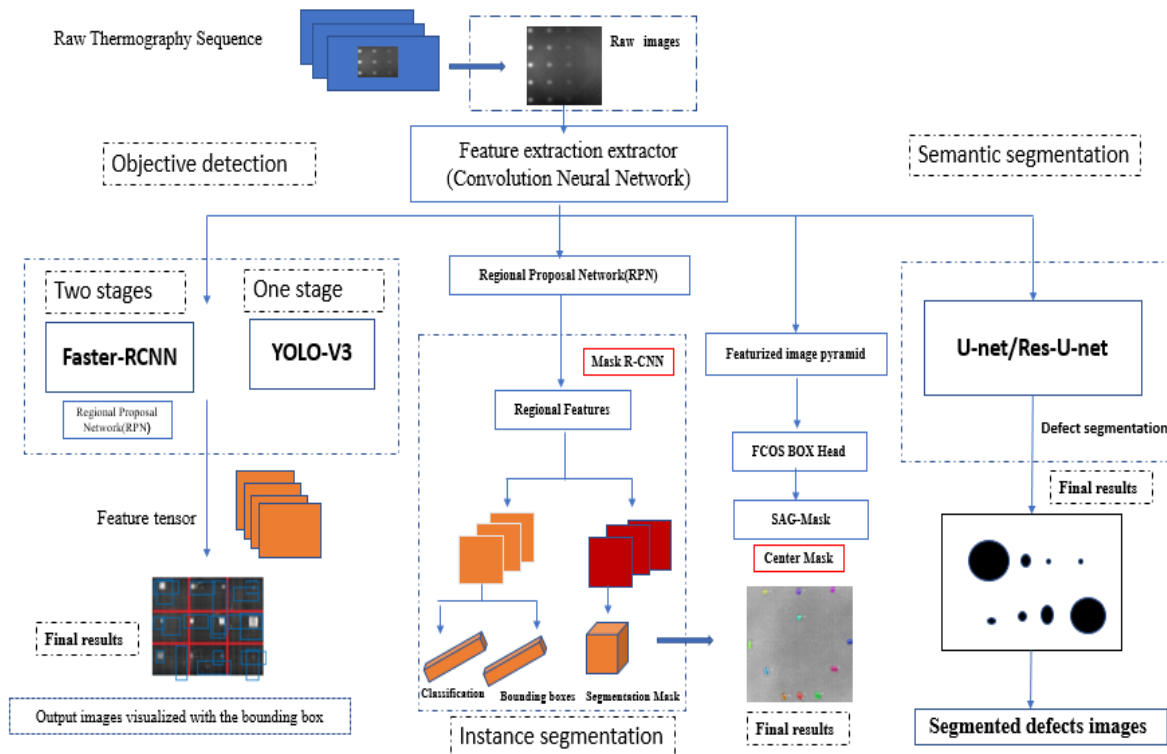


Figure 3. 5 Three types of Deep Learning methods (Objective detection; Instance segmentation; Semantic segmentation)

#### A. Objective localization algorithms

- Method 1: Real time defect localization (YOLO-V3)

YOLO-v3 is a proposed supervised deep learning algorithm which has excellent detection capability both on the large or small objectives due to its concatenation through merging the features from the earlier layer with the features from the deeper layer especially during the infrared nondestructive evaluation with an automatic defect detection task (subsurface defects case).

Processing images with YOLO v3 is quite fast and simple, allowing defects to be detected and localize directly. In order to perform the feature extraction, residual networks and successive  $3 \times 3$  and  $1 \times 1$  convolutional layer are localized in YOLO v3 in Figure 3.6. The skip connections mechanism achieved by residual networks through multiple residual units [9-10], which was proposed to improve the performance of objective detection and also solve the gradient vanishing issue. In this research, the YOLO v3 based deep architecture neural network are proposed to perform the detection of defects (at various sizes). This algorithm includes the implementation of three steps. First, the pictures are resized as the input size; Then an entire convolutional network is run on these pictures; Lastly, we threshold the detection results based on the model confidence scores.

In the Figure 3.7, it is shown an example of an original image(a) and a detected image (b) from YOLO-V3 network. The CNN could be able to distinguish the components which has similar thermal pattern with defects during the processing of thermal diffusion, which indicated the supervised learning method (YOLO-V3) is less influenced by the boundary information in the components.

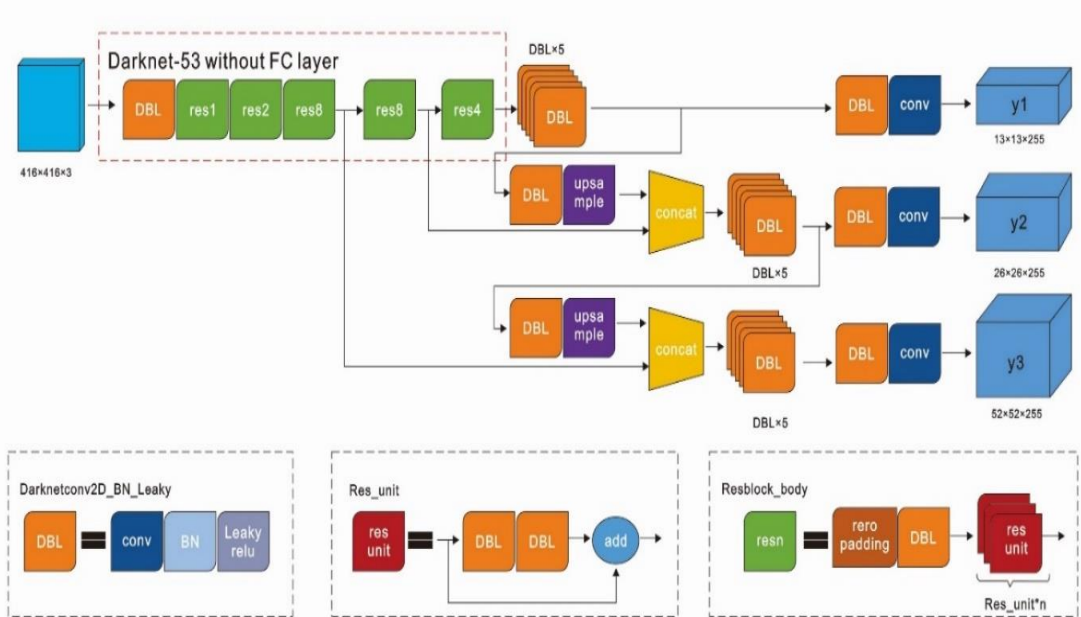


Figure 3. 6 The architecture of Residual units in Yolo-v3



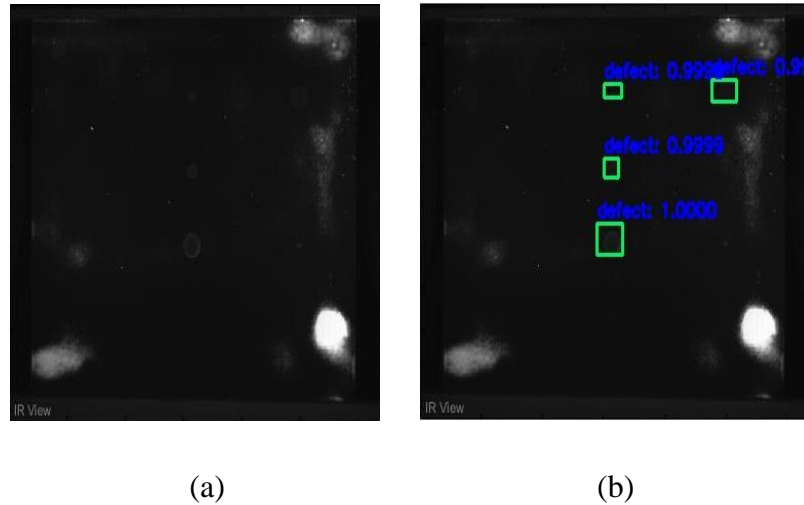


Figure 3. 7. An example of method A (a) the original thermal image; (b) the detected image

- Method 2: Real time multiple stages defect localization model (Faster-RCNN)

Faster-RCNN is a real time detector which achieved satisfying accuracy with several previous object localization applications in NDT [29]. In 2018, the Faster-RCNN has been used for crack detection in an eddy current thermography diagnosis system. The neural network based on a deep architecture was proposed to deal with the problem of accurate crack detection and localization via the preprocessing unsupervised method (Principal Component Analysis).

The deep architecture of Faster-RCNN is composed of several modules (Figure 3.8):

1. a fully convolutional network which included five blocks of basic convolutional layers, a Relu layer with pooling layer to extract feature from the input images;
2. a region proposal network (RPN) connected with the fully convolutional network to obtain the region of interest (RPI);
3. a Fast-RCNN detector using the feature region extracted in the (1)-(2), to achieve bounding box regression and SoftMax classification.

The Faster R-CNN trained from multi-properties rather than the regular unsupervised method was limited with respect to certain properties that the defect information contained. An example image detected from Faster-RCNN and a corresponding original thermal image are shown in Figure 3.9.

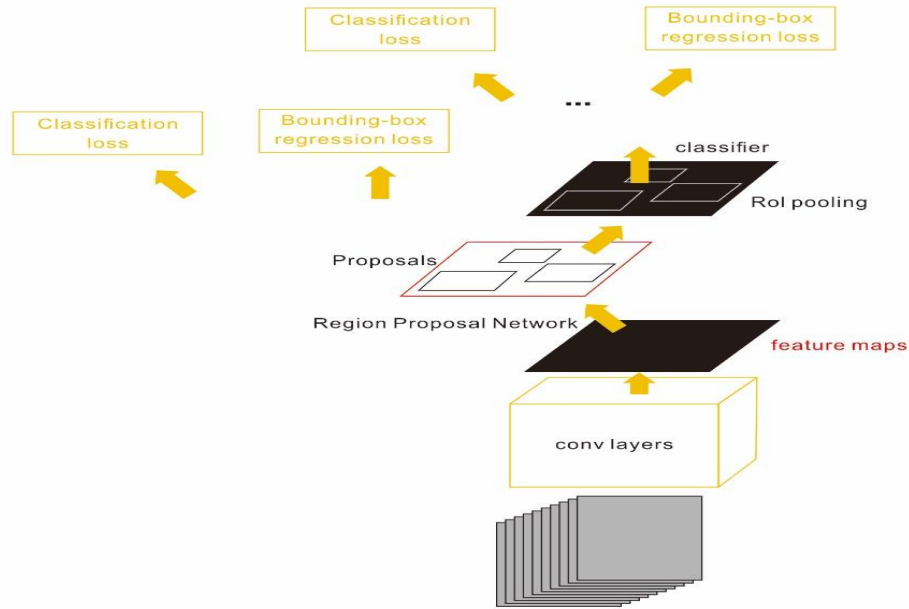


Figure 3. 8 Faster-RCNN defect detection for infrared data

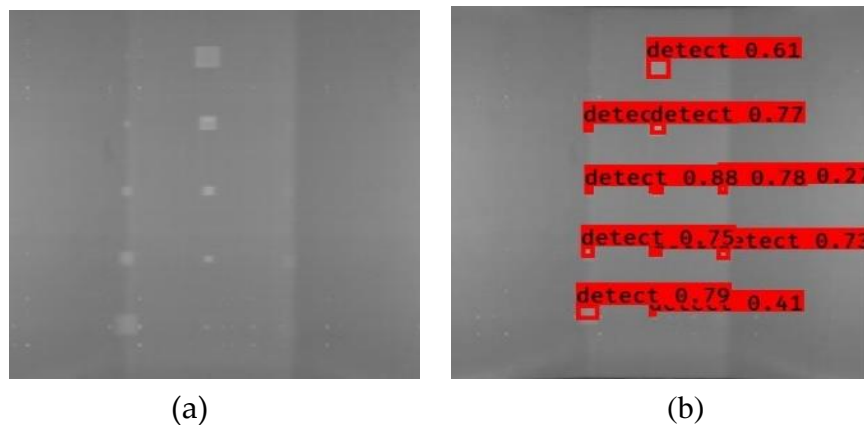


Figure 3. 9 An example of method B (a) the original thermal image; (b) the Faster-RCNN detected image

### C. Semantic defect segmentation method

- Method 3: U-net for semantic defect segmentation

The U-net is an excellent auto-encoder format model to handle the training data with dimensionality reduction and data augmentation. It is worth evaluating the performance of semantic segmentation by U-net after extracting objective features from the temporal infrared sequence. In the previous article [30], the U-net has been employed for the segmentation of

wildland and forest fires as a deep-fire convolutional network obtaining very good performance.

The convolutional architecture of U-net is inspired from the auto-encoder network architecture as indicated in Figure 3.10. Contracting path maps from the original image to a low dimension vector by extracting meaningful feature representations and the expansive path reconstructs the output of the desired feature maps. The contracting path is composed of a group of convolutional blocks: convolutional layers; rectified linear unit (ReLU) [31]; max pooling (dimension reduction). The expansive path included groups of reconstruction blocks to up-sample the feature: up-conv (half-reduce the feature channels); concatenation with a feature map from cropping in the contracting path and so on. In the final layer, the feature vectors are classified into the target number of the class by 1x1 convolution. Moreover, this architecture relies heavily on data augmentation for its performance, which is explained in below. The data augmentation strategy from the U-net architecture also brings a significant benefit for the performance for the training. Due to the characteristics of the spatial thermal temperature sequence, the infrared thermal profile for the defect and non-defect pixels can be distinguished based on the labeling to force implementation of the supervised learning method (U-net segmentation).

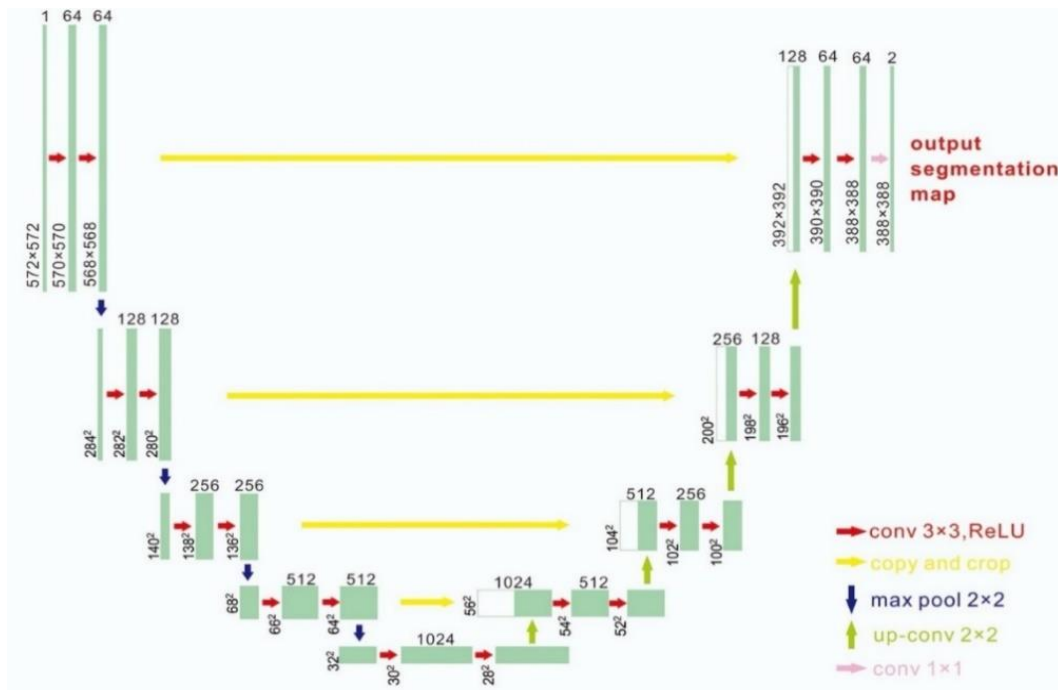


Figure 3. 10 U-net Model structure

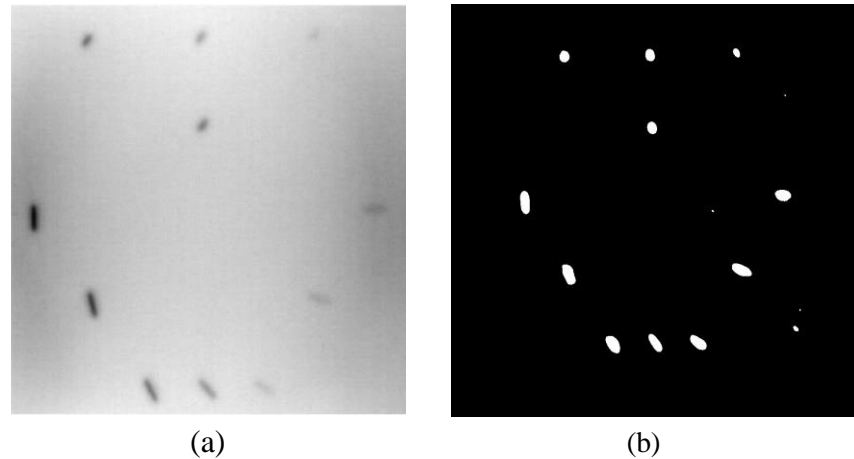


Figure 3. 11 An example of method C (a) the original thermal image; (b) the segmented image

During the cooling period of the thermal data, a temperature change curve over time is obtained on the given image sequence. Therefore, each single thermal frame is fed into this model at the pixel level and the thermal image can gradually capture the physical properties of temperature variation by U-net. The input values of U-net are particular thermal temporal evaluation vectors from each pixel. The output label is set either as 1 or 0 corresponding to the defect or non-defect region. During the validation stage, an obtained thermal sequence is selected as the input data after de-background and normalization. The output is a segmented image reconstructed from the predicted value as shown in Figure 3.11(b). The Figure 3.11(a) is the corresponding original thermal image.

- Method 4: Res-U-net for defect semantic segmentation

It is worth investigating comparatively to evaluate thermal sequence databases based on these different defect segmentation methods. As indicated in Figure 3.12, Res-U-net is an adapted novel encoder/decoder structure evolved from U-net in combination with several particular structures: residual connections [32]; atrous convolutions [33]; pyramid scene parsing pooling [34]. Res-U-net can infer sequentially the boundary of the objects, the distance transforms of the segmentation mask, the segmentation mask and a colored reconstruction of the input.

Due to the fact that Residual blocks in Res-u-net are able to remove vanishing and exploding gradients [35] to a great extent to improve the implementation efficacy of the learning mode and to achieve the pixel level of segmenting of defects and classification, it was compared with other state-of-the-art DL algorithms. The Res-Unet original was performed on the mono-temporal aerial images for the task of semantic segmentation. The framework adapted here for segmenting defects included: A Res-Unet framework and a corresponding novel loss

function: Dice loss [36]. This reliable framework can perform semantic segmentation resulting in high resolution images. In order to avoid the overfitting, the Res-Unet relied on the data augmentation strategy as well. Each image has been rotated to the angle, zoom in/out, flip and so on to enlarge the datasets of Res-U-net. In Figure 3.13, a segmented sample from Res-Unet (b) and the corresponding raw images(a) are shown.

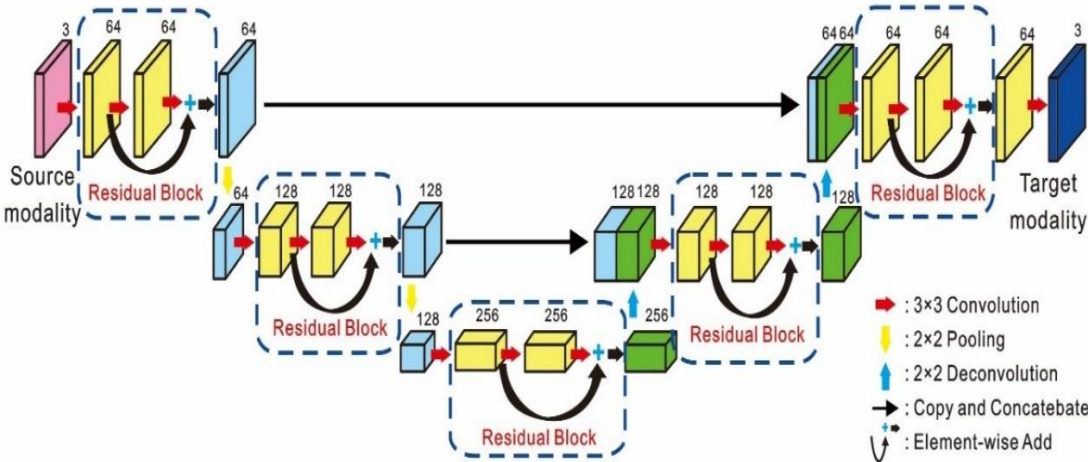


Figure 3. 12 Res-U-net model structure

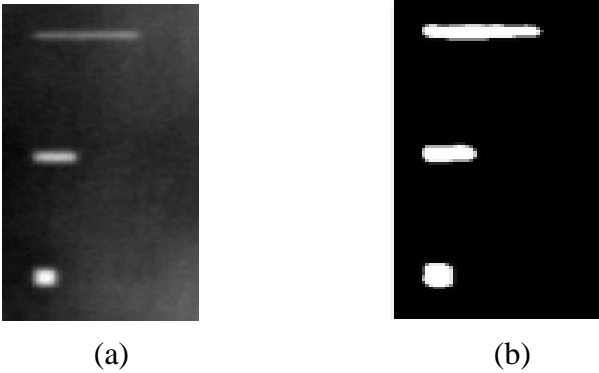


Figure 3. 13 An example of method 4 (a) the original thermal image; (b) the segmented image

D. Instance defect segmentation algorithm

- Method 5: MASK-RCNN for defect segmentation

The Mask-RCNN detection procedure can be considered as either an object detection function or object segmentation function. Compared with the semantic segmentation, the instance segmentation associates each pixel of an image with an instance label. It can forecast a whole segmentation mask for each of those objects and predict which pixels in the input image correspond to each object instance. It also reduces the restriction to the position of defects rather than predicting a group of bounding boxes for the defects. Mask R-CNN is a classical instance segmentation method extended intuitively from Faster-RCNN, which is an end to end trainable model to achieve pixel-to-pixel alignment segmentation between inputs and outputs of a convolutional backbone architecture. ROI Align preserves spatial orientation of features with no loss of data for extraction over the entire image of the network. This approach efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance.

Each thermal image was fed into the backbone convolutional network from Mask R-CNN, once some learned region proposal was obtained from the backbone network. These features projected learned region proposals onto convolutional feature maps. Mask-RCNN using ROI aligning [37] to warp our feature from the convolutional feature map into the right shape then outputs into two different branches. As shown in Figure 3.14, there are two different branches providing an output of predicted results. The top branch (blue line box) is a classification score of categories of region proposals and a bounding box for regression of coordinates in the output. In addition, at the bottom (red line box), a segmentation mask is predicted by the model for each of those region proposals to classify for each pixel in that input region proposal whether or not it is an object. Figure 3.15 provides an example of an original image from pulsed thermography (a) and a segmented image from Mask RCNN (b).

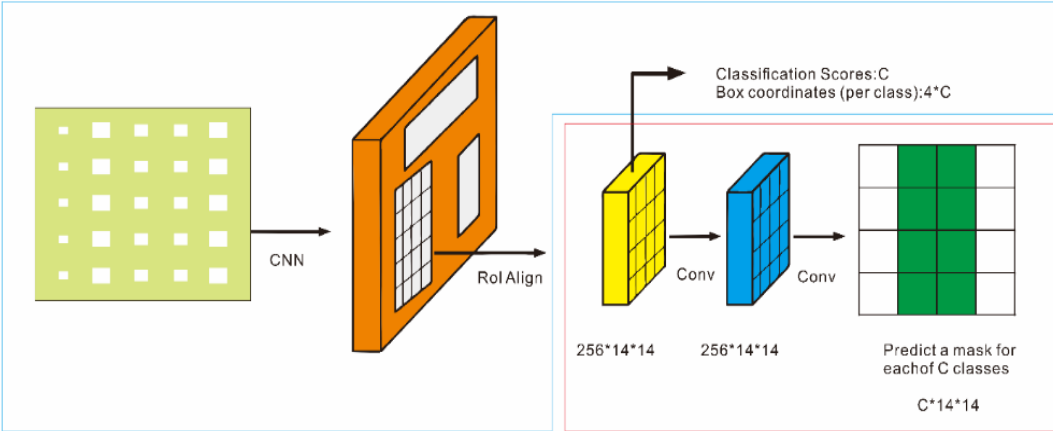


Figure 3. 14 Mask- RCNN Processing Architecture

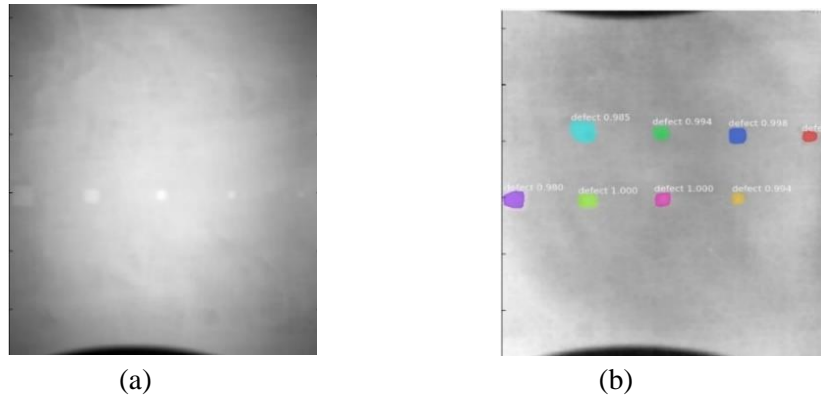


Figure 3. 15 An example of method 5 (a) the original thermal image; (b) the detected image

- Method 6: Central-Mask for defect segmentation

Since the Mask-RCNN relies on the pre-defined anchors, it influenced slowed down for the speed and accuracy in detection. Central-Mask is a simple yet efficient real time anchor-free instance segmentation. Based on the structure, Central-Mask could be regarded as a novel spatial attention-guided mask (SAG-Mask) branch adding on free anchor one stage object detector (FCOS) [38]. A segmentation mask head located on each detected box with the spatial attention map that helps to aim attention at informative pixels and suppress noise. Figure 3.16 show the overview architecture of Center Mask. A feature pyramid extractor combines with the FCOS box head to predict classification scores and bounding box regression. A spatial attention-guided mask (SAG-MASK) predicts the segmentation map for the defects based on a spatial attention module [39] from each bounding box, which focuses on meaningful pixels and eliminates the noised influence. Central Mask achieves a faster speed and surprising accuracy better than other state of the art instance segmentation approaches (Mask-RCNN). In this work, we adapted the Central-Mask network for feature extraction and defect segmentation. The main goal is to precisely detect and analyse defect information from the thermal images. The core strategy from this network is to extract the meaningful thermal pattern from the sequence for each specific defect. Figure 3.17 show a raw thermal image(a) and a corresponding segmented thermal image (b) from Center-Mask. Each defect is precisely localized and segmented by the Mask.

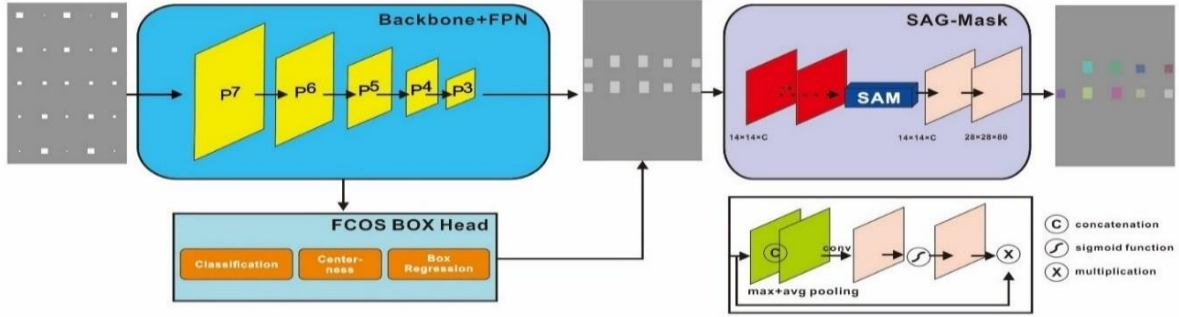


Figure 3.16 The structure of Center-Mask

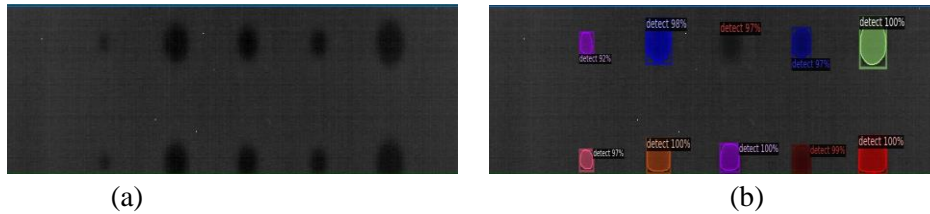


Figure 3.17 An example of method 5 (a) the original thermal image; (b) the detected image

### E. Regular infrared defect detection algorithm

- Method 7: Absolute thermal contrast (ATC) with global threshold (GT)

In combination with a global threshold method (GT), the ATC was adapted for the procedure of segmenting defect areas. The vital concept of this method was to compare the grey level of the pixel in the image coordinated  $(x, y)$  with the average grey level of a sound region of the sample and it is often adapted in infrared image processing. Equation (3.3) describes how this method works: where  $T_{atc}$  is the grey level in the ATC image in the coordinate  $(x, y)$  of the ATC image.  $T_d(x, y)$  is the average grey level of the group pixels in the defect region and  $T_s(x, y)$  is the average temperature of a nearby sound region.

$$T_{atc} = T_d(x, y) - T_s(x, y) \quad (3.3)$$



Figure 3.18 An instance of method 7 applied on the thermal image (a) the original thermal image; (b) the detected image



Figure 3.18 provides an example of the segmentation with this method: (a) the original thermal image from pulsed thermography; (b) the corresponding segmented image in method 7. This method made it possible to reduce the effect from non-uniform heating and remove some thermal pattern noises.

### 3.9 Experimental results and implementation details

#### 3.9.1. Training

The training procedure for deep learning models was set according to the following principles for different neural network architecture parameters adjusted based on the Pytorch framework. The training processing was conducted on a GeForce GTX1080TI about 30 min. The operating system is set as: Ubuntu 16.04. The framework of the learning model is set as: Darknet. CPU: i7-7700k. Memory: 16GB, GPU: NVIDIA GeForce GTX1080TI. For each modeling training procedure and hyperparameters setting, we configured the parameters as shown below in Table 3.2.

Table 3. 2 The training detailed from each DL model

Deep learning model	Training detailed information
U-net	1. A high learning momentum (0.99); 2. The weight decay is set as - 0.005; 3. A learning rate is 0.005; 4. The number of epochs is 200; 5. The input sequences and the corresponding segmentation images are used to train with U-net with min-batch gradient descent implantation from Pytorch deep learning framework.
Res-Unet	1. The Res-Unet in this work was training based on the MXNET[25] deep learning library; 2. A batch size of 256 with mammal gradient aggregation [26]; 3. The number of epochs is 300 ; 4. The weight decay is set as - 0.00066; 5. Adam optimizer with initial rate 0.005; 6. A multi-dimension learning momentum [27]; $(\beta_1, \beta_2)$ is (0.9,0.09) and the initial learning rate is 0.0005.
YOLO-V3	1.The Momentum was implemented as optimizer during the training; 2. The learning momentum was 0.9 ;3. the learning rate was set as 0.001; 4. The backbone is adapted Darknet51; 5. The weight decay was 0.0005; 6. The maximum batches size of iteration was 50200.
Faster-RCNN	Batch size 256; Overlap threshold for ROI 0.5; Learning Rate 0.001; Momentum for SGD 0.9; Weight decay for regularization 0.0001

Mask-RCNN	1. Network training used Resnet50 as backbone; 2. The mini mask size is 28x28; 3. The weight decay is set as - 0.0001; 4. The loss weight is equal for each class and mask ( RPN class, RPN bounding box, MRCNN class, MRCNN bounding box and MRCNN mask) ; 5. The learning momentum is 0.9 and learning rate is 0.0003; 6. Training of the first 20 epochs of network heads, was followed by the training of all network layers for 80 epochs, the model weight.
Center Mask	Stochastic Gradient Descent (SGD) for 90K iterations (200 epoch) with a mini-batch of 2 images and initial learning rate of 0.01; a weight decay of 0.9 and a momentum of 0.01, respectively. All backbone models are initialized by ImageNet pre-trained weights.

### 3.9.2 Evaluation metrics

F-score and the probability of detection [40] are introduced to analysis the capability of detection of each detection deep learning model, which is being interpreted by equations (3.4)-(3.7). The precision means the ratio from the cases that actually contain the defects over the cases that are recognized by the system that contains the defects, which represent how accurate the system is in identifying the defects. The recall means the system correctly recognized the defects over the cases that actually contained the defects. The precision and recall values heavily depend on the confidences scores that the system is setting. The F-score is a method to estimate the detection and segmentation capability from these algorithms.  $\beta$  is a value to represent the weight between the Precision and Recall value. In this work, the Recall is a metric which is more influential in evaluating the performance. Therefore,  $\beta$  is equal to 2. The POD reveals the accuracy of the method to detect the defects, which are always calculated at a specific confidence score value. Although the POD keeps the same mathematical format as the Recall in the equation, POD represents a further explanation in quantifying research with NDT inspectors. In this work, we set the threshold for CTS at 75% for POD metric.

$$Precision = \frac{TP}{TP+FP} \quad (3.4)$$

$$Recall = \frac{TP}{TP+FN} \quad (3.5)$$

$$POD = \frac{TP}{TP+FN} \quad (3.6)$$

$$F \text{ score} = (\beta^2 + 1) \frac{Precision \times Recall}{(\beta^2 \times Precision) + Recall} \quad (3.7)$$

Where TP is true positive, FN is the false negative representing the number of the defects which have not been detected. Meanwhile, FP is the false positive defects representing the defects which are wrongly detected as defects, when they are in fact not defects.

Moreover, the confidence threshold score (CTS) has been defined as a standard for measuring the accuracy of detecting corresponding objects in a given data set. CTS is a simple measurement standard that can be used for any task that yields a prediction range (bounding boxes, segmented maps) in the output regarding the ground truth.

### 3.9.3 Learning curves

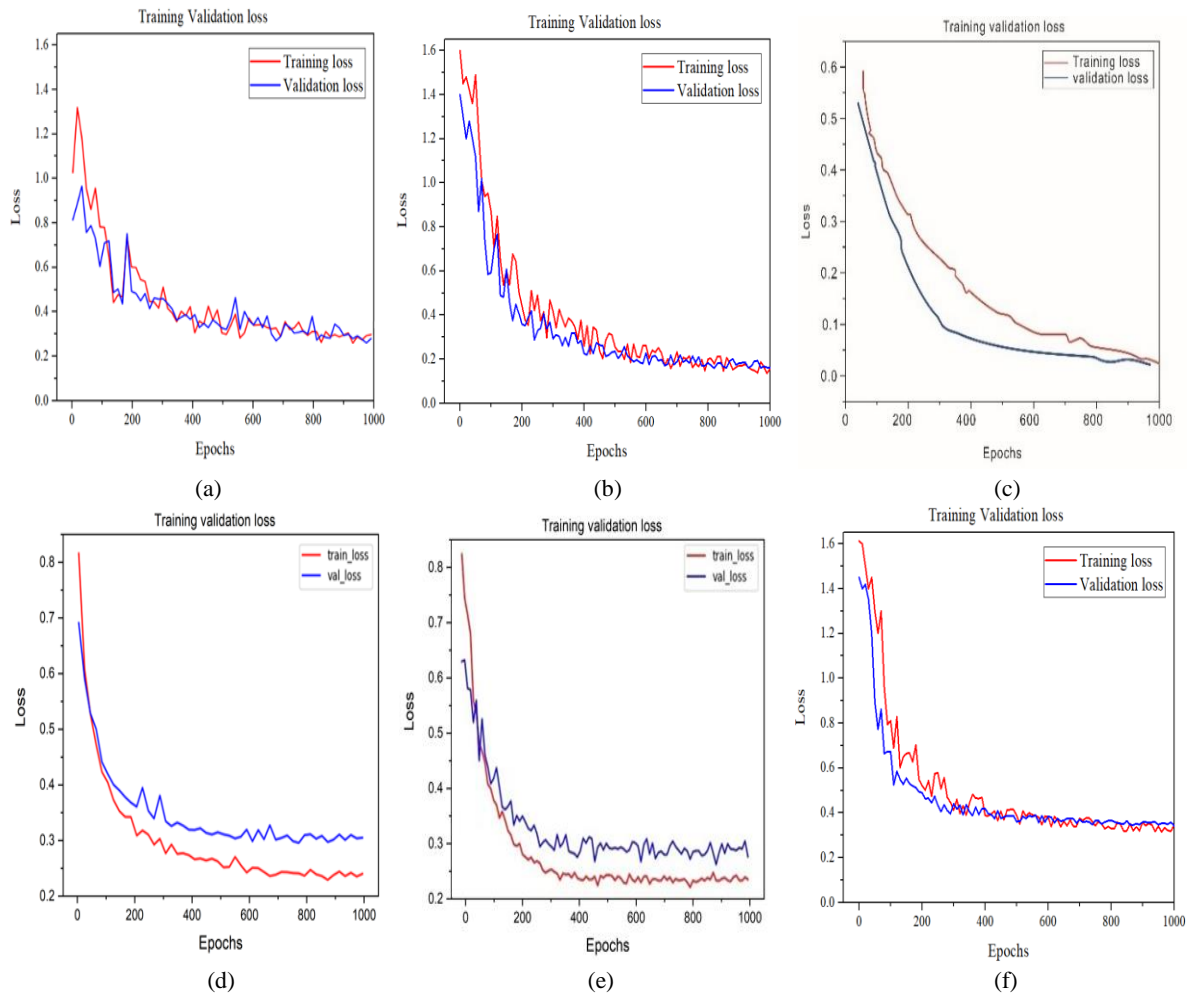


Figure 3. 19 The loss curves in each deep segmentation models (a) Mask-RCNN; (b)Center-Mask ;(c) YOLO-V3 (d) Res-U-net; (e) U-net (f) Faster-RCNN

In Figure 3.19 (a)-(f), each deep learning model was trained for 1000 epochs respectively. Figure 3 .19 (a) shows the average loss curve for training and validation process for the Mask-

CNN model. The Training loss curves decreased while the number of iterations increased. It can be seen that the loss dramatically decreases during the first 200 batches, then gradually flattens out around 0.225 as the batch number of the iterations increases and then remains steady. The validation loss coverage involves a similar loss. This indicates that the performance of Mask-RCNN was promising during the training procedure. In Figure 3.19 (b), the loss curves of the Center-Mask model have a similar momentum to that of the Mask-CNN loss but more smoothly. The training loss stably decreased as well while the whole number of iterations increased, and then converged around 0.341.

In comparison with Figure 3.19(a)(b), the four other DL models in Figure 3.19(c)-(f) seem to maintain a similar momentum. The average curve of the training loss became more dramatic and oscillating decreases in the first 500 epochs and then it flattens out late. The loss curves of steel stabilized at a value lower than 0.5 after 500 epochs. As a result, based on the obtained model, the loss from the 6 different kinds of deep learning models, this furtherly indicated an impressive performance during the whole procedure (training and validation) when it is applied on defect segmentation and localization of composite materials.

#### 3.9.4 Detection results

This model provided the shape and location of each defect detection results based on the labeled images with ground truth.

In table 3.3(f), the noise of the input image is the main factor affecting the segmentation results. As indicated in the U-net result, the segmented image is not clear. The segmentation boundary is still blurry. A preprocessed image from principal component analysis (PCA) was added in the validation database to verify whether the segmentation effect will be better after denoising in the Res-U-net model training. From the results, it seems the performance has been improved to some extent and the test result of Resnet-U-net gave a better performance than the original U-net.

Table 3.3 also shows the visualized results from 6 deep learning algorithms. Specifically, in sample (g), the defect feature from the sample indicated clearly, these deep learning methods show excellent defect detection capability. However, it is obvious that the comparison methods (semantic segmentation) have a substantial shortcoming. These method results are affected by the non-defect area in sample(g), whereas the Resnet-U-net can be conducted without false detection. Compared with Resnet-U-net, the original U-net are more sensitive to fix patterned- noise and non-uniform heating from thermography due to the higher false detection rates in the result of sample (g). Therefore, U-net cannot detect the specific thermal

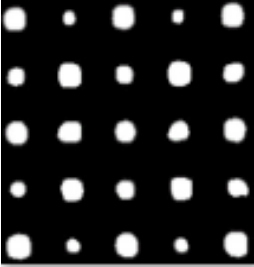
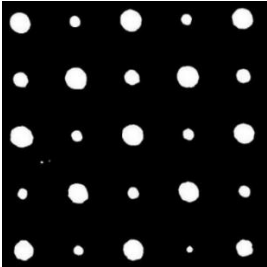

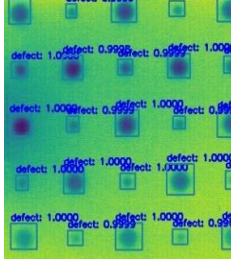
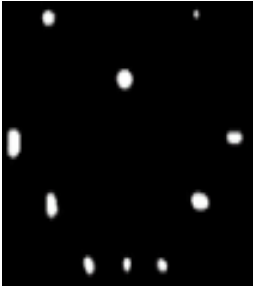
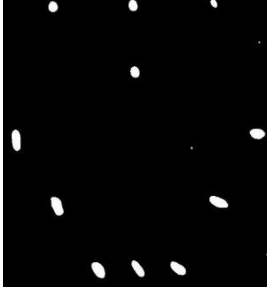
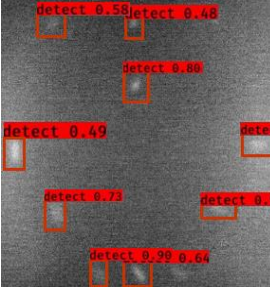
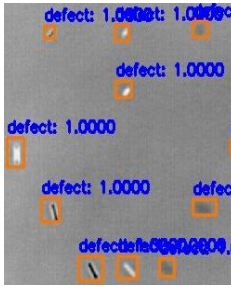
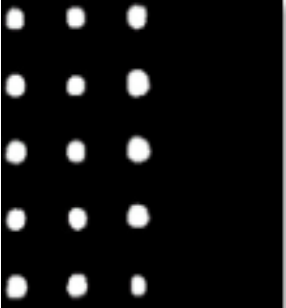
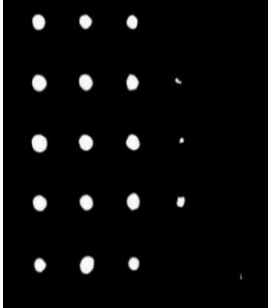
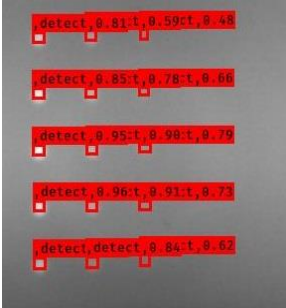
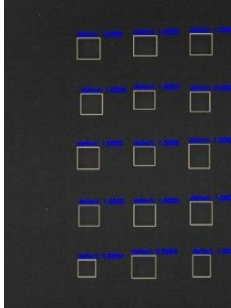
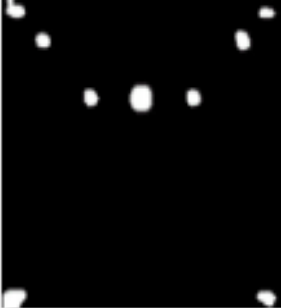
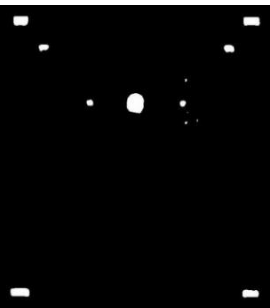
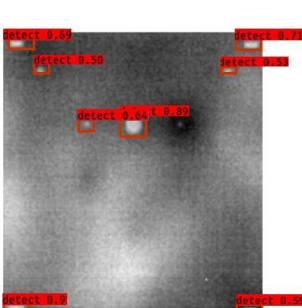
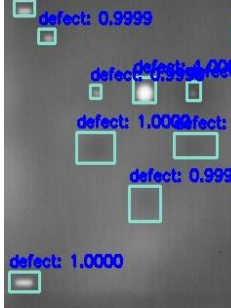
data very well because U-net is too insensitive to defect information. Note that the introduced model Res-U-net can ensure correct detection while effectively prohibiting noise interference.

In terms of sample(d)(Steel) and sample(e)(Plexiglass), it is still quite challenging to detect the defects and abnormal areas because the background and noisy information represent a high percentage around the defect information region in the sample. The result of sample (d) indicated that Faster-RCNN failed in detecting the less visible defects. For the model of Faster-RCNN, although it introduces a hierarchical structure of deep architecture to extract semantic information in the images, there is still failure to distinguish the boundary noise information from the Steel sample. On the other hand, YOLO-V3 is slightly more effective in comparison with Faster-RCNN based on the detected results on 8 evaluation samples (mAP =0.75 IOU metric). This further illustrates the introduced model YOLO-V3 leads to good identification accuracy as a single stage detector in comparison with the other state of the art methods.

For the instance segmentation method, the segmented images (Center-Mask; Mask-RCNN) show some indistinguishable results from the ground truth. Several types of defects are detected which include the shapes of circle, square, rectangle. Table 3.4 shows the detection results of the defects by training using the instance segmentation model: Mask-RCNN /Center-Mask model. Particularly, since the training database is composed of regular shapes and permanent angles with circle and square shapes, the testing results in Table 4 show that the irregularly distributed defects with multiple angles are detected accurately, which indicated that the Mask-RCNN/Center-Mask spatial detection model can enhance detection performance based on instance segmentation of pixel to pixel alignment.

As a result, it is not enough to only acknowledge the semantic information, but more impactful to know how to obtain the low semantic information from defects under the interference of objective noisy conditions. In contrast, for the instance segmentation models, not only did this illustrate a better segmentation performance for the Plexiglass samples, but also it has excellent detection capability for the Steel and CFRP sample.

Table 3. 3 Results with semantic segmentation and object localization algorithms

	Res-U-net	U-net	Faster-RCNN	Yolo-v3
(a)				
(b)				
(c)				
(d)				

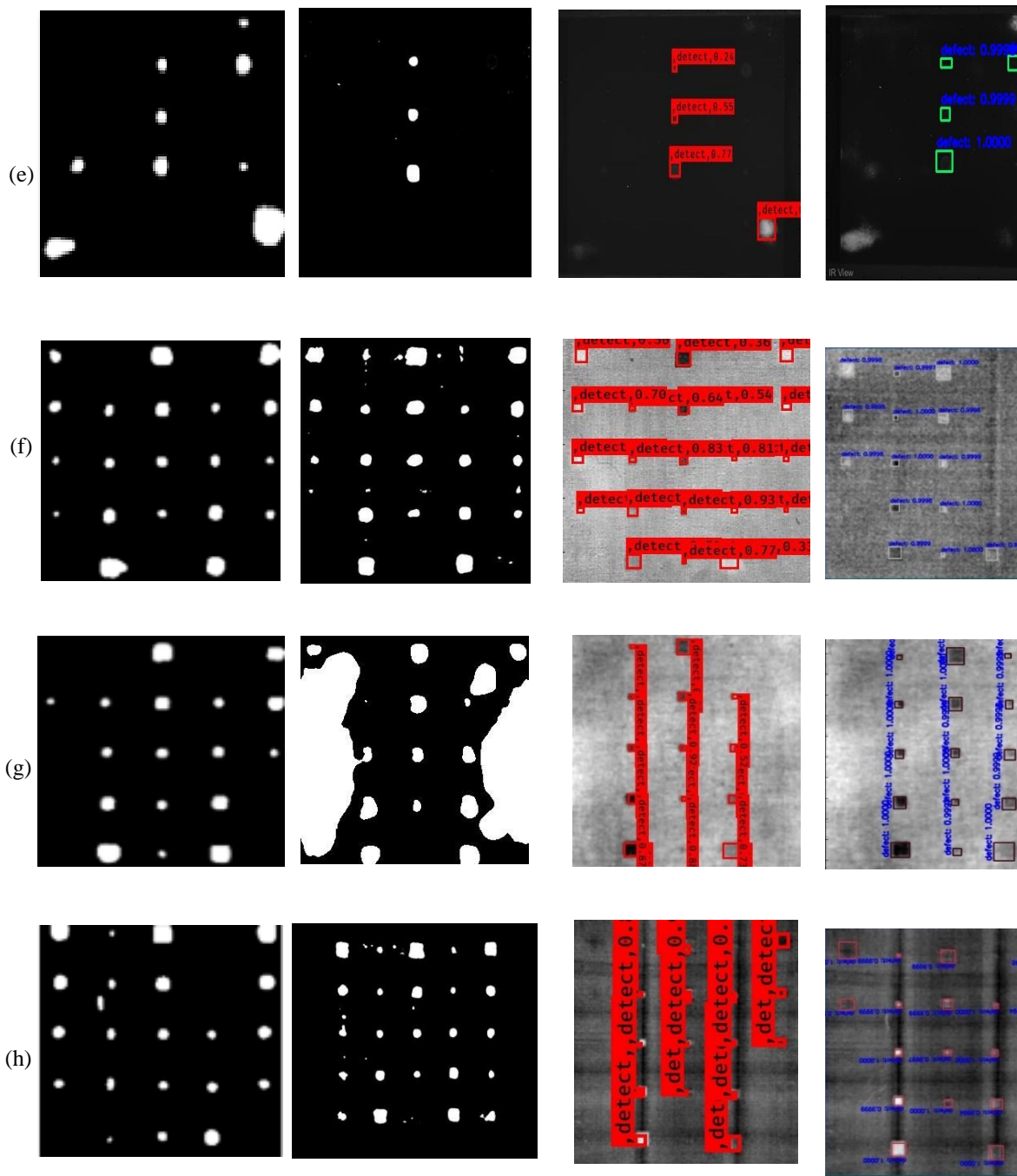
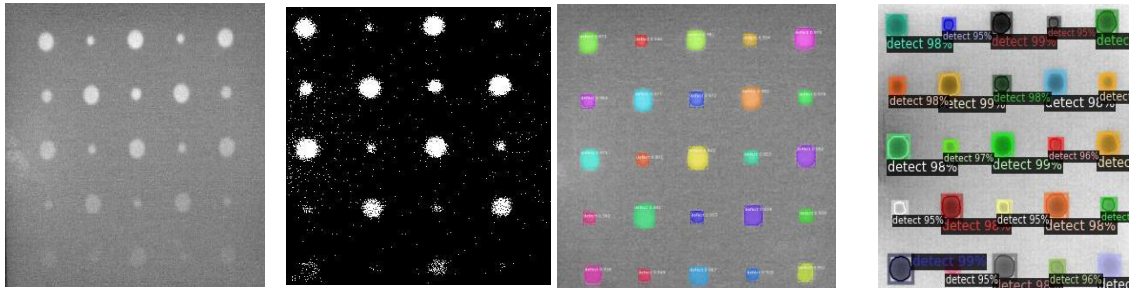


Table 3. 4 Segmentation results with instance segmented algorithms; Absolute Thermal Contrast (ATC); Raw

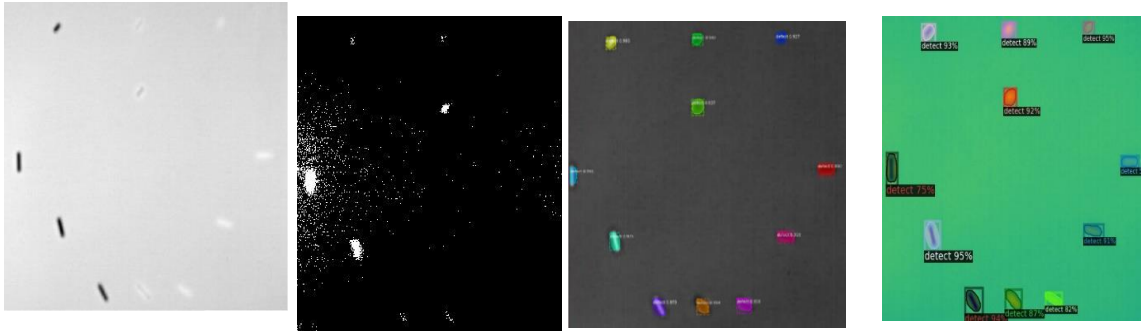
Raw	ATC	Mask-RCNN	Center-Mask
-----	-----	-----------	-------------



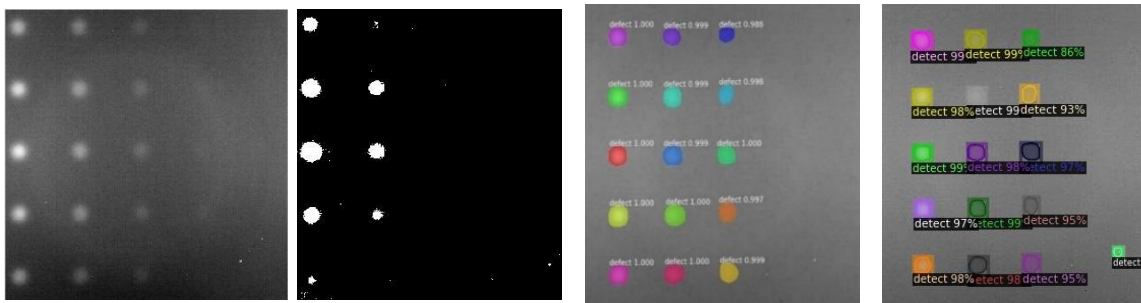
(a)



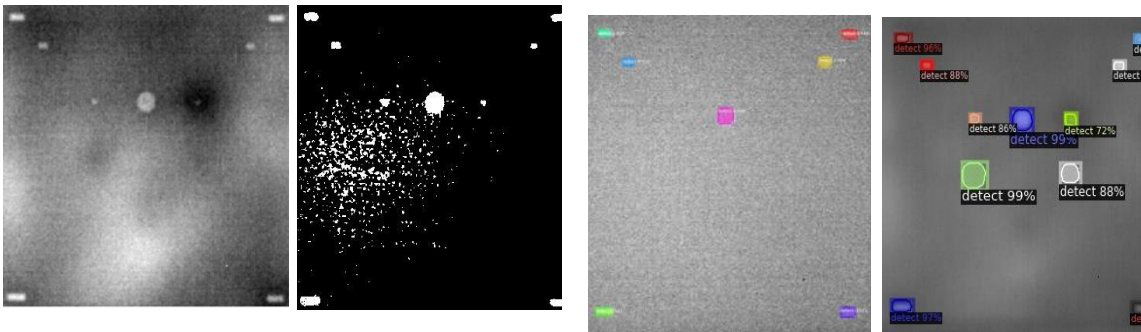
(b)



(c)

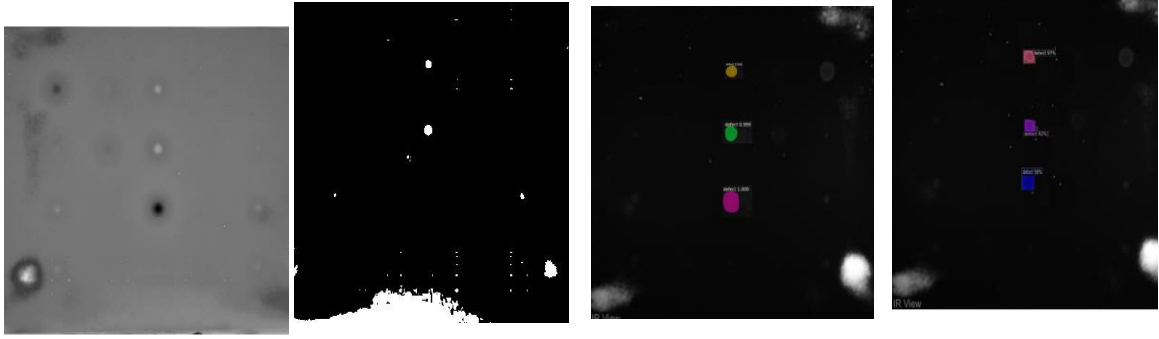


(d)

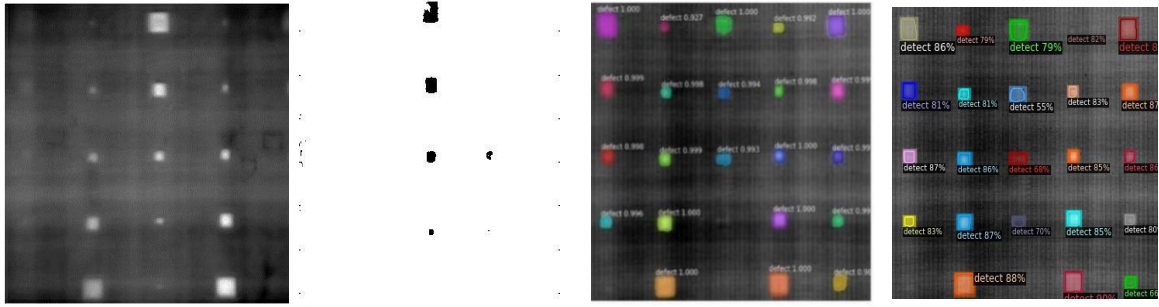




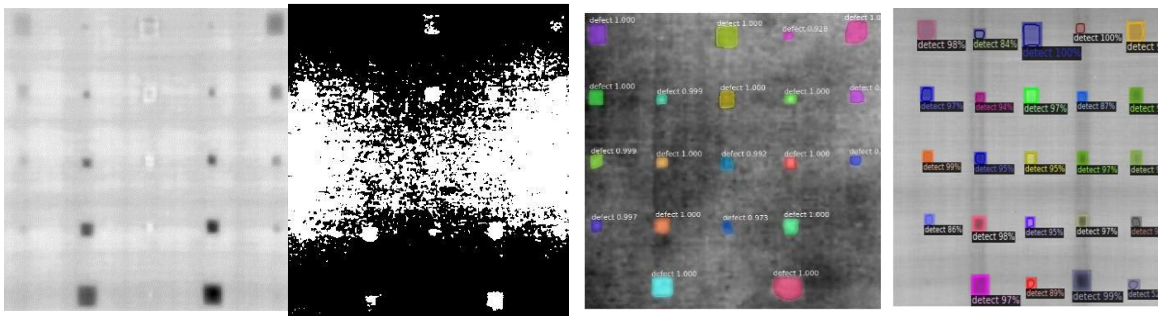
(e)



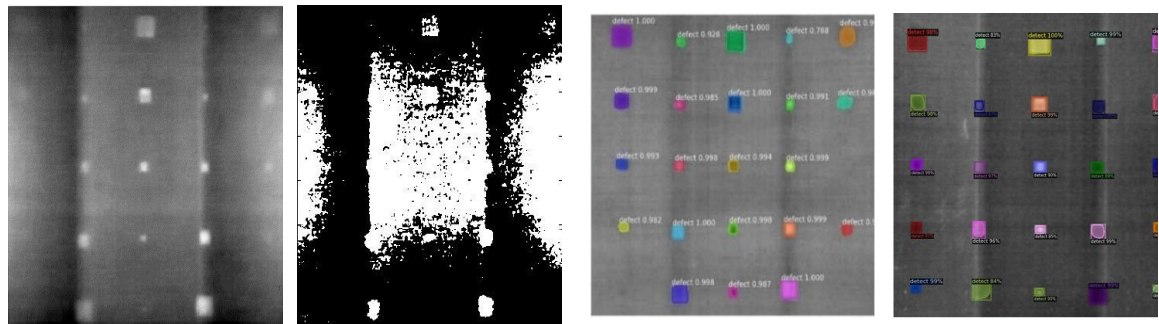
(f)



(g)



(h)



To further evaluate the robustness of the learning model in comparison with the state of the art: ATC, samples (a) to (h) were further adapted to carry out a detailed analysis. For instance, in samples (a)-(e) of plexiglass and steel, in comparison with the introduced method from instance and semantic segmentation, the poorer detection of the comparison methods (ATC) is obvious since the semantic information is unclear and the segmented defects are not obvious. Then, for samples (f) to (h) from CFRP which were limited by the accuracy of IR camera, the detection results of Absolute thermal contrast (ATC) are still disappointing, whereas for the DL methods the results are far superior to ATC. Therefore, for the regular and irregular shaped specimens, the overall performance of the DL methods is significantly better than all the state-of-the-art methods (ATC and global threshold).

### 3.9.5 Reliability assessment using probability of detection (POD)

The reliability assessment metric of subsurface defects detection- Probability of detection [29] has been further assessed to quantify the performance of these six DL models in this task which can be expressed as a function of aspect ratio through a POD curve. Each curve has been plotted in Figure 3.20 with respect to the aspect ratio (*size/depth*) to indicate the quantitative analysis for various sizes and depths of defects with deep learning models. The results of probability of detection (POD) are based on the defect regions detected from the deep neural network methods based on the referenced ground truth.

Figure 3.20 (CTS = 0.75) indicates the final POD scores obtained from all the samples (a)-(h) for each DL algorithm. The POD of the instance segmentation method Center-Mask based approach has a notable performance and the highest POD scores in comparison to other approaches which represent the highest detectability. Then the staged objective localization methods (YOLO-V3; Fast-RCNN) have a faster and medium detection accuracy. The semantic segmentation method (U-net; Res-U-net) obtained less accuracy due to the fixed pattern noise and non-uniform heating from the infrared thermal data. However, all of these six DL models surpass the state-of-the-art method (Absolute thermal contrast: ATC) and function automatically.

The results from different samples for POD validation of methods are indicated in Table 5 (CTS = 0.75). The results are compared using the thermal frames acquired from the Pulsed thermography (PT) on each Plexiglass/CFRP/Steel sample in Section 2.3. In Table 3.5, the instance segmentation method (Center-Mask) still shows an acceptable segmented result for all samples provided. Since the instance segmentation method can capture feature differences over each pixel, the pretrained model (Center Mask; Mask-RCNN) has the intrinsic capability to segment defects from background information based on a learning and labeling process. For the semantic segmentation method, as discussed previously, the original U-net model

fails to detect the deeper defects region, such as some defect regions of sample (c), since the spatial features from the infrared sequences in defect regions are difficult to obtain due to the non-uniform heating and fixed pattern noise from the boundary region in this case. In contrast, it is worth noting that the introduced method from the instance segmentation model (Center-Mask) obtained the best performance in sample (c) in comparison with the other methods.

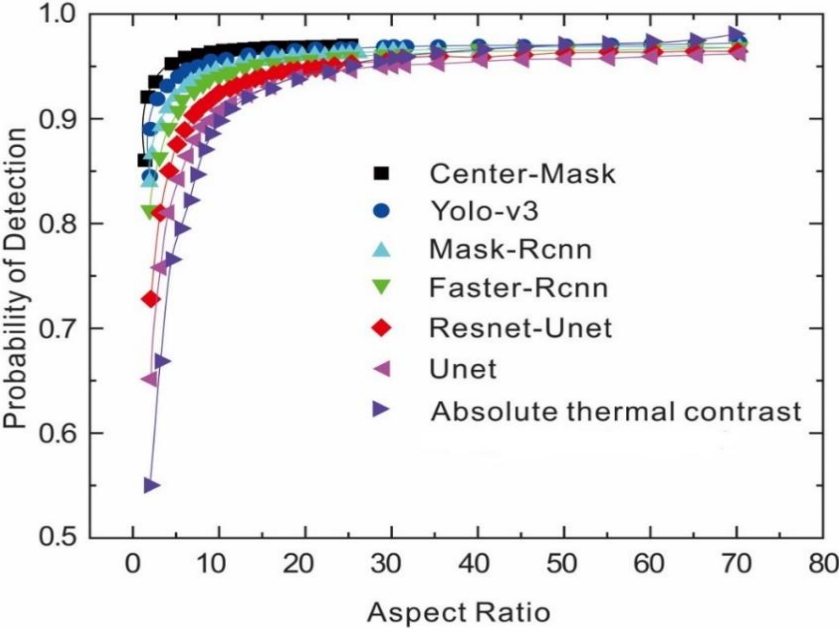


Figure 3. 20 Different detection methods with CFRP sample

The results from samples (a)-(h) analysed by the POD evaluation metrics are indicated in Table 3.5. The performance is compared with thermal images and sequences collected from pulsed thermography experiments on each specimen. It is clear that on samples (a)–(d), the YOLO-V3 and Center-Mask show commensurable results, while the other four algorithms (U-net; Res-U-net; Mask-RCNN; Faster-RCNN; ATC) obtain less accurate performance. Due to the hierarchical structure of deep architecture in neural networks, each DL model has a strong capability to separate low-degree background information from raw thermal sequences (fix-pattern noise; non-uniform heating) and extract defects feature components. However, it is worth mentioning that the semantic method(U-net) model still fails to detect when it encounters a challenging situation (non-uniform heating), just like sample (e), while the instance segmentation (Center-Mask) maintains high performance in POD evaluation.

Table 3. 5 Different detection methods of results with individual sample

Methods	POD of different samples							Sa
	Sample(a)	Sample(b)	Sample(c)	Sample(d)	Sample(e)	Sample(f)	Sample(g)	mpl e(h)
Mask-RCNN	0.89	0.91	0.87	0.84	0.74	0.89	0.95	1
Center-Mask	1	0.94	0.92	0.84	0.82	0.91	0.90	1
U-net	0.85	0.84	0.83	0.81	0.71	0.86	0.83	0.86
Res-U-net	0.89	0.90	0.90	0.84	0.76	0.85	0.81	0.92
Faster-RCNN	0.90	0.93	0.92	0.93	0.92	0.95	1	0.89
YOLO-V3	0.92	0.94	0.94	0.95	0.95	0.96	0.92	0.89
Absolute thermal contrast with global threshold	0.65	0.71	0.73	0.78	0.81	0.73	0.67	0.65

Furthermore, the results validated on the total databases from 8 representative samples (a)-(h) are indicated in Table 3.6. These results in Table 3.6 indicated that the Center-Mask learning model achieves the best performance comparable to that of the other bench mark detection methods (semantic segmentation; defect localisation). This network may be able to obtain the relatively stable performance due to the fact that it involves a deep architecture to extract the features driven by the tasks and instance masks.

Table 3.6 indicated the Precision, Recall and F-score analysis result from each algorithm (CTS = 0.75). The Pr represents the Precision value and the Re represents the Recall value. The F-score of comparison on specimen (a) to (c) is relatively high for each proposed DL model. However, specifically for sample(g), the F-score of Unet is 57.2%, which indicates that Unet only partially detects defects due to noise influences. Whereas the instance segmentation method (Mask-RCNN; Center-RCNN) can reach 76.5%, 83% in sample(g). From sample (h), the F-score of semantic segmentation (U-net, Res-Unet) can only achieve

71.8%, 77% in several, and the remaining two comparison methods give the following results: Center-Mask (86%) and Yolo-v3 (79%).

Table 3. 6 Total detection results with different deep learning segmentation algorithms

Samples	Evaluations	Methods						
		Mask-RCNN	U-net	Res-Unet	Faster-RCNN	Yolo-v3	Center-Mask	ATC
A	Precision	0.46	0.45	0.45	0.47	0.40	0.45	0.30
	Recall	1	0.81	0.89	0.90	1	1	0.75
	F-scores	80%	70%	74%	76%	76%	80%	57.6%
B	Precision	0.41	0.50	0.46	0.43	0.52	0.55	0.25
	Recall	0.91	0.84	0.87	0.88	0.92	0.94	0.69
	F-scores	73.2%	74%	73.8%	72%	80%	82%	51%
C	Precision	0.45	0.44	0.47	0.49	0.57	0.59	0.28
	Recall	0.87	0.85	0.83	0.84	0.91	0.92	0.63
	F-scores	83%	72%	71.9%	73%	81%	82%	50%
D	Precision	0.46	0.47	0.49	0.40	0.59	0.60	0.22
	Recall	0.84	0.81	0.83	0.83	0.90	0.94	0.78
	F-scores	73.3%	71.6%	72.8%	72%	81.4%	84.4%	52%
E	Precision	0.41	0.40	0.50	0.52	0.60	0.66	0.38
	Recall	0.80	0.75	0.79	0.80	0.82	0.82	0.66
	F-scores	67.2%	64%	70%	72%	76%	78%	57%
F	Precision	0.42	0.35	0.46	0.41	0.64	0.67	0.38
	Recall	0.91	0.80	0.85	0.85	0.82	0.91	0.61
	F-scores	73.7%	57.2%	74.9%	70%	78%	85%	54%
G	Precision	0.49	0.42	0.60	0.41	0.65	0.61	0.46
	Recall	0.89	0.63	0.79	0.90	0.94	0.92	0.69
	F-scores	76.5%	57.2%	74.9%	73%	87%	83%	62%
H	Precision	0.42	0.59	0.55	0.42	0.54	0.64	0.31
	Recall	0.87	0.76	0.85	0.89	0.89	0.95	0.75
	F-scores	71.64%	71.8%	77%	73%	79%	86%	58%
<b>Average</b>	F-scores	74.8%	67.25%	70.66%	72.62%	79.8%	82.55%	55.2%

\* The precision and recall values always tend to be negatively correlated in the evaluation of DL models, in this project we placed more emphasis on the recall values and comprehensive F-scores to assess the number of defects which successfully detected and model performance.

The average F-scores for all eight specimens, involving the comparison methods from 6 different DL algorithms (Faster-RCNN, YOLO-V3, U-net, Res-UNet, Mask-RCNN, Center-Mask) provide 72.62%, 79.8%, 67.25%, 73.66%, 74.8% and 82.55% defect detection capability respectively. The instance segmentation method (Center-Mask) gives the highest capability for detection on average. As a result, the instance segmentation method is relatively better than the semantic segmentation methods in terms of detection ability due to noise influence from thermography in this circumstance.

In this experiment, the samples from three types of materials are divided into different geometric distributions (regular and irregular shape defects). Due to the different geometric architecture of the regular and the irregular defects, this leads to inconsistent thermal diffusion. Correspondingly, this inconsistent thermal diffusion causes the data distribution to be quite different. It can be concluded based on the comparison of the algorithms that it can be difficult for a single DL model to detect all defects effectively. Thus, we adapt the 3 types of deep learning algorithms separately training for the different type of samples and also compare the results with other state of the art methods.

### 3.9.6 Mean average precision (mAP)

Average Precision (AP) [41] is also an indicator that is analysed for the relationship between the precision and recall values. In this section, the mean average precision (mAP) metric is introduced to further analyze the top 4 detection models ranking in POD analysis in Section 6.5 (Center Mask; Mask-RCNN; YOLO-V3; Faster-RCNN).

The average precision (AP) is calculated based on the indicated bounding boxes and the different confidence thresholding scores from the DL models in object localization and instance segmentation. The total AP and the precision-recall plots obtained when 4 different deep learning architectures are adapted are compared in Fig. 21 (including all the confidence thresholding score values). The detection results show that the Center-Mask still displays the best performance. Taking the 1000 infrared images from the 8 eight representative specimens, defects can be detected and the Recall and Precision can be calculated by the confidence scores, as shown in Figure 3.21 below. The AP of all 1000 thermal images is 75.05%. For instance, in instance segmentation model, the AP is calculated as 75.33% and 81.06% respectively, for Mask-RCNN and Center-Mask (Fig. 21 (a) and (b)). Therefore, the detection performance is better in Center-Mask than in Mask-RCNN. For the object localisation model, the AP is calculated as 76.63% and 71.06%, for YOLO-V3 and Faster-RCNN respectively (Fig. 21 (c) and (d)). Therefore, the detection performance is better in YOLO-V3 than in Faster-RCNN.

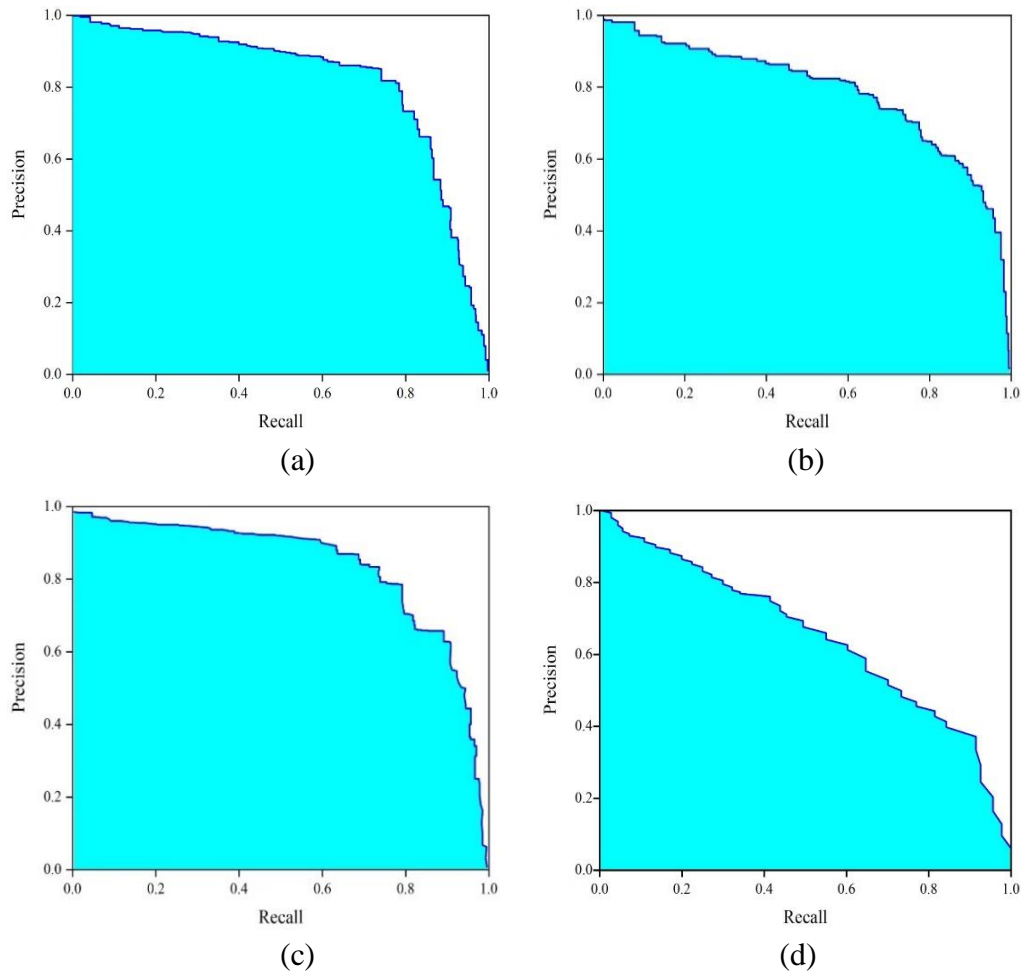


Figure 3. 21 The mean average precision curve from each deep segmentation model: (a) Mask-RCNN; (b)Center-Mask; (c) YOLO-V3; (d) Faster-RCNN

### 3.9.7 Running time complexity

Further, frames per second (fps) [36] was introduced as an idea to certify how many images can be processed in a unit (1s) time by each deep learning model in order to analyse the running time complexity for the model. In Figure 3.22, the running time complexity of each model has been indicated which illustrates the average time to detect or predict a defect in each frame(picture) from DL model: The higher the value in the graph, the faster speed the DL algorithm has.

Based on the analysis of Figure 3.22, it can be seen that the objective localization approach significantly achieved the fastest speed among the models. For the instance segmentation detector, Center-Mask has increased the time per frame from the state-of-the-art method ATC: 0.5 fps to 12 fps. Then Mask-RCNN also achieved a time per frame of 5fps which increased

the processing speed significantly in comparison with the regular thermal threshold segmentation method (ATC) in thermography.

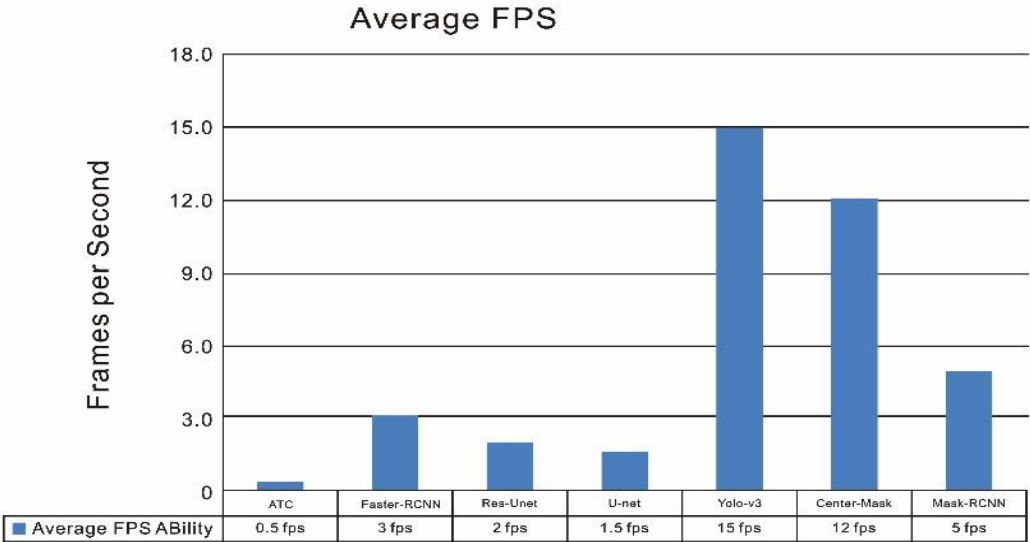


Figure 3. 22 Average frame per second for each deep learning model

Moreover, we further analyzed the results in comparison with other state of the art networks and YOLO-V3 still obtained the fastest running time speed to process images due to the reason that it is a one-stage real-time detector and has a much faster speed than other detectors (such as Mask-RCNN; Faster-RCNN). The RCNN methods are relatively slow since these models are two-stage procedures (Region Proposal Network (RPN); ROI pooling). Whereas, as indicated previously the POD curves in Figure 3.20 (Section 6.5), Center-Mask still achieved the highest POD scores during the whole validation process based on the different aspect ratio values (size/depth). Therefore, in this work, Center-Mask is the most promising to obtain the highest accuracy but YOLO-V3 is the most efficient which has the faster time frame speed.

**3.10. Results analysis and discussion**

The deep segmentation models gave attractive results for the Plexiglass/CFRP/Steel materials defects identification evaluation. This project focused on building and fine-tuning on the training parameters for those particular defects. To improve the accuracy of the detection model, the manner in which the dataset is built has a significant impact.

According to the results obtained, the following analyses and points of this experiment have been concluded below:



1. To implement a robust detection model, the databases must include enough samples. One way to effectively improve is to increase the size of the dataset by including the multiscale images. A database composed of images on different scales (larger or smaller), enables the training to be sensitive to those new dimensions. This would increase the robustness of the deep segmentation algorithms facing larger defects, as well as improve the results on blurry pictures. To help reduce false alarms in the algorithm results and be more convenient for the user, implementing different types of labels is necessary. In the case of this particular project, each section was labeled with a defect in the spatial segmentation training (Mask-RCNN; U-net; Res-U-net). The proposal is to add different classifications. For example, including the name of the shape of the defect: circle, triangle or some false positive cases (lighting spots, scratches) would be beneficial. This would allow the algorithm to not detect these shapes as a defect, and thus, reduce the number of false alarms.
2. Another critical point in this experiment to be considered is the marking process. In comparison to other objective detection methods, Mask-RCNN/Center-Mask especially involves a pixel-based marking approach which could mark the defects accurately, as opposed to marking a considerable area around each defect. It can rapidly and easily annotate the object without the bounding boxes restrictions in most cases. In comparison with an instance segmentation method, U-net and Res-Unet are the auto-encoder format DL models that can be trained based on each pixel level to semantically segment defect pixels from sound pixels. However due to the burden of tackling massive temporal data of thermal frames, U-net and Res-Unet have less time efficiency and high time complexity on the thermal data in comparison to the instance segmentation model. Therefore, to build and create more diverse and representative training samples is the key point in the future work in this research. There are several ways in which the size of the dataset can be effectively increased. Through data augmentation involving rotation, horizon flipping, and vertical shifts, the deep neural network model could learn the transformations further. By having different scales of larger or smaller training images, the learning procedure will be more sensitive to those new dimensions. This would also enhance the robustness of the algorithm to train for the detection of large defects and improve the results of grayscale images.
3. In addition, the specific training gave results for specific defects in the academic samples. In this work, training only involved using square, circle, and rectangle defects of Plexiglass, CFRP and Steel samples. The detection results indicate that similar defects could be detected on other types of training samples. But the results also show that if the learning model is tested on other defects that the model did not learn on, it would not be

an accurate system to rely on. Hence, to use the deep learning algorithm for training, we should clearly define the type of sample we are working on and enlarge the robustness of the system to learn this type of sample during the neural network training procedure. In addition, due to the time limitation, we simply labeled all the visible defects of each sample in this experiment. However, if we want to extract the feature map completely for each defect area, the positioning of less visible defects in infrared data will be a significant but challenging issue in further research.

4. Another disadvantage is that the objective localization algorithms are being influenced by the labeling process. Although fast and efficient to use, the bounding boxes also led to some restrictions in most cases. As can be seen, when the circle is present in bounding box, this involves a defect which is totally bounded by the box. However, this shows that although the entire defect is contained, the bounding box also extracted the non-defect area which possibly introduces multiple errors and less accuracy in the results. The proposal is to make a pixel-based labeling in order to achieve integrity in the image segmentation which would only label the defects and not a considerable area around each defect. This proposition can be further clarified by segmentation methods. The results presented here lead to a more reliable defects characterization with Pulsed thermography (PT).
5. A good defect characterisation is essential to not replace parts that could yet be used and to not leave critically damaged components without the needed repair. Therefore, these results are important, especially e.g. in the designing of autonomous diagnosis NDT systems, which can make decisions regarding the integrity of the inspected part by themselves.
6. Future work includes: (a) tests that can be performed with the instance segmentation method and other NDT techniques based on images like stereography and holography; (b) the best technique, method instance segmentation method (Center-Mask), which can still be improved by tuning the network parameters; (c) since the CNN technique achieves excellent performance, other network architectures must be tested and compared in the future to specify the best intelligent tool for defect measurement with infrared images.

### 3.11. Conclusions

In this work, six spatial deep learning models, involving: instance segmentation (Mask R-CNN; Center-mask). Autoencoder format semantic segmentation (U-net; Res-u-net), and the object localization model (YOLO-V3; Faster-RCNN) are applied for defect detection in infrared thermography. The evaluated results and analysis from different geometric specimens of Plexiglass, CFRP and Steel specimen with different aspect ratios (*size/depth*)

are indicated in Section 6. Each POD curve is related to the particular defect sizes which assess the quality of the results to land smoothly in the case of catastrophic failure results. These Spatial deep learning models are separately and comparatively discussed in brief. Future work will focus on the detection of more complicated structured materials through the modification and combination of different spatial and transient deep learning models.

## References

- [1] *Maldague X. Nondestructive evaluation of materials by infrared thermography. London: Springer; 1993.*
- [2] *Janssens O, Van de Walle R, Loccufer M, et al. Deep learning for infrared thermal image-based machine health monitoring[J]. IEEE/ASME Transactions on Mechatronics, 2017, 23(1): 151-159.*
- [3] *Ibarra-Castanedo C, Maldague X P V. Infrared thermography[M]//Handbook of technical diagnostics. Springer, Berlin, Heidelberg, 2013: 175-220.*
- [4] *Maldague X, Galmiche F, Ziadi A. Advances in pulsed phase thermography[J]. Infrared physics & technology, 2002, 43(3-5): 175-181.*
- [5] *Rajic N. Principal component thermography[R]. DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION VICTORIA (AUSTRALIA) AERONAUTICAL AND MARITIME RESEARCH LAB, 2002.*
- [6] *Pilla M, Klein M, Maldague X, et al. New absolute contrast for pulsed thermography[C]//Proc. QIRT. 2002, 5: 53-58.*
- [7] *Bison P, Cadelano G, Grinzato E. Thermographic Signal Reconstruction with periodic temperature variation applied to moisture classification[J]. Quantitative InfraRed Thermography Journal, 2011, 8(2): 221-238.*
- [8] *Weng J, Zhang Y, Hwang W S. Candid covariance-free incremental principal component analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8): 1034-1040.*
- [9] *Adegbola T. Pattern-based unsupervised induction of Yorùbá Morphology[C]//Proceedings of the 25th International Conference Companion on World Wide Web. 2016: 599-604.*
- [10] *Lakshmi A V, Gopitilak V, Parvez M M, et al. Artificial neural networks based quantitative evaluation of subsurface anomalies in quadratic frequency modulated thermal wave imaging[J]. Infrared Physics & Technology, 2019, 97: 108-115.*

- [11] Huang G B, Chen Y Q, Babri H A. Classification ability of single hidden layer feedforward neural networks[J]. *IEEE Transactions on Neural Networks*, 2000, 11(3): 799-801.
- [12] Huang S, Cai N, Pacheco P P, et al. Applications of support vector machine (SVM) learning in cancer genomics[J]. *Cancer Genomics-Proteomics*, 2018, 15(1): 41-51.
- [13] Fachrurrozi M, Fiqih A, Saputra B R, et al. Content based image retrieval for multi-objects fruits recognition using k-means and k-nearest neighbor[C]//2017 International Conference on Data and Software Engineering (ICoDSE). IEEE, 2017: 1-6.
- [14] Dai X, Duan Y, Hu J, et al. Near infrared nighttime road pedestrian's recognition based on convolutional neural network[J]. *Infrared Physics & Technology*, 2019, 97: 25-32.
- [15] Rubio Guillamón J. Bridge Structural Damage Segmentation Using Fully Convolutional Networks[D]. *Universitat Politècnica de Catalunya*, 2018.
- [16] Sirikham A, Zhao Y, Nezhad H Y, et al. Estimation of damage thickness in fiber-reinforced composites using pulsed thermography[J]. *IEEE Transactions on Industrial Informatics*, 2018, 15(1): 445-453.
- [17] Susa M, Maldague X, Boras I. Improved method for absolute thermal contrast evaluation using Source Distribution Image (SDI)[J]. *Infrared physics & technology*, 2010, 53(3): 197-203.
- [18] Zhao Z Q, Zheng P, Xu S, et al. Object detection with deep learning: A review[J]. *IEEE transactions on neural networks and learning systems*, 2019, 30(11): 3212-3232.
- [19] Garcia-Garcia A, Orts-Escolano S, Oprea S, et al. A review on deep learning techniques applied to semantic segmentation[J]. *arXiv preprint arXiv:1704.06857*, 2017.
- [20] Hafiz A M, Bhat G M. A survey on instance segmentation: state of the art[J]. *International Journal of Multimedia Information Retrieval*, 2020: 1-19.
- [21] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. *arXiv preprint arXiv:1804.02767*, 2018.
- [22] S. Ren, K. H. R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*, 2015.
- [23] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

- [24] Diakogiannis F I, Waldner F, Caccetta P, et al. Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 162: 94-114.
- [25] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//*Proceedings of the IEEE international conference on computer vision*. 2017: 2961-2969.
- [26] Lee Y, Park J. Centermask: Real-time anchor-free instance segmentation[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020: 13906-13915.
- [27] Sun J G. Analysis of pulsed thermography methods for defect depth prediction[J]. 2006.
- [28] Liu B, Zhang H, Fernandes H, et al. Experimental evaluation of pulsed thermography, lock-in thermography and vibrothermography on foreign object defect (FOD) in CFRP[J]. *Sensors*, 2016, 16(5): 743.
- [29] Hu J, Xu W, Gao B, et al. Pattern deep region learning for crack detection in thermography diagnosis system[J]. *Metals*, 2018, 8(8): 612.
- [30] Akhloufi M A, Tokime R B, Ellassady H. Wildland fires detection and segmentation using deep learning[C]//*Pattern Recognition and Tracking XXIX. International Society for Optics and Photonics*, 2018, 10649: 106490B.
- [31] Agarap A F. Deep learning using rectified linear units (relu)[J]. *arXiv preprint arXiv:1803.08375*, 2018.
- [32] Wang F, Jiang M, Qian C, et al. Residual attention network for image classification[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 3156-3164.
- [33] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. *arXiv preprint arXiv:1706.05587*, 2017.
- [34] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 2881-2890.
- [35] Grosse R. Lecture 15: Exploding and vanishing gradients[J]. *University of Toronto Computer Science*, 2017.
- [36] Soomro T A, Afifi A J, Gao J, et al. Strided U-Net model: Retinal vessels segmentation using dice loss[C]//*2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2018: 1-8.

- [37] Su L, Wang Y W, Tian Y. *R-SiamNet: ROI-Align Pooling Baesd Siamese Network for Object Tracking*[C]//2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, 2020: 19-24.
- [38] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. *Fcos: Fully convolutional one-stage object detection*. In ICCV, 2019
- [39] Zhu X, Cheng D, Zhang Z, et al. *An empirical study of spatial attention mechanisms in deep networks*[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6688-6697.
- [40] Wehling P, LaBudde R A, Brunelle S L, et al. *Probability of detection (POD) as a statistical model for the validation of qualitative methods*[J]. *Journal of AOAC International*, 2011, 94(1): 335-347.
- [41] Robertson S. *A new interpretation of average precision*[C]//Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. 2008: 689-690.

## **Part II. Synthetic Data for defect segmentation in deep learning**

The following one chapters will present one published paper concerning the application of infrared thermography for Non-Destructive Testing & Evaluation applied in procedure, which published in Journal of Big Data and Cognitive Computing: Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data

### **General explanation:**

Deep learning's Achilles heel is the training dataset. The training dataset must be accurate enough and contain enough images and exceptions (e.g., occlusions) so as to allow the algorithm to learn reliable features of interest. If the dataset does not contain enough images and cases of possible situations or if the data contain too much noise or artifacts, the algorithm may not be accurate enough or even may learn to identify the wrong features (in the cases of CM feature learning methods).

However, during the reality inspection in Infrared NDT, the composite materials (CFRP; GFRP) maintain high costing expense for the real raw data collection. The flexible and alternative way for saving expensive costs meanwhile achieving the highest quality inspection would be introducing a synthetic data generation of thermal data for the training in order to benefit from automatic defect detection and identification from the deep learning models combining with infrared thermography.

Therefore, in this study from the chapter, we will mainly focus on introducing and adapting inexpensive synthetic data merging with a certain amount of the experimental database for training the neural networks in order to achieve the compelling performance from a limited collection of the annotated experimental infrared data of a real-world practical thermography experiment.

## **Chapter 4 Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data**

*The results of this study were published in special issue “Machine Learning and Data Analysis for Image Processing” in the Big Data and Cognitive Computing Journal in February, 2021, cited 1 times up to now.*

### **4.1 Résumé**

Dans l'évaluation de la qualité (QE) du domaine de la production industrielle, la thermographie infrarouge (IRT) est l'une des techniques les plus cruciales utilisée pour évaluer les matériaux composites en raison de ses propriétés de faible coût, d'inspection rapide de grandes surfaces et de sécurité. L'application de réseaux neuronaux profonds tend à être une direction proéminente dans le contrôle non destructif (CND) de l'IRT. Pendant la formation du réseau neuronal, le talon d'Achille est la nécessité d'une grande base de données. La collecte d'énormes quantités de données de formation est une tâche très coûteuse. Dans le domaine des CND avec apprentissage profond, les données synthétiques contribuant à la formation en thermographie infrarouge restent relativement inexplorées. Dans cet article, des données synthétiques provenant des modèles d'éléments finis standard sont combinées avec des données expérimentales pour construire des référentiels avec des réseaux neuronaux convolutifs basés sur la région du masque (Mask-RCNN) pour renforcer le réseau neuronal, apprendre les caractéristiques essentielles des objets d'intérêt et réaliser une segmentation automatique des défauts. Ces résultats indiquent la possibilité d'adapter la fusion de données synthétiques peu coûteuses avec une certaine quantité de la base de données expérimentale pour l'entraînement des réseaux neuronaux afin d'obtenir des performances convaincantes à partir d'une collection limitée de données expérimentales annotées d'une expérience pratique de thermographie dans le monde réel.

#### **Contributing authors:**

**Qiang Fang (Ph.D. Candidate):** a part of the experiment planning, data collection, data analysis, designing and implementing the algorithm. Moreover, testing their accuracy and robustness throughout the process and writing the manuscript.

**Clemente. Ibarra. Castanedo:** the experiment planning, data collection, mineral preparation, revision and correction of the manuscript.

**Xavier Maldague (The research director):** supervision, revision and correction of the manuscript.



**Other contributors:** Annette Schwerdtfeger (research officer): manuscript preparation.

# **Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data**

Qiang Fang<sup>1</sup> \*, Clemente Ibarra-Castanedo<sup>1</sup> and Xavier Maldague<sup>1</sup>\*

<sup>1</sup> Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Université Laval, 1065, av. de la Médecine, Québec, QC G1V 0A6, Canada;

\* Correspondance: qiang.fang.1@ulaval.ca (Q.F.); Xavier.Maldague@gel.ulaval.ca (X.M.)

## **4.2 Abstract**

In quality evaluation (QE) of the industrial production field, infrared thermography (IRT) is one of the most crucial techniques used for evaluating composite materials due to the properties of low cost, fast inspection of large surfaces, and safety. The application of deep neural networks tends to be a prominent direction in IRT Non-Destructive Testing (NDT). During the training of the neural network, the Achilles heel is the necessity of a large database. The collection of huge amounts of training data is the high expense task. In NDT with deep learning, synthetic data contributing to training in infrared thermography remains relatively unexplored. In this paper, synthetic data from the standard Finite Element Models are combined with experimental data to build repositories with Mask Region based Convolutional Neural Networks (Mask-RCNN) to strengthen the neural network, learning the essential features of objects of interest and achieving defect segmentation automatically. These results indicate the possibility of adapting inexpensive synthetic data merging with a certain amount of the experimental database for training the neural networks in order to achieve the compelling performance from a limited collection of the annotated experimental data of a real-world practical thermography experiment.

**Keywords:** automatic defect detection; infrared thermography; deep learning algorithms; Non-Destructive Evaluation (NDE); supervised learning; image processing; data augmentation; Finite Element Models (FEM)

## **4.3 Introduction**

Quality evaluation is playing a fundamental role for modern industrial production and manufacturing processing. The demand for the inspection of materials with respect to the possible presence of defects, damage, and flaws has increased due to the wide use of composite materials and metals in industries such as aerospace. The method used to inspect defects via manual, visual evaluation involving humans can be hampered by the human fatigue and subjectivity. In order to meet the need of high-quality production and maintain

the stringent high-quality level, an advanced inspection system is becoming more and more essential for structural health monitoring and production lines application. Automated quality control [1] can be applied in the industrial field to facilitate the consistent and efficient inspection. The high inspection rate and the inspection procedure without human involvement are the main advantages of automated inspection systems [2].

Non-Destructive Testing (NDT) [3] encompasses a group of inspection techniques that are intended to assess the integrity of an object without causing any kind of modification or permanent damage. Infrared testing (IRT) is an NDT technique in which the differences in thermo-physical properties between the inspected object sound material and possible surface and/or subsurface defects are exploited through the use of an infrared camera. However, automatically identifying defects in materials via IRT still remains an ongoing and challenging task in the image processing domains.

For the defect identification and detection, several state-of-the-art defect detection algorithms have been proposed in previous literature. These included Faster-Region based Convolutional Neural Networks (Faster-RCNN) [4], YOLO-V3 [5], Autoencoders structured neural network [6], and conditional monitoring (CM)-based feature learning methods [7,8]. Faster-RCNN and YOLO-V3 detectors were used to automatically localize flaws in a thermography diagnosis system. These methods show satisfied performance on the defect localization when multi-properties were contained in defect information. However, some issues still need to be improved for the whole marking process in these methods. For example, bounding boxes from YOLO-V3 and Faster-RCNN offer some restrictions in most cases, although they are fast and easy to localize. The area of non-defect always occupied a high percentage of area from bounding boxes, which may possibly introduce multiple errors during the whole training process. The Autoencoders structured neural network [6] was proposed as an unsupervised learning method to automatically extract features from intelligent faults. This method has made impressive research progress during the thermography data processing. However, the encoder–decoder format (Autoencoders structured) algorithm may have unavoidable limitations in faults extraction from thermographic data processing due to weakness signals, complex noise interference, data limitation, etc. The regular CM based-feature learning methods derived from the idea of Convolutional Neural Network (CNN) feature transformations and transfer learning [9], which have shown strong capability as a pattern to detect the flaws or defect location naturally and obtained an excellent performance. Due to the data limitation as well as high noise inference from the experimental infrared thermal databases, it is still quite complicated to train a model that has strong robustness to detect irregular and complex flaws in the CM defect diagnosis system.

Therefore, in this research, an automatic instance segmentation and identification algorithm in deep learning (Mask-RCNN) [10] is introduced for defect segmentation in an automatic infrared detection system. The defect detection procedure can be regarded as either an object detection task [11] or object segmentation task [12]. In the object detection task, the objective is to fit the bounding box localized around each defect in the image. In the object segmentation task, the objective is based on pixel-level classification to distinguish each pixel if it is detected. Compared with the earlier segmentation strategies [13] [14], the Instance segmentation associated each pixel of an image with an instance label [15]. It can predict a whole segmentation mask for each of those objects and predict accurately which pixel in the input image corresponds to each object instance. It also reduces the restriction concerning the position of defects rather than predicting a group of bounding boxes for the defects.

Meanwhile, as discussed in the state-of-the-art approach [7,8], deep learning's Achilles heel is the training dataset. The training dataset must be accurate enough and contain enough images and exceptions (e.g., occlusions) so as to allow the algorithm to learn reliable features of interest. If the dataset does not contain enough images and cases of possible situations or if the data contain too much noise or artifacts, the algorithm may not be accurate enough or even may learn to identify the wrong features (in the cases of CM feature learning methods). During recent years, the exploitation of synthetic data [16] during the training and validation of deep neural networks has emerged as a popular topic. In this work, synthetic data generated with Finite Element Models (FEM) is used during the training process to greatly reduce the high expenses involved in real experiments in infrared thermography. The synthetic data generated by the simulator are based on the same parameters with real specimens in non-realistic ways, being able to force the neural network to learn the essential features of the object of interest [17]. The generated features from synthetic data can cover or overlap the amount of data distribution that are not sufficiently represented in the original experimental dataset [18]. The results also show the possibility of using inexpensive synthetic data for the deep neural network training but avoiding the necessity of collecting large amounts of hand-annotated experimental data.

In this research, a small amount of IRT experimental data together with FEM synthetically generated IRT data will be used to train a deep spatial characteristic model (Mask-RCNN) for the segmentation of defects. This proposed method can effectively achieve the defect identification and segmentation with the cases of data limitation. The contribution of this work can be illustrated in the following:

- 1) By adapting a data augmentation strategy through the Synthetic Data Generation Pipeline (Finite Element Modeling), the proposed method effectively improves the performance of segmentation (capability for feature extraction as well as reducing the noise interference).

- 2) An instance segmentation is introduced for defects segmentation and identification for each object of defects with different specimens to predict each irregular shape of defects instance in the input images at the pixel's level.

The remainder of the paper is structured as follows: Section 2 provides the pulsed thermography and related experimental set-up. Section 3 gives a detailed explanation of the Mask-RCNN defects detection system and the synthetic data generation pipeline. Section 4 describes the main features and dataset for evaluation. Section 5 provides the experimental results and implementation stages, including the comparison with states-of-the-art object detection algorithms (YOLO-V3; Faster-RCNN). Section 6 furtherly discusses and analyzes experimental results. Section 7 concludes this paper.

#### 4.4 Thermophysical Consideration

In pulsed thermography (PT), the surface of the inspected specimen is exposed to a heat pulse using an energy source such as photographic flashes, as indicated in Figure 4.1. A heat pulse can be represented as the combination of several periodic waves having different frequencies and amplitudes. After the thermal front reaches the surface of the specimen, a thermal front travels from the surface through the specimen. As time elapses, the normal path of the thermal front changes due to the presence of the subsurface discontinuity, and leads to the thermal contrast at the specimen's surface, which can be monitored with an infrared (IR) camera. The one-dimensional solution of the Fourier equation of conduction for a Dirac heat pulse propagating through a semi-infinite isotropic solid has the form [19]:

$$T(z, t) = T_0 + \frac{Q}{\sqrt{kpct}} e\left(-\frac{z^2}{4at}\right) \quad (4.1)$$

where  $Q$  [ $\text{J}/\text{m}^2$ ] is the energy absorbed from the surface, and the initial temperature is  $T_0$  [K].  $p$  [ $\text{kg}/\text{m}^3$ ] is density. The specific heat at constant pressure is  $c$  [ $\text{J}/\text{kgK}$ ].  $K$  [ $\text{W}/\text{mK}$ ] is the thermal conductivity that indicated the energy transformation rate by the materials.  $\alpha$  [ $\text{m}^2/\text{s}$ ] is the thermal diffusivity to measure the material ability to conduct heat, and  $Z$  [mm] is the depth of the defect. At the surface ( $z = 0$ ), the temperature evolution can be written as follows [20]:

$$T(0, t) = T_0 + \frac{Q}{\sqrt{kpct}} \quad (4.2)$$

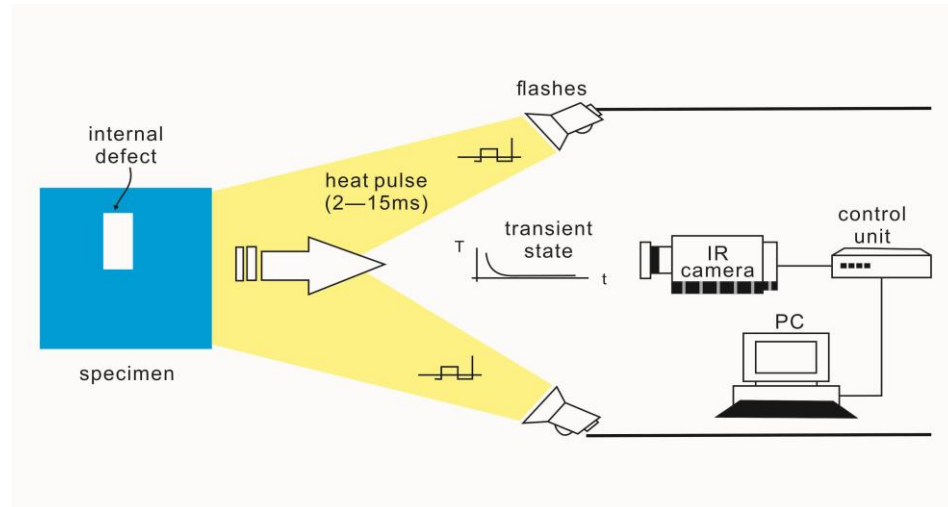


Figure 4. 1 Pulsed thermography experimental setup optical excitation.

From Equation (4.2), it can be concluded that the temperature evolution at the surface following a Dirac heat pulse follows a monotonous decrease with a slope of  $-1/2$  for areas without defects, whilst areas with defects will diverge more or less from this behavior depending on their actual thermo-physical properties.

#### 4.5 Automatic Defect Segmentation Strategy

As indicated in Figure 4.2, a detection system training with, for most intents and purposes, data generated by a computer simulation can be seen as synthetic data; experimental data has been proposed to segment and identify defects in thermal images. This design of this detection system is based on Mask R-CNN. The infrared thermal sequences are acquired by pulsed thermography (PT). Then, both the raw and synthetic sequences are preprocessed with the bilateral filtering and the thresholding technique. Therefore, two different characteristic datasets (raw data; synthetic data) are generated respectively, to be trained with Mask-RCNN in different combination cases. In the final step, delamination is identified through the deep region convolutional neural network and then visualized with the bounding box or instance segmentation map.

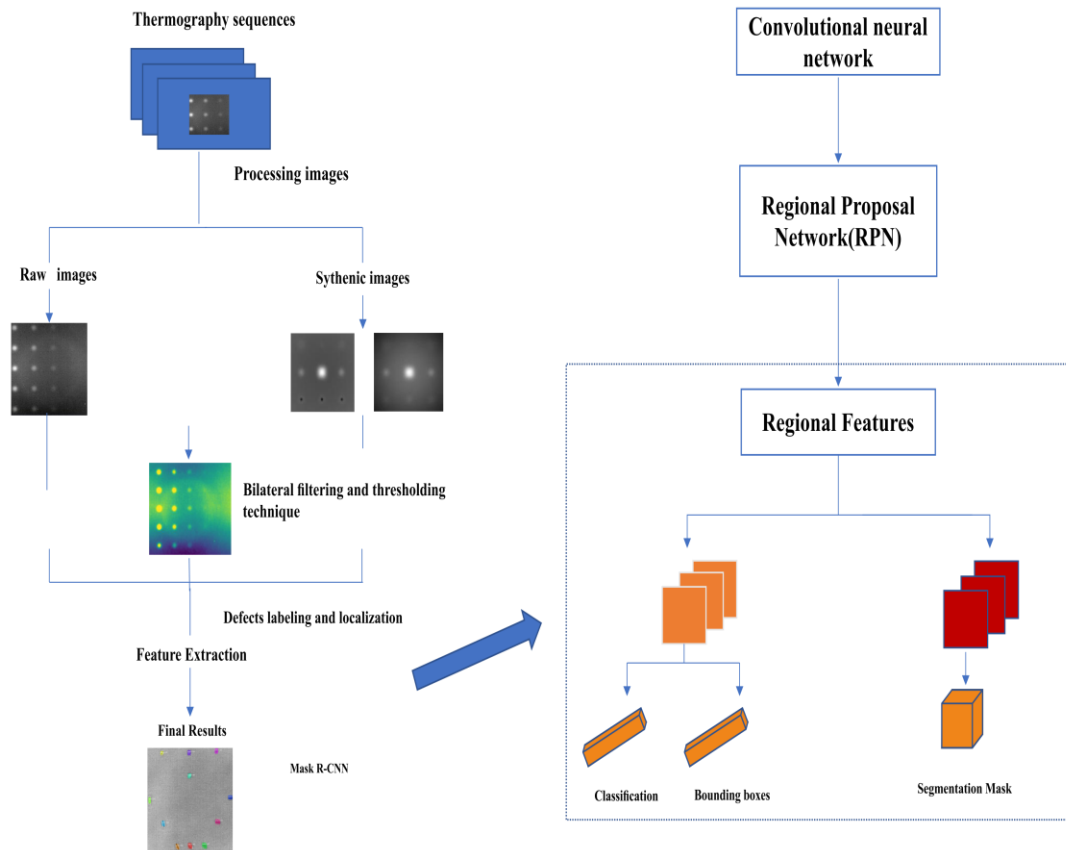


Figure 4. 2 Proposed segmentation strategy.

#### 4.5.1 Mask-RCNN

Convolutional Neural Networks (CNN) are utilized as a supervised feature extractor to analyze the defect localization and segmentation from materials. The Region based-CNN (R-CNN) [21] move toward the object bounding box detection method based on each region of interest (ROI). Faster-RCNN [11] have further advanced the network via learning the attention mechanism with a Region Proposal Network (RPN). Mask R-CNN [4] extends from Faster-RCNN through a mask branch construction. It is an instance segmentation algorithm, which recognizes object boundaries at the pixel level and designs pixel-wise segmentation with alignment of pixel-to-pixel among the input and output. Mask R-CNN adopts the same two-stage with Faster-RCNN architecture. During the first stages, the input images were scanned, and the proposals were generated via the Region Proposal Network (RPN). In the second stage of detection, it included two total branches. At the first branch (the blue line box at the top in the Figure 4.3), it predicted classification scores to tell what is the category corresponding to that region of proposal (whether or not it is background). Each of the

bounding box coordinates that regressed off the region proposal coordinates was indicated. At the second branch (the red line box at the bottom in the Figure 4.3), it added a binary mask regarding each ROI as a mini semantic segmentation network [22] to classify for each pixel in that input region proposal whether or not it is an object. Mask R-CNN is intended to address the instance segmentation task and aims at adjusting the plentiful hyper-parameters from the neural network. This model predicts a certain number of bounding boxes for the defects, while each defect region is segmented within the bounding boxes at the same time. The architecture is shown in Figure 4.3. The loss function of Mask-RCNN in Equation (4.3) consists of five different terms:

1. RPN\_class\_loss: The performance of objects can be separated from background via RPN;
2. RPN\_bounding\_box\_loss: The performance of RPN to specify the objects;
3. Mrcnn\_bounding\_box\_loss: The performance of Mask R-CNN specifying objects;
4. Mrcnn\_class\_loss: The performance of classifying each class of object via Mask R-CNN;
5. Mrcnn\_mask\_loss: The performance of segmenting objects via Mask R-CNN.

In these five terms, when the cost loss values are smaller, then performance improves.

$$L_{cost} = L_{(rpn\_class)} + L_{(rpn\_bbox)} + L_{(mrcnn\_bbox)} + L_{(mrcnn\_class)} + L_{(mrcnn\_mask)} \quad (4.3)$$

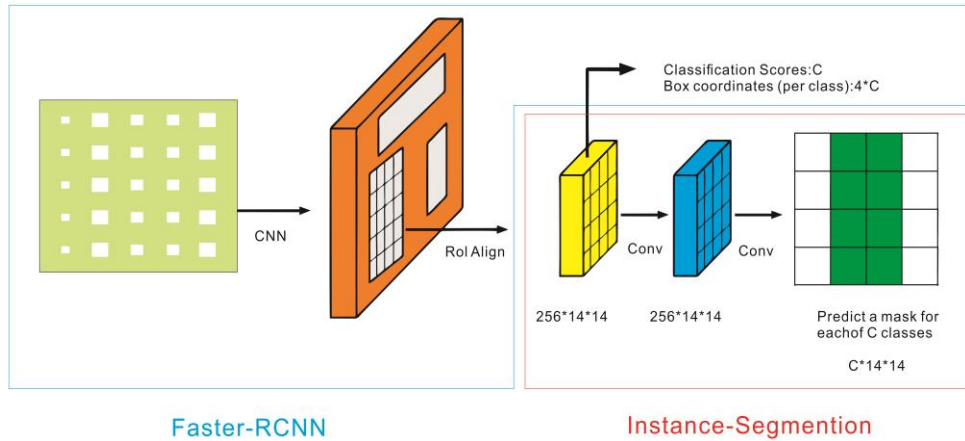


Figure 4. 3 Mask-RCNN processing architecture [6].

#### 4.5.2 Synthetic Data Generation Pipeline

Synthetic data are computed information rather than data collected from real-world events [16]. The main application involves using synthetic data as a training dataset. The use of synthetic data provides several benefits; one of the most interesting is that it is possible to

generate large databases to train, modify, and create new variables and features on individual data cases. Another advantage is that labeling is quite straightforward and precise.

Finite Element Modeling (FEM) has become a practical tool to evaluate the thermal response of pulsed thermography. It is also intended for predicting and simulating the results of Thermal NDT (Non-Destructive Testing) experiments and allows the solving of three-dimensional heat conduction problems in samples containing surface and subsurface defects. This plexiglass sample (30 cm  $\times$  30 cm) was designed with 25 flat bottom holes (FBH), which are located at increasing depths in either circles or square shapes, as indicated in Figure 4a. Figure 4b shows the pulsed thermography result of a plexiglass synthetic sample from FEM simulation at  $t = 106.5$  s after heating, which indicates defects more distinctly compared to the corresponding experimental frame at the same time shown in Figure 4.4 c. A good correspondence temperature of the region is observed, which verifies that synthetic modeling in FEM is in outstanding accordance with the real experiment [23]. The detailed implementation procedure for using deep neural networks with FEM is described in Figure 5. The training of synthetic data generated by FEM provides clean and automated supervision data compared to the other numerical methods such as boundary element methods [24] and finite volume methods [25], where most other numerical computational methods are highly non-linear and have uncertainty issues with respect to quantification and simulation. In addition, the learning process can be abundantly simplified since both synthetic and experimental data are governed by the same heat transfer principles. Specific feature positioning and extraction for less visible defect regions can be referenced from the synthetic data.

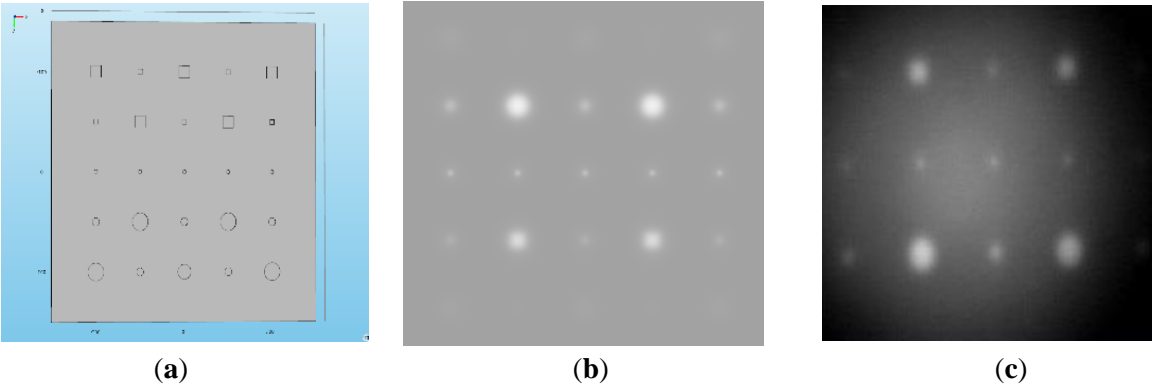


Figure 4. 4 (a) Finite Element Modeling (FEM) 3D model; (b) Simulated thermogram at  $t = 106.5$  s; (c) Real experimental data  $t = 106.5$  s.



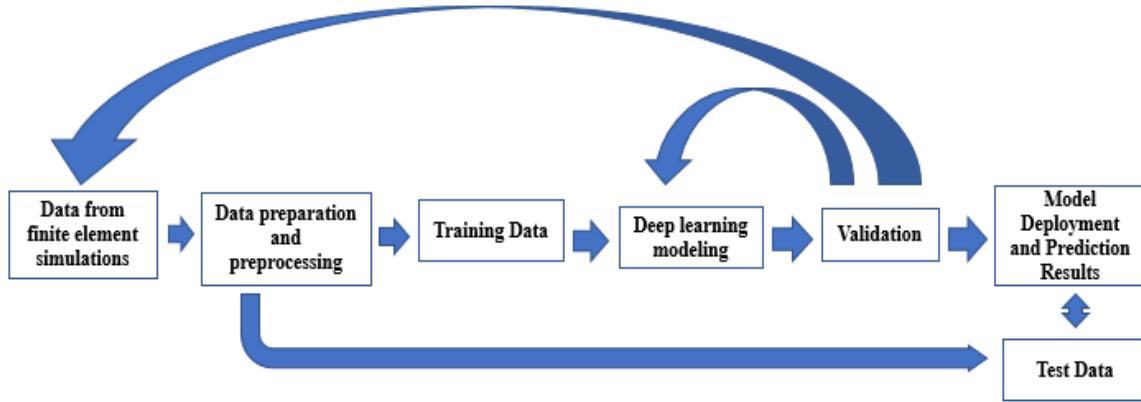


Figure 4. 5 Proposed workflow to train with a deep learning model based on the data generation by Finite Element Modeling.

#### 4.5.3. Automatic Preprocessing Stage

The main purposes of the preprocessing stage (Figure 4.6) are to make defects more distinguishable from the cluttered background of the thermal sequence and to normalize data. The bilateral filtering and thresholding techniques present in [26] were adapted for the preprocessed stage in this work so as to be the reference database. In step 1, the bilateral filtering removes noise from the thermal image while preserving feature edges. Secondly, two thermal criteria are applied in a thermal image sequence acquired from pulsed thermography: (1) 1st thermal criterion: automatically determine the instant just before the appearance of the first thermal footprint belonging to a defect (reference point); (2) 2nd thermal criterion: automatically find the intersection point between the overlap of each of the thermal image histograms after the reference point and the thermal image histogram corresponding to the reference point. In the end, the pixel values above the corresponding intersection point are equal to this value. Pixel values corresponding to the thermal images after the reference point are multiplied by a scale in Equation (4). The final result of this method is believed to be fitted with the convolutional neural network stage thereafter.

$$Scale\ Value = \frac{(Max\_pixel\_value - Min\_pixel\_value)_{corresponding\ to\ the\ 'n'\ thermal\ after\ Step1}}{(intersection\ point - Min\_pixel\_value)_{corresponding\ to\ the\ 'n'\ thermal\ after\ Step1}} \quad (4.4)$$

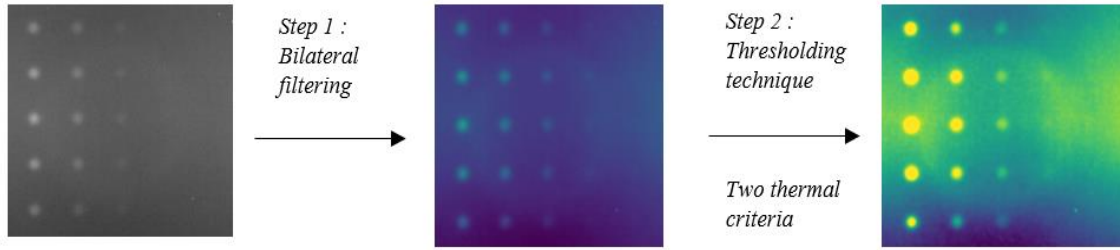


Figure 4. 6 Scheme of preprocessing stage

#### 4.6 Dataset and Features

The experimental samples and databases are acquired from two different materials: plexiglass and CFRP (carbon fiber-reinforced polymer). Two independent experimental groups were set up based on these materials. The whole database consisted of 500 images ( $446 \times 446$  pixels each). Six sub datasets were formed respectively from the two materials: plexiglass (A; B; E) and CFRP (C; D; F), which collected specific instants (5 s; 15 s; 25 s...) from 20 different thermal sequences during the temperature evolution in pulse thermography for both regarding training and validation data in order to avoid redundant information.

Based on the two types of images, 200 synthetic thermal images were generated with COMSOL, and 200 raw thermal images were collected from pulsed thermography experiments. Four groups of training databases were created based on the images we collected in the previous steps:

1. Database A, C: (Original database) 100 raw thermal images from thermal sequences with corresponding time;
2. Database B, D: (Mixed database) 100 raw thermal images with 100 new synthetic images; both selected from the same corresponding time;

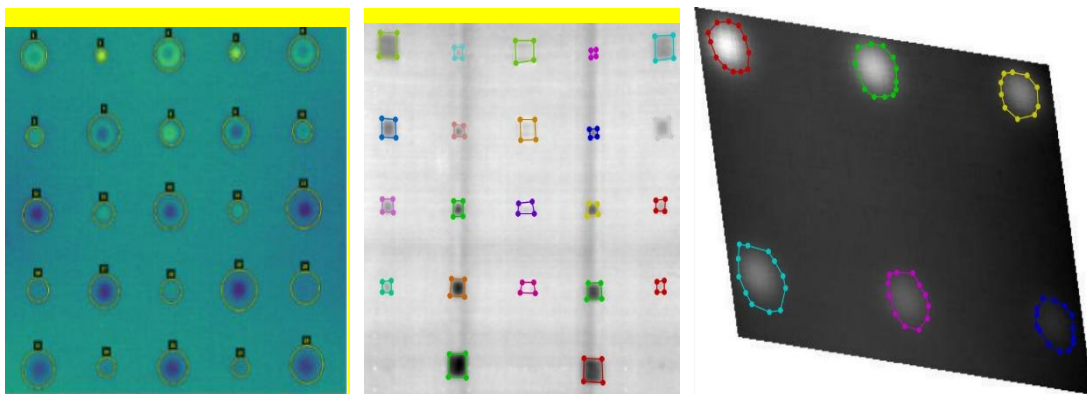


Figure 4. 7 Labels for preprocessed sample image.

All thermal images in the four training databases (A; B; C; and D) were preprocessed by the method in Section 4.5.3 in order to enhance the defect contrast.

Two raw (without preprocessing) thermal databases (E; F) (each consisting of 50 images) were used to validate the trained model (Mask-RCNN). These sequences, which have different shapes and depths of defects, were compared with the sequences during training. For sake of consistency, the raw thermal plexiglass database E was validated on the learning model trained on plexiglass (A, B), whilst the raw thermal CFRP database F was used to validate the learning model trained on CFRP (C, D). The defect labels and instance are shown in Figure 4.7, which includes the following shapes: circle, multilateral; square, rectangle; multi-angle of defect instances. LabelMe Image Annotator [27] was employed to label each defect region on the images; this allowed giving a simple and standard manual annotation for all of the images.

## 4.7 Experimental Results and Implantation Details

### 4.7.1. Evaluation Metrics (Average Precision (AP) and Probability of Detection (POD))

Average precision (AP) [28] is introduced to evaluate the performance capability of the detection model, which is analyzed by Equations (4.5) and (4.6).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4.5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4.5)$$

where TP is the true positive rate. FP is the false positive rate, which indicated that defects are in the non-defects regions and are falsely detected as defects. FN is the false negative rate, representing a certain number of the defects that failed to be detected during this experiment. Precision represents the accuracy of the prediction from the performance of the model. Recall represents the performance of the model to find all the possible positive cases in top priorities predictions.

In this work, a definition of Probability of Detection (POD) [29] (Equation (4.7)) also has been introduced to measure how reliable the non-destructive testing automatic inspector will be for a given case. POD is a specific metric especially for characterizing and quantifying defect analysis in NDT methods.

$$\text{POD} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (\text{Confidence score}=0.75) \quad (4.7)$$

The mathematical equation of *POD* keeps the same mathematical format as Recall in Equation 6, but they have different descriptions and explanations. In Equation (5.7) of *POD*, *TP* refers to true positive cases, which represent flaws in the system that the method can identify; *FN* refers to a false negative case, which indicates that the system has a flaw but the method cannot identify it. As a result, *POD* indicated how many flaws can truly be detected in a system from the whole amount of positive detected flaw cases.

In addition, we also introduced a definition herein for the further discussion of *POD* and defect classification-confidence score [30]. The confidence score can illustrate to which degree it is possible that the actual region boundaries overlap with the predicted defect region boundaries. In this work, our confidence score is set up as 0.75, which was selected as an acceptance criterion to define the percentage of the ground truth boundaries overlap with the predicted defect region boundaries for segmentation.

Therefore, the Probability of Detection (*POD*) assesses the performance of the Mask-RCNN (with synthetic data/without synthetic data) from the non-destructive evaluation-based inspection (NDE), which could determine the capability of the detection as a function of defect type. The probability of detection can form a core mathematical equation for examining and evaluating of the materials, which is also recognized as the quantitative evaluation tool involving the human factor and the various inspection parameters.

#### 4.7.2. Main Results Analysis and Discussion

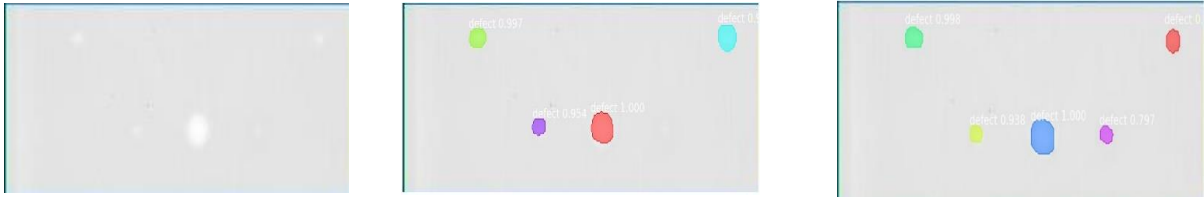
The operating system is set as Ubuntu 14.04; CPU: i7-5930k; Memory: 64GB. Each training processing was conducted on a GPU (NVIDIA GeForce GTX 1080Ti) and required approximately 30 min. Some main hyper-parameters and training parameters are set as below: (1) The Network training used Resnet101 as a backbone; (2) The learning momentum is 0.9 and learning rate is 0.0003; (3) The first 30 epochs were trained on network heads; then, all network layers were trained for 1030 epochs, the model weight (in h5 format) used was COCO (could be replaced with the original weight model file). This could be improved if a similar binary segmentation weight model was used; (4) The weight decay is 0.0001 and mini mask size is 56×56.

#### 4.7.3. Segmentation Results and Learning Curves

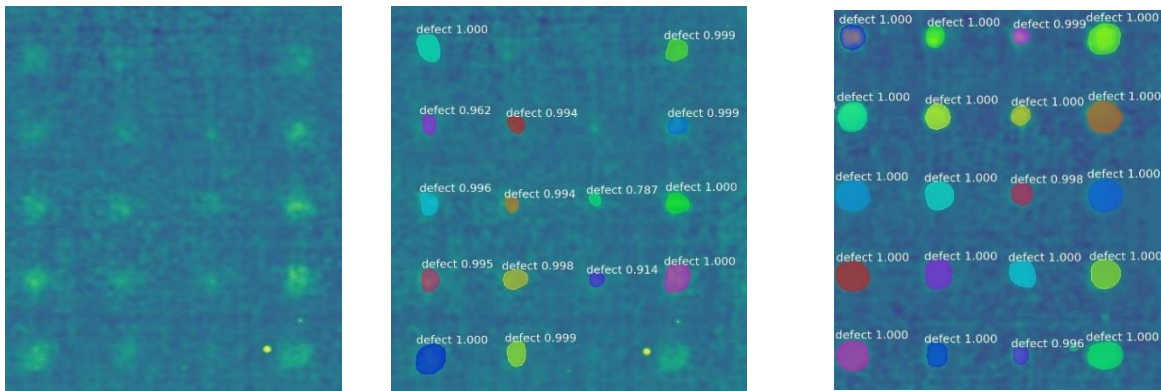
As shown in Figure 4.8 below, the segmentation results of five examples Figure 8a–e from the Mask-RCNN model are indicated. The segmented results in the middle column were the validated results from the model trained on raw images after preprocessing (Database A, C). The segmented results in the right column were the validated results from the model trained on raw images merged with synthetic images after preprocessing (Database B, D). Meanwhile, in the first three rows, the results from plexiglass specimens (Validation

Database E) are shown. The last two rows show the results for CFRP specimens (Validation Database F). The segmented defects from obtained images clearly show the improvement when trained on raw images merged with synthetic images after preprocessing in Figure 5.8. Although each segmented region has different color, all the colorful regions in the images represent true positive detective cases.

(a) Example 1 from plexiglass specimen (left: ground truth; middle: training without synthetic data; right: training with synthetic data)



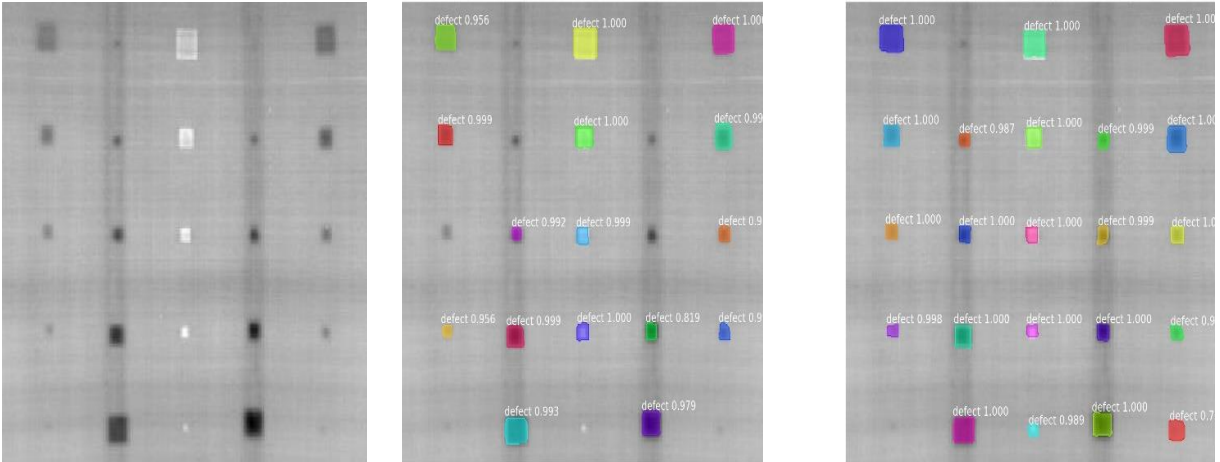
(b) Example 2 from plexiglass specimen (left: ground truth; middle: training without synthetic data; right: training with synthetic data)



(c) Example 3 from plexiglass specimen (left: ground truth; middle: training without synthetic data; right: training with synthetic data)



(d) Example 1 from CFRP specimen (left: ground truth; middle: training without synthetic data; right: training with synthetic data)



(e) Example 2 from CFRP specimen (left: ground truth; middle: training without synthetic data; right: training with synthetic data)

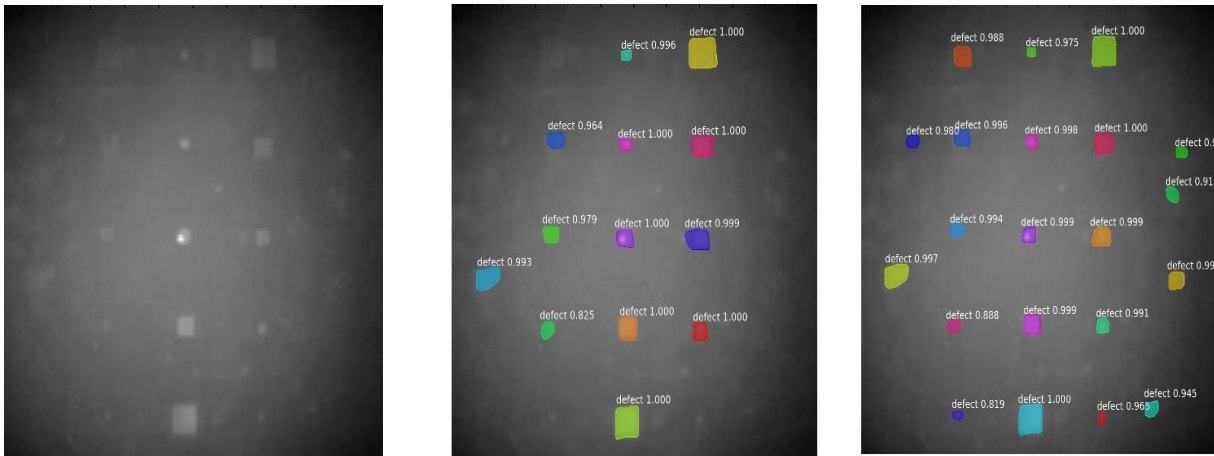


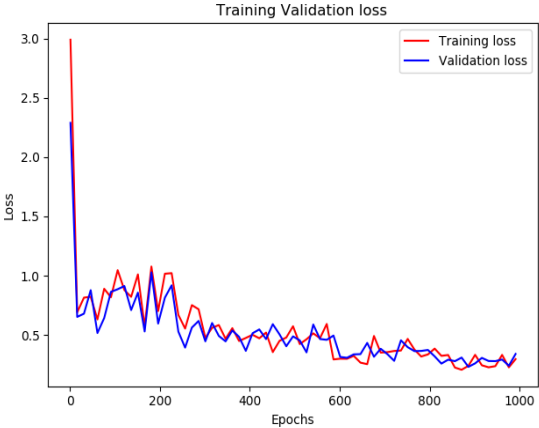
Figure 4. 8 The best obtained validation results of Mask-RCNN segmentation on different training databases. From left to right: original images, training on the preprocessed raw images database, training on the mixed database (preprocessed data from synthetic and raw images). From the first three rows to the last two rows: plexiglass (a–c), carbon fiber-reinforced polymer (CFRP) (d–e);

In addition, we can see the segmented results from the last column in Figure 4.8. When the model is trained with raw images (in the middle column), then the model is able to detect 12

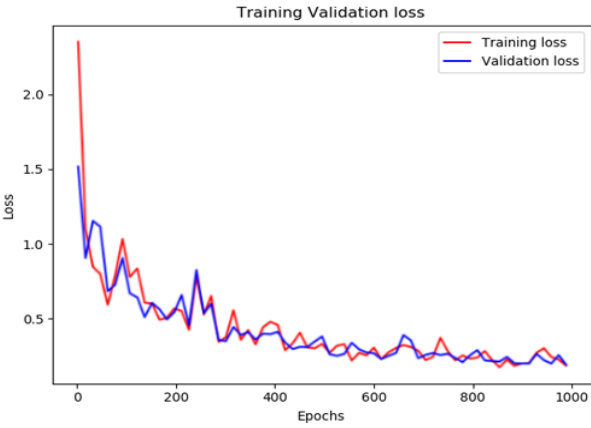
defects out of 25 defects in the whole specimen with limited raw images during the validation. Each detected defect is covered by a solid mask. It is notable that the model produces one false positive defect case by highlighting one non-existing defect on the image and 13 false negative defects (less visible defects, which are difficult to detect). Further, in the right column, when the training of the databases was increased with synthetic data, the trained model is able to detect 17 defects in total. However, four false positive and false negative cases still appear on the image.

The learning loss indicated how well the learning model performed. In this experiment, Figure 4.9 a, b showed the acquired average loss curve of the training and validation on Database A, B. The loss has a certain range of oscillation during the first 800 epochs in both Figure 4.9 a, b; then, it decreases less rapidly after 800 epochs until it flattens out. The cost curve is decreasing as the batch number increases and converges approximately from 0.25. Correspondingly, Figure 4.9 c, b indicates the learning curve with training and validation on Database C, D. The learning curve of the training and validation for Database D became more stable compared with the curves in Figure 4.9 c. Although the validation loss from Figure 4.9 c is instantly unstable around the point of the 380th epoch, the learning curves of training and validation (Database C, D) eventually converge around 0.2 around the 1000th epoch. Meanwhile, the model accuracy of the validation from Database A, B, C, D in this experiment is respectively 0.990, 0.996, 0.98738, and 0.98738. We obtained the model accuracy from four different training databases: B (0.996) > A (0.990) > C (0.987) = D (0.987). As a result, the obtained model accuracy from four groups of databases is impressive and shows the good performance during the training and validation when the Mask-RCNN model is used for defect segmentation of plexiglass and CFRP specimen in this project.

(a) Learning loss from Database A



(b) Learning loss from Database B



(c) Learning loss from Database C

(d) Learning loss from Database D



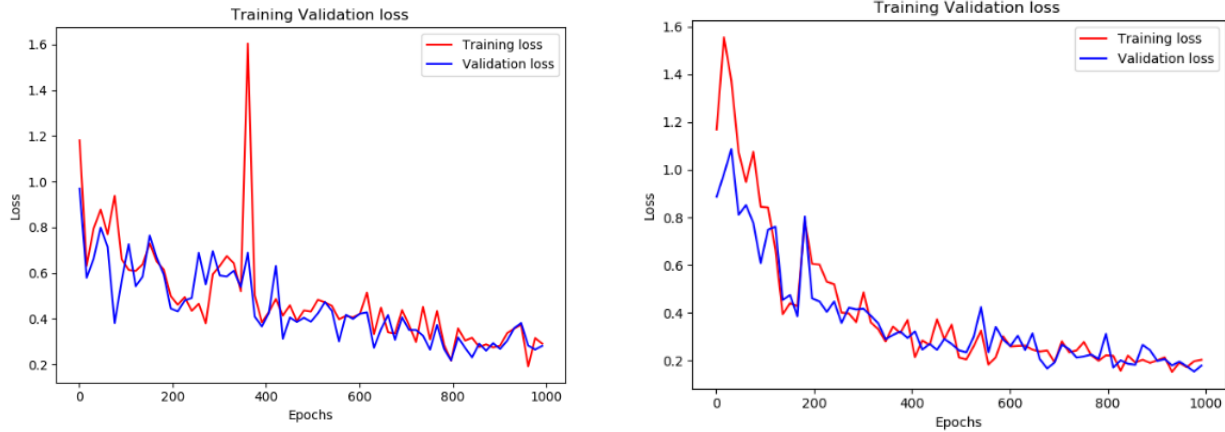


Figure 4. 9 The average learning loss for two types of specimens: plexiglass (a, b); CFRP (c, d).

#### 4.7.4. Precision–Recall Curves (PR Curves)

Precision–recall curves [31] for the Mask-RCNN architectures trained on four different groups of databases are shown in Figure 4.10. This figure indicated two PR curves for training sets of plexiglass specimens (preprocessed raw images database (A); synthetic images merging with raw images database from preprocessing (B)); an independent pure raw images validation set without preprocessing (E)) and two PR curves for training sets of CFRP specimens (preprocessing raw images database (C); synthetic images merging with raw images database from preprocessing (D); an independent pure raw images validation set without preprocessing (F)). The mean average precision (mAP) values in databases A, B, C, and D respectively reach 68.66, 76.40, 70.29, and 72.68 as shown in Figure 10.

During the training of Database (B, D), Mask-RCNN has seen limited preprocessed raw thermal images (100 images) merging with preprocessed synthetic data (beyond pretraining of the early layers on COCO [32]). The mAP values in the mixed database (B, D) are higher than those of Databases (A, C) ( $B (76.40) > A (68.66)$ ;  $D (72.68) > C (70.29)$ ). From the plot of the PR curves, Database B achieves higher precision than Database A consistently for all of the recall values in the Mask-RCNN model. On the other hand, database D consistently obtained lower precision than database C between the recall values from 0.5 to 0.55. This helps explain that the plexiglass database obtained the higher mAP difference compared with CFRP between the training of raw images and the training from databases merging with synthetic data.



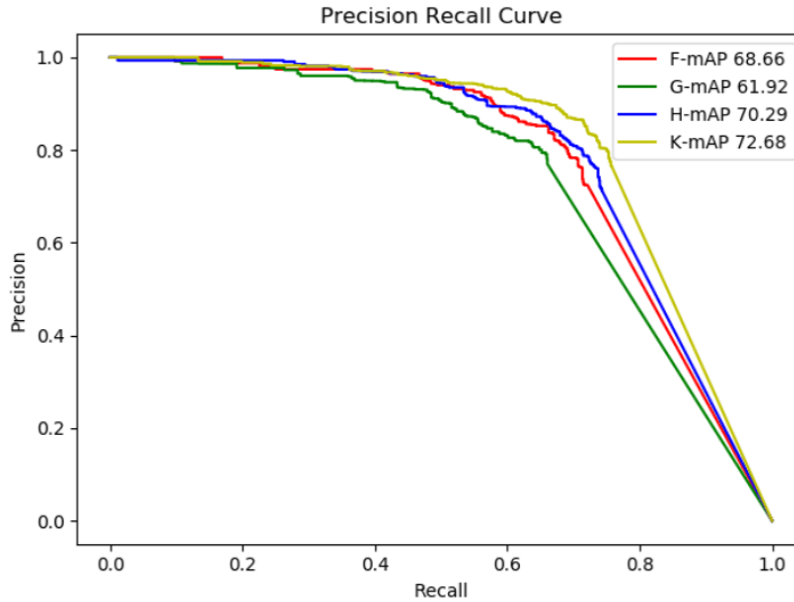


Figure 4. 10 Different detection results with four groups of datasets (two types of materials).

In each region of interest (ROI), this experiment obtained better results with the performance metric of precision and recall values compared with another cited deep learning IRT project [5,33]. Especially, the plexiglass obtained a superior performance of mAP from Mask-RCNN when it was merged with synthetic data for training compared with the highest mAP value obtained from previous research. This compelling performance also illustrates the ability of synthetic data to bridge the reality gap and demonstrates that merging with synthetic data for training can improve the accuracy in both types of material (CFRP; plexiglass). As a result, the training model merged with synthetic images could be able to outperform databases with pure raw thermal databases in deep learning IRT projects.

#### 4.7.5. Evaluation with Probability of Detection

The Probability of Detection has been successfully applied for analyzing the detection capability of materials from Mask-RCNN. In this study, the POD was classified as a function of an aspect ratio  $r = \text{diameter (D)}/\text{depth (d)}$ . Each POD point on the curve corresponds to a specific aspect ratio from a particular diameter (D) and depth (d). As indicated in Figure 4.11a, the red POD curve (Database B) maintained the highest POD curve from the whole aspect ratio, which indicated that the Mask-RCNN with synthetic data detected the largest percentage of defects in plexiglass specimens. Simultaneously, the POD of Database B is better than the corresponding POD curves (Database A), confirming that synthetic data merged with raw data can contribute to achieve better performance as compared to the model

trained with raw data only. Based on these results, the POD of the databases that merged with synthetic data (B; C) has a better detection capability than any other database without synthetic data (A; D), revealing that merging the raw databases with synthetic data could be a reliable procedure for a deep learning model (Mask-RCNN), enhancing the capability of automatic defect segmentation and identification. The pink POD curve (C), which represents the CFRP database training with synthetic data, has an overall higher detection probability than the blue curve (D) (CFRP database training without synthetic data). However, in comparison with the POD curve (A; B), the POD curve (D) shows a lower detection performance than the POD curve B due to data diversity and data augmentation capability (different defect shapes and multiple angles, diverse formatted defects) that PLEXI databases have. As a result, PLEXI databases outperformed CFRP databases on the probability scores in this case.

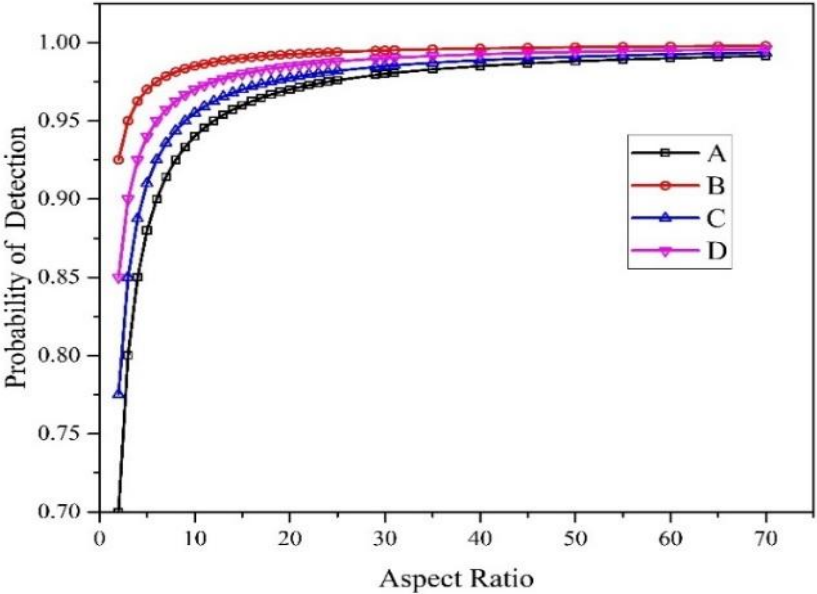


Figure 4. 11 Probability of distribution curve of different methods for processing on CFRP samples (a)/PLEXI samples (b) (confidence score = 0.75).

4.7.6. Defect Classification Analyses

As indicated in Table 4.1 below, the detection with defects in each training database (A, B, C, D) by the Mask-RCNN algorithm in pulse thermography has been investigated with the objective to measure the global accuracy in each database [34]. Four confusion matrixes from training databases A, B, C, and D have been indicated, where TP = True Positive, FP = False Positive, TN = True Negative, and FN = False Negative. Each column of the matrix stands

for the objectives in the actual class; simultaneously, each row of the matrix stands for the objectives in the predicted class.

Table 4. 1 Class mark list—comprehensive results.

Database		A/E *		B/E *	
<b>Actual Class</b>					
Class		Defect	Non-defect	Defect	Non-defect
Predicted	Defect	TP: 1785	FP: 229	TP: 2060	FP: 199
Class	Non-defect	FN: 456	TN: 291	FN: 181	TN: 321
Database		C/F *		D/F *	
<b>Actual Class</b>					
Class		Defect	Non-defect	Defect	Non-defect
Predicted	Defect	TP: 1442	FP: 257	TP: 1610	FP: 225
Class	Non-defect	FN: 358	TN: 296	FN: 190	TN: 328

\* Training database/Testing database.

The results from detection have shown an excellent performance when the original database was merged with synthetic data based on the number of well-classified labels. The approximately average prediction of each database is shown in the Table 4.1. We set up the confidence score threshold as 75% of probability of detection to distinguish TP, FP, TN, and FN cases.

As mentioned earlier, the objective in this research is to automatically extract and segment the features (i.e., defects) that could be found in each thermal frame. Each classification may either be a defect or a non-defect region in this task. In data science, an effective metric to validate the performance of a detection algorithm widely used in the deep learning applications is the confusion matrix [35]. In the confusion matrix, each matrix is set up  $2 \times 2$ , where 2 is the number of classes in the dataset. The row elements of a confusion matrix represent the classes to which the features of the image belong, either to the feature region of a defect or a non-defect. Its columns, on the other hand, represent the classes given by the Mask-RCNN during the segmentation process.

Generally, in this case, based on Table 4.1 (confidence score = 75%), the testing Database E with the training Database A, 1785 defects were segmented correctly, 456 were missed, 229 non-defect regions were detected as defects, and 291 non-defect regions were correctly labeled as non-defects in this experiment. In the case of testing Database E with the training Database B, 2060 defects were detected correctly, 182 were missed, 199 non-defect regions were detected as defects, and 321 non-defect regions were correctly labeled as non-defect in this experiment. In the case of the testing Database F with the training Database C, 1442

defects were detected correctly, 358 were missed, 257 non-defect regions were detected as defects, and 296 non-defect regions were correctly labeled as non-defect in this experiment. In the case of the testing Database F with the training Database D, 1610 defects were detected correctly, 190 were missed, 225 non-defect regions were detected as defects, and 328 non-defect regions were correctly labeled as non-defects in this experiment. Then, it can be concluded from the confusion matrix that after merging with synthetic data, Mask-RCNN leads to a superior performance in defects detection than the results obtained previously.

In addition, the global accuracy Acc given in the corresponding training databases (A–D) and test databases (E, F) in Figure 4.12 is defined in Equation (8). The implementation of the deep learning model (Mask-RCNN) used for the whole validation process is set according to the principle as follows: the specific testing Dataset (E) corresponding to the training data (A, B); the specific testing data (F) corresponding to the training data (C, D) respectively.

$$Acc = \text{Average total accuracy} = \frac{\text{correctly classify regions}}{\text{Total number of evaluated regions}} \tag{4.8}$$

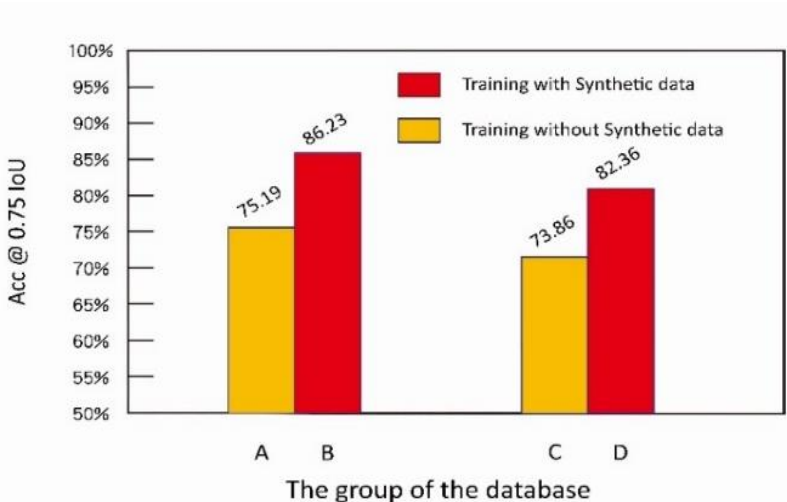


Figure 4. 12 The total performance of accuracy with Mask-RCNN on CFRP and PLEXI samples with/without synthetic data.

4.7.7 The Comparisons with State-of-the-Art Deep Learning Detection Algorithms

In this session, YOLO-V3 and Faster-RCNN have been selected as the state-of-the-art object detectors to compare with the performance of Mask-RCNN when it is merged with the synthetic data or without during the training. A CFRP specimen has to be validated on four different training situations from Database C, D in Section 4 for three types of object

detectors in Figure 4.13, which included (a) Mask-RCNN training with Database C; (b) Mask-RCNN training with Database D; (c) YOLO-V3 training with Database C; and (d) Faster-RCNN training with Database C.

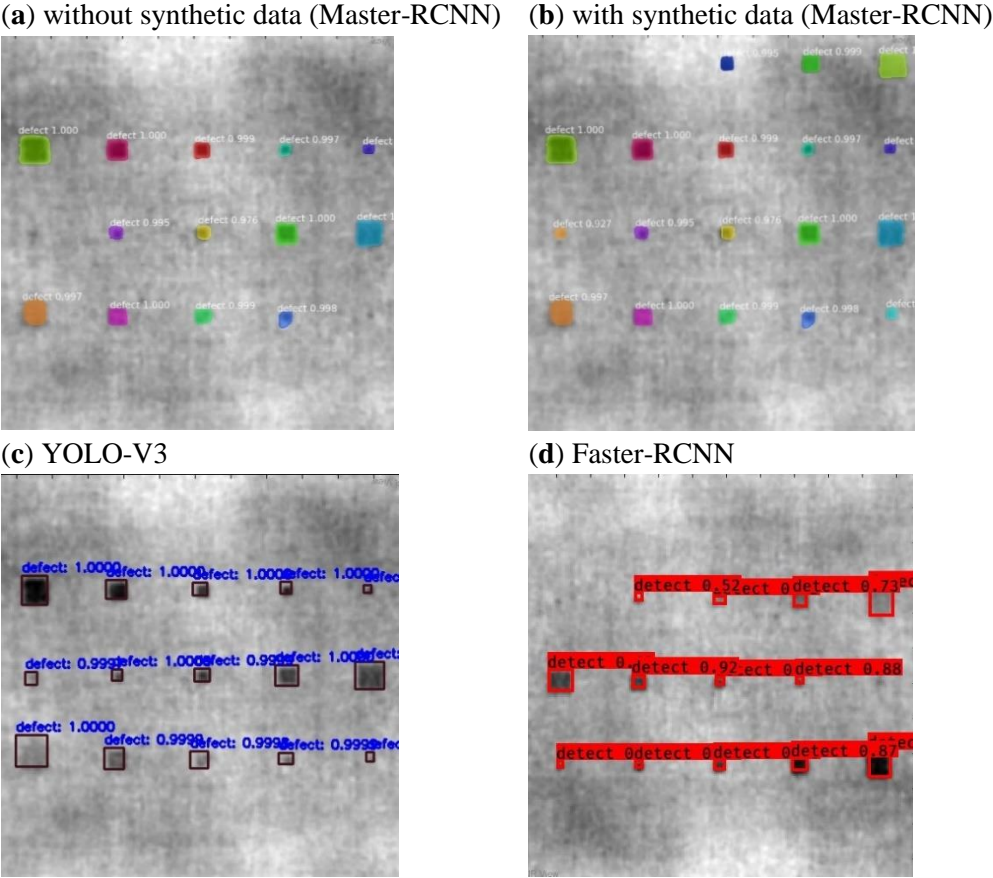


Figure 4. 13 Detection results on a reprehensive CFRP specimen provided by different objective detection algorithms or scenarios (a) Master-RCNN without synthetic data; (b) Master-RCNN with synthetic data; (c)YOLO-V3; (d) Faster-RCNN.

It can be seen from the Figure 4.13 that when Mask-RCNN was only training with experimental data in Figure 5.13 a, it obtained similar accuracy based on comparing the results those of with YOLO-v3 in Figure 13c and Faster-RCNN in Figure 13d. However, after the training databases merged with synthetic data in Figure 13 b, the results from Mask-RCNN outperformed the other three different situations.

On the other hand, frames per second (fps) [36] has been introduced as a concept to verify how many pictures can be processed in a unit time (one second) by an objective detector in

order to evaluate the running time complexity for each detector. In Table 4.2, the time complexity of three deep learning algorithms have been illustrated. In this experiment, during the training processing being conducted on GPU (NVIDIA GeForce GTX 1080Ti YOLO-V3, three deep learning detectors YOLO-V3, Mask-RCNN, Faster-RCNN running are 15 fps, 5fps, and 1fps, respectively. As a result, comparing with Mask-RCNN and Faster-RCNN, YOLO-V3 has the highest running time speed to process images due to the reason that YOLO-v3 is a one-stage real-time detector, and it has a much faster speed than other detectors (Mask-RCNN; Faster-RCNN), which are two-stage procedures (Region Proposal Network (RPN); ROI pooling). However, as indicated in Figure 4.14, Mask-RCNN still obtained higher POD scores during the whole validation process when it merged with synthetic data based on the different ratio values (size/depth).

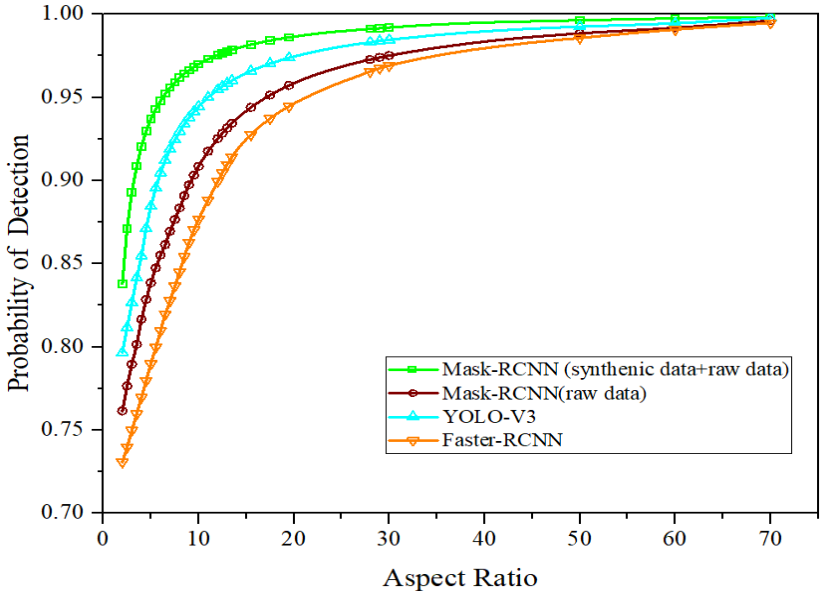


Figure 4. 14 Probability of distribution of different deep learning methods on CFRP databases (confidence score = 0.75).

Table 4. 2 Running time complexity comparison with state-of-the-art methods.

Running Time Complexity	YOLO-V3	Mask-RCNN	Faster-RCNN
Frame per second (FPS)	15	5	1

**4.8. Result Analysis and Discussion**

The main objective of synthetic data in this project is to enlarge the datasets when you only have limited raw experimental data. Synthetic data can more easily generate a dataset with a large amount of variety. Small raw training databases merging with synthetic data can help

detect most of the defects successfully while avoiding expensive experiments. As we can see from Section 4.7.4, Figure 4.10, which shows the performance of the definite and regularly used evaluation metric (precision and recall), higher average precision gives better results when trained on the data merging with synthetic data than when trained on the raw data only. Therefore, we are able to boost the model performance by feeding the synthetic data to merge or replace the actual one.

This confirmed that in different situations, when the synthetic data merged with the sample, it can actually benefit the learning network, enabling it to learn and understand the pattern by (a) generating more significant samples than the original available data, which is limited; (b) providing more data of any specific minority cases (for example, the highly expensive thermal data of the CFRP samples, which is difficult to obtain) in order to avoid the underrepresentation or underperformance situation for automatic defects segmentation.

Comparing with the up-to-date CNN defect detection work [4–8] in infrared thermography as mentioned in introduction, the Mask-RCNN model merging with synthetic data has become the most flexible method that has the capacity to obtain the improvable performance. Especially in pulsed thermography, too little data might be available and obtained in the realistic experiment. It is always too expensive and time-consuming to training with the additional thermal data during the training and defect detection by the deep learning algorithm. Therefore, the generated synthetic data show their ability to support this case. As indicated in Figure 4.11 of Section 4.7.5, the POD curve of Database D exceeds the POD curve of Database C from all aspect ratio values (size/depth), which illustrated that synthetic data have the potential to immensely improve performance in regard to defect detection for CFRP and plexiglass specimens. On the other hand, in the defect classification analysis in Figure 4.12 from Section 4.7.6, the global accuracy  $Acc$  increasing predominately from the testing results when the trained system performed with Database B and Database D, which further demonstrated that a trained system merged with synthetic data may beneficial for detecting and identifying defects.

As shown in the results from different objective detectors indicated by the POD evaluations from Figure 4.14 (Section 4.7.7), it is obvious that YOLO-V3 obtained comparable accuracy with the Mask-RCNN method during the whole validation process, while the Faster-RCNN gave mediocre performance. However, after merging with synthetic data, our proposed method shows excellent performance because the data augmentation strategy has a strong effect on the capability of the Mask-RCNN model to capture less visible defects and separate low-rank non-defect information from the thermal data, after which the meaningful features remain. It is worth mentioning that the proposed method (Mask-RCNN) maintains the highest accuracy in POD evaluation when it encounters the challenging and lowest (0–5) aspect ratio value (size/depth). With all task-driven objective detectors based on the deep

learning architecture, the proposed method keeps the highest accuracy results over the other state-of-the-art algorithms. This excellent performance is thanks to the capability of the synthetic data generation pipeline's data augmentation strategy to segment the defects from the multi feature distribution.

Furthermore, as mentioned in Section 4.5.2, the COMSOL is capable of creating high-quality synthetic thermal data from scratch on specific data points and the same heat transfer principles setting from pulsed thermography. However, based on the results, there is still some room for improvement. The defects from synthetic data (COMSOL) may still have blurry boundaries and unclear shapes that influence the training results from learning. Mask-RCNN is robust to the limited size of synthetic data and is more likely to overfit unless the data increase the augmentation (rotating, flipping images, or random Gaussian noise). We applied the following data augmentation strategy: flip, rotation, scale, crop and adding the Gaussian noise to each individual synthetic and raw image. We stopped all the training during the time of the performance, nearly saturating with the overfitting.

In additional, in this task, we used Resnet101[37] as a feature extractor with weights initialized from COCO in Mask-RCNN. Since the weights from COCO obtained the features trained by a group of datasets that contain basic lines and shapes of the objectives (circle, square, rectangle), it can be beneficial to the defect detection to a certain degree even without training with thermal images. So, the selection of beneficial initial weights from the pretrained models could be a key training strategy to apply for the future direction of research in deep learning IRT.

In future work, the other available thermography methods in the literature will be implemented with Mask-RCNN and a synthetic data generation pipeline to further exploit its possibility to enhance its capability of defect detection. A modified version of the Mask-RCNN network will be introduced for enhancing its training performance. The further plan can be also illustrated as follows: (a) enhancing the training performance by adjusting the key parameters and structure in Mask-RCNN; (b) exploitation of variety software in Finite Element Modeling for generating synthetical data for defect detections for comparison with COMSOL; (c) Mask-RCNN with a synthetic data generation pipeline test with other NDT techniques to evaluate the performance of defect analysis.

#### **4.9. Conclusions**

In summary, we present a method where the deep learning architecture is combined with a small amount of synthetic thermal data for defect segmentation. The proposed architecture is based on a Mask-RCNN that has been applied in natural image segmentation. We collected our thermal database from the experimental results and synthetic data in the Finite Element



Model (COMSOL). The network yields a better performance when it is fused with synthetic data for training. Meanwhile, different types of composite materials (CFRP; plexiglass) with defects have been validated via the proposed method to reveal the performance of detection. The proposed algorithm, Mask-RCNN without modification, is in line with the main objective in this work, which is to assess whether synthetic data can improve the detection accuracy of defects on the CFRP or plexiglass composite samples. Further quantitative analysis of the diameter/depth ratio with POD evaluation in defect detection with a modified structured Mask-RCNN network will be carried out in future work.

## References

- [1] Ferguson, M.K.; Ronay, A.; Lee, Y.T.T.; Law, K.H. *Detection and Segmentation of Manufacturing Defects with Convolutional Neural Networks and Transfer Learning*. *Smart Sustain. Manuf. Syst.* 2018, 2, doi:10.1520/SSMS20180033.
- [2] Maldague, X.; Krapez, J.C.; Poussart, D. *Thermographic nondestructive evaluation (NDE): An algorithm for automatic defect extraction in infrared images*. *IEEE Trans. Syst. Man Cybern.* 1990, 20, 722–725.
- [3] Rajic, N. *Principal Component Thermograph; Defence Science and Technology Organisation Victoria (Australia) Aeronautical and Maritime Research Lab: Fisherman's Bend, VIC, Australia, 2002.*
- [4] Hu, J.; Xu, W.; Gao, B.; Tian, G.Y.; Wang, Y.Z.; Wu, Y.C.; Yin, Y.; Chen, J. *Pattern deep region learning for crack detection in thermography diagnosis system*. *Metals* 2018, 8, 612.
- [5] Fang, Q.; Nguyen, B.D.; Castanedo, C.I.; Duan, Y.X.; Xavier, M. *Automatic Defect Detection in Infrared Thermography by Deep Learning Algorithm*. Available online: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11409/2555553/Defects-detection-in-infrared-thermography-by-deep-learning-algorithm/10.1117/12.2555553.short?SSO=1> (accessed on 26 February 2021).
- [6] Jia, F.; Lei, Y.; Guo, L.; Lin, J.; Xing, S. *A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines*. *Neurocomputing* 2018, 272, 619–628.
- [7] Janssens, O.; Van de Walle, R.; Loccufier, M.; Van Hoecke, S. *Deep learning for infrared thermal image based machine health monitoring*. *IEEE/ASME Trans. Mechatron.* 2017, 23, 151–159.
- [8] Zhang, Y.; Fjeld, M. *Condition Monitoring for Confined Industrial Process Based on Infrared Images by Using Deep Neural Network and Variants*. Available online:

[https://www.researchgate.net/publication/340599990\\_Condition\\_Monitoring\\_for\\_Confined\\_Industrial\\_Process\\_Based\\_on\\_Infrared\\_Images\\_by\\_Using\\_Deep\\_Neural\\_Network\\_and\\_Variants](https://www.researchgate.net/publication/340599990_Condition_Monitoring_for_Confined_Industrial_Process_Based_on_Infrared_Images_by_Using_Deep_Neural_Network_and_Variants) (accessed on 26 February 2021). *Big Data Cogn. Comput.* 2021, 5, 9 20 of 21

- [9] Torrey, L.; Shavlik, J. *Transfer learning*. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI global: Hershey, PA, USA, 2010; pp. 242–264.
- [10] He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. *Mask r-cnn*. In *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017*; pp. 2961–2969, doi:10.1109/ICCV.2017.322, doi:10.1109/ICCV.2017.322.
- [11] Ren, S.; He, K.; Girshick, R.; Sun, J. *Faster r-cnn: Towards real-time object detection with region proposal networks*. In *Proceedings of the Advances in Neural Information Processing Systems, Montreal, Canada, 7–12 December, 2015*; pp. 91–99, doi:10.1109/TPAMI.2016.2577031.
- [12] Long, J.; Shelhamer, E.; Darrell, T. *Fully convolutional networks for semantic segmentation*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015*; pp. 3431–3440, doi:10.1109/CVPR.2015.7298965.
- [13] Ronneberger, O.; Fischer, P.; Brox, T. *U-net: Convolutional networks for biomedical image segmentation*. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
- [14] Badrinarayanan, V.; Handa, A.; Cipolla, R. *Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling*. *arXiv* 2015, arXiv:1505.07293.
- [15] Ren, M.; Zemel, R.S. *End-to-end instance segmentation with recurrent attention*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017*; pp. 6656–6664.
- [16] Tremblay, J.; Prakash, A.; Acuna, D.; Brophy, M.; Jampani, V.; Anil, C.; To, T.; Cameracci, E.; Boochoon, S.; Birchfield, S. *Training deep networks with synthetic data: Bridging the reality gap by domain randomization*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, Utah, USA, 23 April 2018*; pp. 969–977, doi:10.1109/CVPRW.2018.00143.
- [17] Richter, S.R.; Vineet, V.; Roth, S.; Koltun, V. *Playing for data: Ground truth from computer games*. In *Proceedings of the European Conference on Computer Vision, Amsterdam, the Netherlands, 8–16 October 2016*; Springer: Cham, Switzerland, 2016; pp. 102–118, doi.org/10.1007/978-3-319-46475-6\_7.

- [18] McCormac, J.; Handa, A.; Leutenegger, S.; Davison, A.J. SceneNet RGB-D: 5M photorealistic images of synthetic indoor trajectories with ground truth. *arXiv* 2016, arXiv:1612.05079.
- [19] Ibarra-Castanedo, C.; Maldague, X.P.V. *Infrared Thermography*. In *Handbook of Technical Diagnostics*; Springer: Berlin/Heidelberg, Germany, 2013; p. 826.
- [20] Maldague, X. *Theory and Practice of Infrared Technology for Non Destructive Testing*; Wiley: New York, NY, USA, 2001; p. 684.
- [21] Girshick, R. *Fast r-cnn*. In *Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015*; pp. 1440–1448, doi:10.1109/ICCV.2015.169.
- [22] Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv* 2017, arXiv:1704.06857.
- [23] Kononenko, O.; Kononenko, I. *Machine Learning and Finite Element Method for Physical Systems Modeling*. *arXiv* 2018, arXiv:1801.07337.
- [24] Beskos, D.E. *Boundary element methods in dynamic analysis*. *Appl. Mech. Rev.* 1987, doi:10.1115/1.3149529.
- [25] Chen, Z.; Xu, Y.; Zhang, Y. A construction of higher-order finite volume methods. *Math. Comput.* 2015, 84, 599–628.
- [26] Garrido, I.; Lagüela, S.; Sfarra, S.; Arias, P. Development of Thermal Principles for the Automation of the Thermographic Monitoring of Cultural Heritage. *Sensors* 2020, 20, 3392, doi:10.3390/s20123392.
- [27] Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* 2008, 77, 157–173.
- [28] Robertson, S. A new interpretation of average precision. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore, 20–24 July 2008*; pp. 689–690., doi:10.1145/1390334.1390453.
- [29] Ahmad, J.; Akula, A.; Mulaveesala, R.; Sardana, H.K. Probability of Detecting the Deep Defects in Steel Sample using Frequency Modulated Independent Component Thermography. *IEEE Sens. J.* 2020, doi:10.1109/JSEN.2020.3021118.
- [30] Rezatofighi, H.; Tsoi, N.; Gwak, J.Y.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of*

the IEEE Conference on Computer Vision and Pattern Recognition, Longbeach, LA, USA, 16–20 June 2019; pp. 658–666, doi:10.1109/CVPR.2019.00075.

- [31] Henderson, P.; Ferrari, V. End-to-end training of object class detectors for mean average precision. In *Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November, 2016*; Springer: Cham, Switzerland, 2016; pp. 198–213.
- [32] Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoftcoco: Common objects in context. In *Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014*; Springer: Cham, Switzerland, 2014; pp. 740–755, doi.org/10.1007/978-3-319-10602-1\_48.
- [33] Manzano, C.; Ngo, A.C.Y.; Sivaraja, V.K.S.O. Intelligent infrared thermography inspection of subsurface defects. Available online: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11409/114090V/Intelligent-infrared-thermographyinspection-of-subsurface-defects/10.1117/12.2558958.short> (accessed on 26 February 2021).
- [34] Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* 2015, 5, 1.
- [35] Deng, X.; Liu, Q.; Deng, Y.; Mahadevan, S. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Inf. Sci.* 2016, 340, 250–261. *Big Data Cogn. Comput.* 2021, 5, 9 21 of 21
- [36] Benenson, R.; Mathias, M.; Timofte, R.; Van Gool, L. Pedestrian detection at 100 frames per second. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012*; IEEE: New York, NY, USA, 2012; pp. 2903–2910, doi:10.1109/CVPR.2012.6248017.
- [37] Sitaram, S., Dessai, A. Classification of Cervical MR Images Using ResNet101. Available online: [https://www.ijresm.com/Vol.2\\_2019/Vol2\\_Iss6\\_June19/IJRESM\\_V2\\_I6\\_69.pdf](https://www.ijresm.com/Vol.2_2019/Vol2_Iss6_June19/IJRESM_V2_I6_69.pdf) (accessed on 26 February 2021).

## **Part III. Defects visibility enhancement and analysis by deep learning algorithms in Infrared thermography**

The following chapter will present one submitted article concerning the exploration of deep learning approach in infrared thermography applied in Non-Destructive Testing & Evaluation. The results of this study were accepted by the Journal of Nondestructive Evaluation in August, 2021.

### **General explanation:**

Due to experimental noise and non-uniform heating in regular inspection from industry, the defective pixels are always difficult to observe through the original thermal sequence based on experimental conditions and the properties of materials. As a result, the data processing and feature extraction method haven played an important role in IRT defect detection and identification which influences the accuracy and detection results.

In the study of this chapter, we will mainly concentrate on further reduce the noise influences and enhanced the objectives in the IR images by a deep learning extraction model. Data augmentation (GANs network) can be able to enlarge the diversity of data and extract features from a small dataset of IR data to improve the performance of the detection model. The technique will interpret how deep learning algorithms extract more information from IR thermography for Nondestructive testing which enables to detection of deeper subsurface defects for which IR signals are limited, etc.

# Chapter 5 Defect detection and visibility enhancement analysis using Partial Least Square-Generative Adversarial Networks (PLS-GANs) in Pulsed Thermography

*The results of this study were accepted by the Journal of Nondestructive Evaluation in August, 2021.*

## 5.1 Résumé

Avec les récents développements en intelligence artificielle, l'apprentissage profond (réseaux de neurones convolutifs) a été étudié pour la gestion de la qualité des matériaux par thermographie infrarouge (IRT). Dans ce travail, nous présentons une approche pour traiter efficacement les données thermiques afin d'améliorer la performance de détection des défauts de l'IRT. Un algorithme d'amélioration des défauts basé sur des réseaux adversariaux génératifs (GAN) a été conçu et mis en œuvre pour analyser qualitativement et améliorer la visibilité des défauts en se basant sur l'augmentation des données à partir d'une approche d'apprentissage profond. Nous avons mis en œuvre une méthode de réduction de la dimension des données basée sur la thermographie des moindres carrés partiels (PLST) fusionnée avec les réseaux GAN (PLS-GAN) pour obtenir une extraction et une visualisation interprétables des caractéristiques, et nous avons comparé les résultats avec les données thermiques de la thermographie pulsée (PT) afin d'évaluer l'efficacité de l'algorithme proposé. En appliquant les PLS-GAN, un petit ensemble de données thermiques peut être capable d'élargir la diversité des données afin d'améliorer la performance du modèle de détection. Les résultats expérimentaux ont été illustrés empiriquement sur les spécimens de référence : Polymères renforcés de fibres de carbone (CFRP). En conséquence, les résultats expérimentaux de détection sur les CFRP ont démontré la faisabilité de la méthode PLST-GANs.

### Contributing authors:

**Qiang Fang (Ph.D. Candidate):** the experiment planning, data collection, data analysis, designing and implementing the algorithm. Moreover, testing their accuracy and robustness throughout the process and writing the manuscript.

**Clemente. Ibarra. Castanedo:** the experiment planning, revision and correction of the manuscript.

**Duan Yuxia:** assisting the experiment planning, preparation

**Jorge Erazo-Aux:** assisting the experiment planning, preparation

**Iván Garrido:** assisting the experiment planning, preparation

**Xavier Maldague: (The research director):** supervision, revision and correction of the manuscript.

**Other contributors: Annette Schwerdtfeger** (research officer): manuscript preparation.

# Defect enhancement and images noise reduction analysis using Partial Least Square-Generative Adversarial Networks (PLS-GANs) in Pulsed Thermography

Qiang Fang 1\*, Clemente Ibarra-Castanedo1, Duan Yuxia2, Jorge Erazo-Aux3,4, Iván Garrido5, and Xavier Maldague1\*

1 Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Université Laval, 1065, av. de la Médecine, Québec, QC G1V 0A6, Canada;

2 School of Physics and Electronics, Central South University, 932 Lushan South Road, Changsha, Hunan 410083, China Abstract

3. Escuela de Ingeniería Eléctrica y Electrónica, Universidad del Valle, Cali 760032, VA, Colombia;

jorge.erazo@correounivalle.edu.co or jherazo@admon.uniajc.edu.co

4. Facultad de Ingeniería, Institución Universitaria Antonio José Camacho, Cali 760046, VA, Colombia

5. GeoTECH Group, CINTECX, Universidade de Vigo, 36310 Vigo, Spain; ivgarrido@uvigo.es (I.G.);

\* Correspondance: qiang.fang.1@ulaval.ca (Q.F.); Xavier.Maldague@gel.ulaval.ca (X.M.)

## 5.2 Abstract

With the recent developments in artificial intelligence, deep learning (convolutional neural networks) has been investigated for the quality management of materials by infrared thermography (IRT). In this work, we present an approach to efficiently process thermal data in order to improve the defect detection performance of IRT. A defect-enhanced algorithm based on Generative Adversarial Networks (GANs) has been designed and implemented qualitatively analysis and improve defects visibility based on data augmentation from a deep learning approach. We implemented a data dimension reduction method based on Partial Least Square Thermography (PLST) merged with GAN Networks (PLS-GANs) to achieve interpretable feature extraction and visualization, and compare result with thermal data of Pulsed Thermography (PT) in order to evaluate the efficacy of the proposed algorithm. By applying PLS-GANs, a small dataset of thermal data can be able to enlarge the diversity of data in order to improve the performance of the detection model. The experimental results were empirically illustrated over the benchmark specimens: Carbon Fiber Reinforced Polymers (CFRPs). Consequently, the experimental detection results on the CFRPs demonstrated its feasibility of the PLST-GANs method.

**Keywords:** Non-Destructive Testing Methods; Deep learning; Pulsed thermography; Defect enhancement; Generative Adversarial Networks (GANs); Partial Least Square Thermography (PLST); Thermal images noise reduction; Data augmentation and generation;

## 5.3 Introduction

Non-destructive evaluation (NDE) [1] consists of a group of methods used for quality assessment of materials in industrial fields such as aerospace due to its non-destructive and comprehensive approach. Infrared thermography (IRT) [2] has been widely used to evaluate



subsurface defects and hidden structures for the quality control of materials due to its rapidity of inspection, the relatively low cost and straightforward operation of these NDT techniques. However, due to experimental noise and non-uniform heating [3], the defective pixels are more difficult to observe through the original thermal sequence based on experimental conditions and the properties of materials. As a result, the data processing and feature extraction method [4] has played an important role in IRT defect detection and identification which influences the accuracy and detection results.

Previously, several signal processing techniques have been successfully applied to IRT for reducing the noise and improving defect visibility. For example, the classical principal component thermography (PCT) [5] is to project the raw inspection data onto a series of orthogonal bases. Through this method, noise and redundant information in the raw inspection data can be effectively reduced. The first few empirical orthogonal functions (EOFs) highly correlate to the presence of defects and can indicate defects more clearly. Pulsed phase thermography (PPT) [6] is another effective technique to improve defect visibility, where the raw inspection data is transformed from the time domain to the frequency domain. In the frequency domain, information of sinusoidal functions is used to construct new images, some of which indicate defects more distinctly than the raw thermograms. In thermographic signal reconstruction (TSR) [7], the new images are constructed using the first or the second-order time derivatives of the fitting data. In these new images, defects are usually indicated more clearly. Note that although many dimension reduction methods are available, all of the approaches have different advantages and disadvantages, and no method outperforms the others in all aspects. In addition, there are always few training samples obtained from these methods due to the limitation of the thermal imager. As a result, the evaluation effects from the thermal thermography have always been less accurate at the same time.

Nowadays, owing to their feature extraction ability, deep learning methods [8] have been increasingly applied to manufacturing processes. As the deep learning methods were adopted widely in the Non-destructive evaluation (NDE) research field recently, the convolutional neural network (CNN) [9] [10] [11] has the most significant influence on feature extraction for the research applied on the industrial applications. The CNN network has a strongly feature extraction capability to process the data from the manufacturing inspection, where one can fundamentally transfer the raw inspected data into the high level extracted features.

In the computer vision field, the generative adversarial network (GAN) [12] is one of the innovative CNN models to deal with the complex non-linear modeling issues with unsupervised learning. The ingenious training between the two main networks of GAN (Generator Network) and (Discriminator Network) can be compared to playing like a mini-max game [13]. Besides, the presence of sufficient accurate images, including exceptions (e.g. occlusion) could allow the

deep learning algorithm to effectively learn reliable features of interest. If the dataset does not contain enough images or contains too much noise, the learning algorithm may not be accurate enough or may even identify the wrong features. Therefore, in order to acquire enough dataset to train the network, identifying defects, the Generative Adversarial Network can be used to generate enough simulated thermal images to reveal defects variations [14] and enrich training data in unsupervised learning. To some extent, the generative data from GAN could be indistinguishable from the real data when the dataset is provided.

The main work focuses on a deep learning/image augmentation method that is beneficial for enhancing the visibility of defects and enlarging the diversity of the original set of images from IRT data analysis methods. Simultaneously, a Partial least square thermography (PLST) [15] is taken as a dimension reduction method to reduce the noise and extract the features. In optical pulsed thermography [16], limited thermal images were recorded in each experiment, while only containing a certain amount of feature information of defects. GAN is a powerful method for generating new data with a similar distribution of the original features.

As a result, a Convolutional based Generative Adversarial Network (GAN) for thermographic data analysis for defect detection in polymer composites has been proposed, where data augmentation can be applied for generating a diverse dataset from the original images. A series type of a Convolutional based GANs (WGAN [17]; DCGAN [18]) merged with the feature extraction method PLST in order to enhance each defect's visibility with respect to the signal-noise ratio (SNR) [19]. Specifically, this generative Partial least square thermography (GPLST) method enhances the defect visibility of composite polymer materials. A GAN network can be a model to generate simulated data in unsupervised learning for generating diverse images based on the raw databases. Although only a few studies have employed infrared thermography from the GAN network of composite materials, this innovative deep learning defect visibility enhancement algorithm can enhance the visibility of defects in sequential thermographic images.

In this work, A PLS-GANs data augmentation thermography model is introduced. GPLST is integrated with an unsupervised encoder-decoder structure based Generative Adversarial Network (DCGAN/WGAN) to transform raw data into higher-level features using nonlinear models. By employing this GAN-data augmentation strategy [20], more informative images are generated to enlarge the diversity of the original set of images. By applying deep convolution GAN [21] to extract useful features to eliminate noise, this strategy can enlarge the diversity and informative data of interpretable features. Consequently, the defect detection results can be visualized using a number of interpretable features. The defect detection performance and thermographic data visibility of thermographic data analysis can also be enhanced to some extent.

The remainder of this paper is structured as follows. Some preliminary principles are mentioned in Section 5.4. Section 5.5 proposes the PLS-GANs based image augmentation method for thermographic data analysis. The detailed analysis, metrics and training parameters are introduced in Section 5.6. In Section 5.7, the effective results generated from the series of GANs network (WGAN; DCGAN) and the results being processed by the PLS-GANs network is provided. Section 5.8 introduces a series of results, discussion and analysis. Finally, the conclusion is outlined in Section 5.9.

## 5.4 Preliminaries

### 5.4.1. Pulsed thermography and thermographic data

Figure 1 shows the basic principles of pulsed thermography [22]. The inspection by PT is based on the application – via radiation heat transfer – of a short and high-power thermal pulse to the specimen surface (input signal). The amount of thermal energy being absorbed by the surface of the sample will create a thermal front that propagates within the material until it reaches internal defects, which alter the heat diffusion flux. This interaction between the heat flux and internal anomalies - regions with different thermal properties in relation to their surroundings - produces dissimilar behaviors in terms of the temperature decay during the cooling process, which can be observed with an infrared camera (output signal). The deployment of this approach is carried out in a transient regime – in contrast to pulsed thermography, which is carried out under steady-state conditions – thereby allowing fast and straightforward data acquisition [23].

The thermal data obtained in the IRT experiment can be recognized as a three-dimensional (3-D) matrix  $n_x \times n_y \times n_t$ . Figure.2 shows that  $n_t$  frames of thermal sequences are obtained during the cooling period, where each frame is the size of  $n_x \times n_y$  pixels. Every pixel has a temperature decline curve measuring at a corresponding time instant. The temperature difference value between the regions of the defect and the non-defect provides detailed thermographic information for defect detection.

### 5.4.2. Materials

To evaluate the robustness and performance of the proposed algorithm, the detailed information of experimental specimens has been conducted as follows:

Three specimens(a)-(c) from Carbon Fiber-Reinforced Polymer (CFRP) were adapted in this work. Table 2 gives comprehensive descriptions of these samples. Each specimen has its own specific geometric graph with the Teflon square insertions of defects at different depths (Sample(a); Sample(b); Sample(c)). Specimens are square-shaped plates with the size of 30cm×30cm×2cm (10 ply carbon-fiber-reinforced polymer). These samples are all CFRP

specimens with natural artificial defects. Note that, each specimen has the same defect geometric distribution, however, these specimens are each a plane with a different trait (Sample(a): flat; Sample(b): curved; Sample(c): Flap) which influences the light illumination and thermal experimental response.

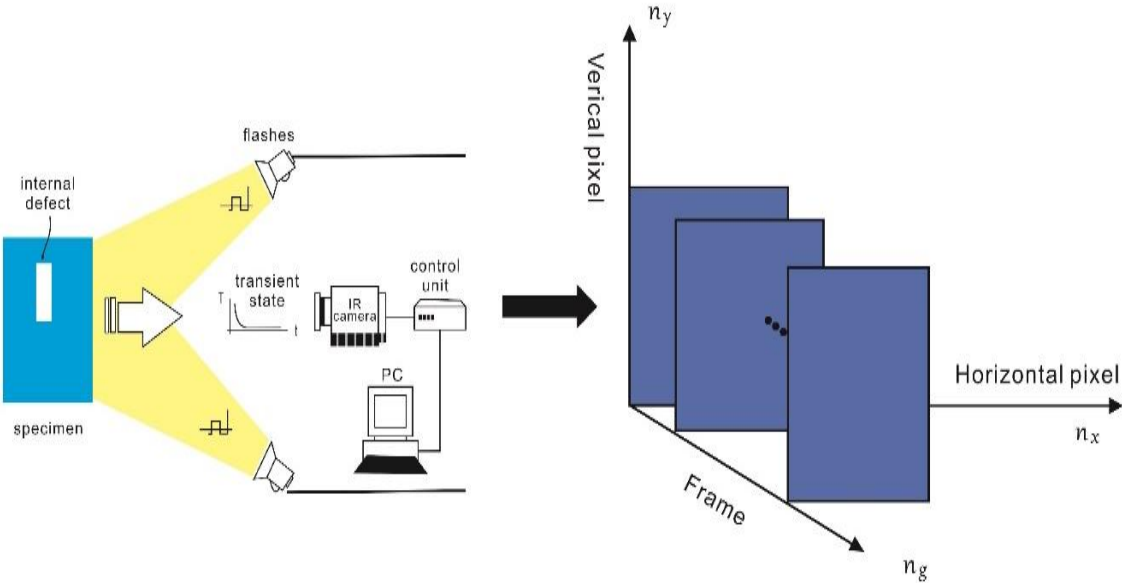


Figure 5. 1 Pulsed thermographic testing using optical excitation system and its three-dimensional thermographic data structure.

Number	Geometrics Specimen	Cross Section	Dimension of Specimen(cm)	Defect diameters(cm)
(a)			30cm*30cm	<p>Five different lateral sizes:  (3mm, 5mm, 7mm, 10mm, 15mm)  with five different depths (0.2mm; 0.4 mm; 0.6mm; 0.8mm; 1.0mm)</p>

(b)			30cm*30cm	<p>Different angle of defects  (<math>0^\circ</math>; <math>7.6^\circ</math>; <math>15^\circ</math>) with different equivalent diameters  (3.4mm, 5.6mm, 7.9 mm, 11.3mm, 16.9mm) and corresponding depths (1.0mm, <math>15^\circ</math>), (0.6mm, <math>7.6^\circ</math>), (0.2mm, <math>0^\circ</math>), (0.4mm, <math>7.6^\circ</math>), (0.8mm, <math>15^\circ</math>)</p>
-----	--	--	-----------	---

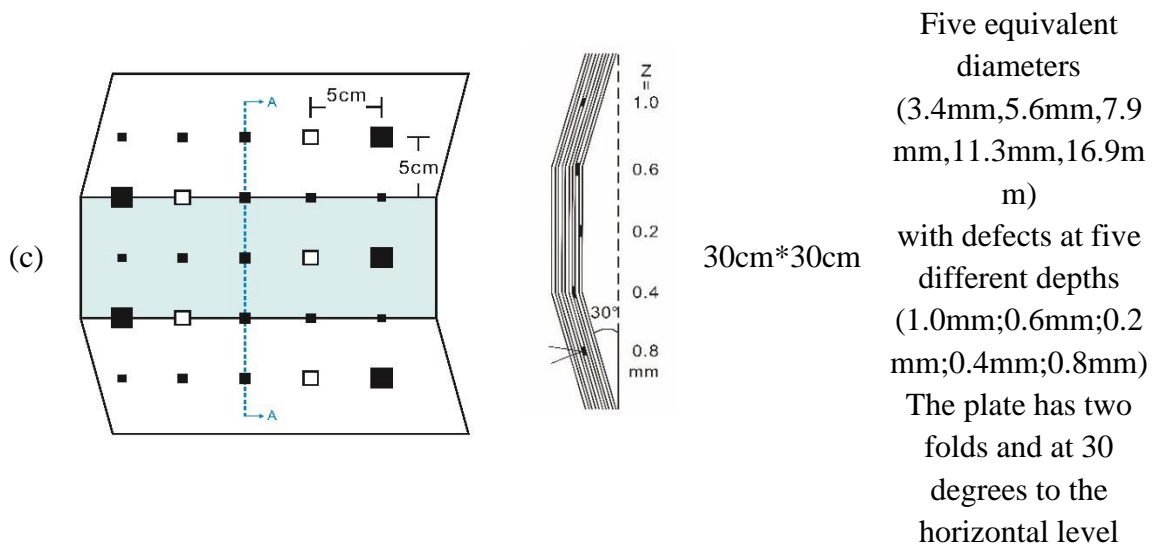


Table 5. 1 The geometrics and parameters distribution of three CFRP specimens

#### 5.4.3. Partial least squares regression (PLSR)& Partial least squares thermography (PLST)

Partial least squares (PLS) [24] is a dimension reduction algorithm to enable predictable modeling for the data. This algorithm also enables multivariate regression (MLR) [25] [26] [27] which is also correlating the information from one data matrix ( $W$ ) to another matrix ( $H$ ). As shown in Figure 2, the fundamental principle of the PLS algorithm and its difference with other linear regression methods [28] have been described.

Compared with most regression methods for dimension reduction which takes all independently x-values to form a linear combination of the feature, the PLS only combines a few linear components from the original x-values, and only these linear combinations will be considered in the whole regression procedure. Therefore, the irrelevant noise and information can be reduced and discarded, only the most relevant and key information will remain and be used for the regression. All variables can be projected down to a limited linear combination and the collinearity issue can be solved in order to obtain a more stable regression.

The PLS regression method could decompose the thermographic PT data sequence obtained during the cooling regime. During dimension reduction, the predictor matrices  $W$  ( $n \times N$ ) and predicted  $H$  ( $n \times M$ ) are decomposed into latent structures (a combination of loadings, scores, and residuals) during an iterative process. The latent structure corresponding to the variation matrix  $H$  is extracted and explained by the latent structure from  $T$  and  $U$ . The underlying model can be described as shown below [26]:

$$W = TP^T + G \quad (5.1)$$

$$H = UQ^T + K \quad (5.2)$$

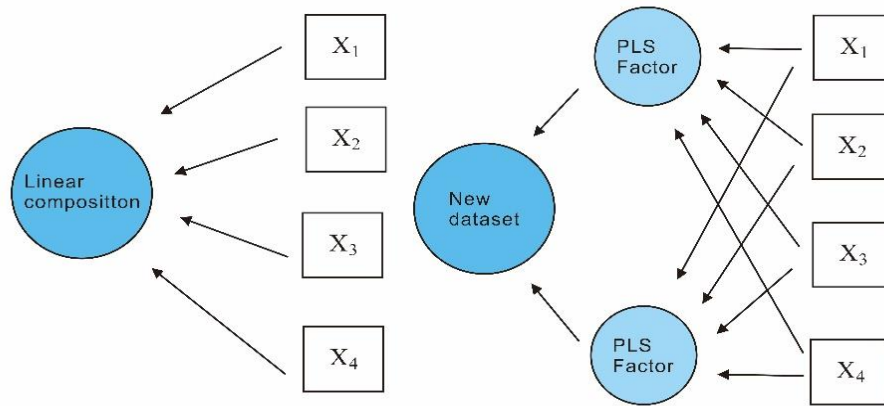


Figure 5. 2 Conceptual illustration of PLS and its comparison with classical linear regression methods

In Equation (5.1) and (5.2),  $T$  ( $n \times a$ ) and  $U$  ( $n \times a$ ) is defined as the score matrix, and its elements are denoted by  $t_a$  ( $a = 1, 2, \dots, A$ ). The scores can be considered as a small set of underlying or latent variables responsible for the systematic variations in  $W$ . The matrices  $P(N \times a)$  and  $Q(M \times a)$  are called loadings (or coefficients) matrices and they describe how the variables in  $T$  relate to the original data matrices  $W$  and  $H$ . Finally, the matrices  $G$  ( $n \times N$ ) and  $K(n \times M)$  are named as residuals matrices which represent the noise or irrelevant information from  $W$  and  $H$  respectively. The feature extraction and components analysis are carried out through the decompositions between  $W$  and  $H$  so as to maximize the covariance between  $T$  and  $U$ .

In NDT, the Partial least squares thermography (PLST) is defined as the same format of PLSR to decompose the raw data into a series of extracted feature components. Unlike PCT (Principal Component Thermography), the latent variables are extracted by a compromised method between the amount of variance summarized by the scores and the degree in PLST to which those scores are correlated with. Due to the co-focus on prediction from PLST, confounding of the targeted objective with other sources of variation in the data is usually more convenient in PLST than in PCT. Therefore, the PLST scores arise as a better predictor of the response than PCT scores.

#### 5.4.4 Generative Adversarial Network (GAN)

##### a. Basic principle of GAN

Generative Adversarial Network (GAN) is proposed by (Goodfellow et al) [12] to be used as a model to generate high-quality synthetic data in deep learning. To some extent, the generative data from GAN could be indistinguishable from the real data when the dataset is provided. As a deep structured generation neural network, GANs play a crucial role in machine learning due to the unique implementation step for confrontation training.

On the other hand, for data generation, the presence of sufficient accurate images, including exceptions (e.g. occlusion) could allow the deep learning algorithm to effectively learn reliable features of interest. If the dataset does not contain enough images, or contains too much noise, the learning algorithms may not be accurate enough or may even identify the wrong features. Therefore, in order to acquire enough dataset to train the network, identifying defects, the Generative Adversarial Network can be used to generate enough synthetic thermal images to reveal defects by enriching training data.

Based on the mathematical description, the adversarial training of GAN [29] [30] [31] can be described as shown in the equation (5.3) in the below. The GAN network consisted of a generator  $G$  and a discriminator  $D$ . During the first step of training, a vector  $z$  is sampled from the prior Gaussian distribution  $P(z)$ , then a generator network  $G$  is designed to learn the feature from the distribution  $P(z)$  through the vector  $z$  randomly. This generated network projects a new output data  $G(z)$ . Another network, which is defined as discriminator  $D$  has the responsibility to distinguish the true data (from training distribution) and fake input (from the generator  $G$ ).  $V(D, G)$  is a loss function for GAN to achieve the min-max game.

In addition,  $x$  represents the real thermal data.  $D(x)$  represents the probability  $x$  to be discriminated by Discriminator  $D$  as the real data;  $D(G(z))$  represents the probability of the  $D$  network to discriminate the generated data from the  $G$  network to see how similar it is with the real data.  $P_{z(z)}$  is the distribution of noise vector  $z$  and  $E$  represents the Expectation value. What the  $G$  network can do is to generate real data as much as possible to deceive the



judgment of the D network. In contrast, the D network is to improve its ability to distinguish between the real and generated data. Once the  $D(x)$  and  $D(G(z))$  values are equal to 0.5, these two networks (G; D) achieve equilibrium [12].

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (5.3)$$

In addition, the whole procedure of the GAN thermal image generation strategy is an unsupervised learning procedure which can be summarized from the Figure 5.5(a)-(c) in the below.

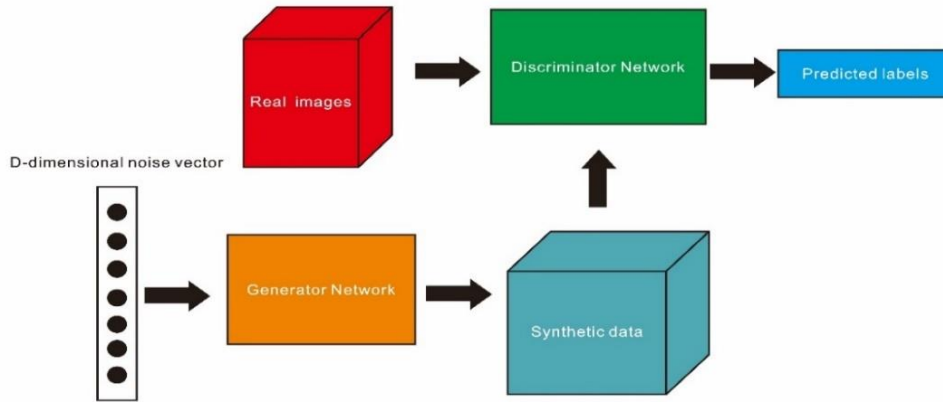
As shown in Figure 5.5 (a) the main GAN image generation strategy, a random D-dimensional noise vector was fed into the Generator Network G (affine transformation), then a series of fake thermograms are generated. In this case, the discriminator D can be treated as a binary classifier to predict corresponding labels for the fake images generated from random noise vectors in the G network and realistic images obtained from experimental pulsed thermography.

In the beginning of each training iteration, the initialization of the generator G and discriminator D has been activated. Then, during the training of the Discriminator network (Figure 5.5(b)), the generation G will be fixed while the discriminator D is updated in order to strengthen its distinguishing capability. The discriminator starts to learn to distribute the different scores to real objects and generate objects. On the other hand, during the training of the Generator (Figure 5.5(c)), the discriminator D will be fixed while updating the generator G (update the parameters of Gradient Ascent), the generator G enhances its capability to generate images similar to real images in order to obtain enough scores judged from the discriminator. During the training of the Generator, the Generator G and discriminator D can be treated as an integrally neural network (first several layers for the Generator; last several layers for the Discriminator; in the middle of hidden layers to be able to obtain generated images).

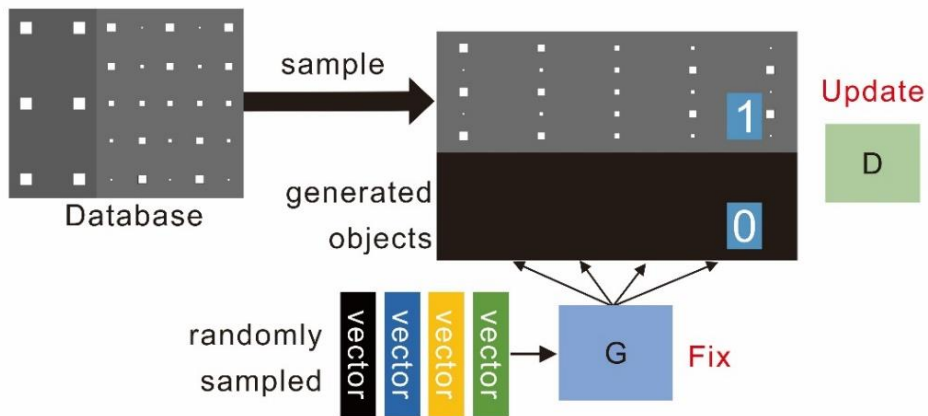
Therefore, one can make an analogy and compare the relationship between the main networks of GAN to a mini-max game as indicated in Figure 5.3 (c) which is played between G (the Generator Network) and D (the Discriminator Network) [32]:

- During the training of the discriminator network, the discriminator learns to assign high scores to real objects and low scores to generate objects;

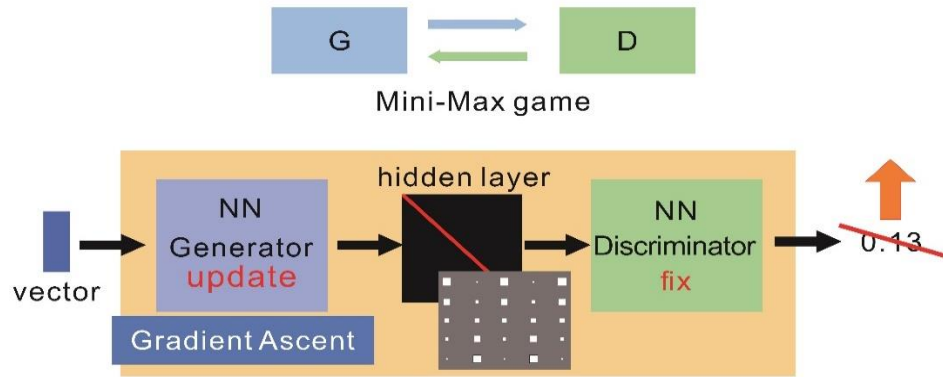
- During the training of the Generator network, the Generator is optimized to create realistically data to “fool” the discriminator. The Discriminator also evolves in its ability to precisely distinguish true and fake data.



(a) The main GAN Images generation strategy



(b) Training of the Discriminator network



(c) Training of the Generator network

Figure 5. 3The implementation steps for GAN network (a)-(c)

## 5.5 Proposed Methodology

### 5.5.1 Deep Convolutional Generative Adversarial Network (DCGAN) & Wasserstein Generative Adversarial Network (WGAN)

In this work, two types of modified versions of GAN networks (DCGAN; WGAN) have been introduced to discuss their capability of defect visibility enhancement when merged with the PLST algorithm for thermal data analysis.

#### A. Deep Convolutional GAN(DCGAN)

A Deep convolutional generative adversarial network (DCGAN) [33] is an extended version of GAN network (an unsupervised learning method). Due to its own capability of feature extraction, DCGAN utilized as a thermographic data augmentation modeling for enhancing the visibility of defects during the detection in composites polymer (CFRP; GFRP), which also has the capability to learn the multi-hierarchy representative feature extracted from the thermography sequences by the Discriminator D and Generator G of in the DCGAN network.

Compared with the original GAN, the detailed structure of the Generator (a) and the Discriminator (b) from the DCGAN as indicated in Figure 5.4, all the max-pooling layer has been replaced with convolutional strides. DCGAN adapted the transposed convolutional for upsampling and eliminated fully connected layers. It uses batch normalization in each layer (except the discriminator's input layer and the generator's output layer). In addition, a Leaky-ReLU has been adopted from the discriminator. The generator produces the images based on a random 100-dimensional vector. Its structure contains four convolutional layers in total which could be able to generate a series of synthetic thermograms. The discriminator

maintains the symmetrical structure with the generator but its output is a number  $X$  between 0 to 1 by the sigmoid function for the possibility to discriminate between the real or fake images.

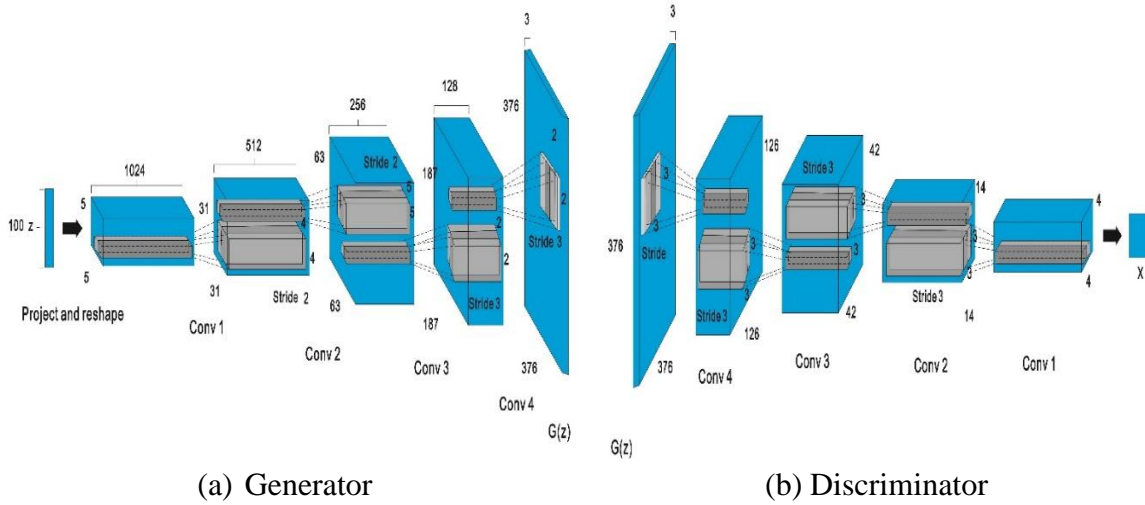


Figure 5. 4 The architecture of DCGAN

## B. Wasserstein -GAN(WGAN)

As shown in the research literature, the training process of GAN is quite challenging. Specially, the DCGAN may not easily converge and the mode may collapse (the generator tends to produce a high similarity of the samples) as well.

In this work, an unsupervised generation model (Wasserstein GAN) [34] has been introduced for the comparison with DCGAN especially for feature extraction and data augmentation of defect visibility enhancement.

The WGAN is a new type of GAN which has the capability to overcome the training difficulty from the regular GAN network and to overcome the potential issue of mode collapse. In addition, the Earth Mover (EM) distance [35] optimum function of WGAN can be defined as follows:

$$W(P_{data}, P_G) = \underset{D \in 1\text{-lipschitz}}{\text{Max}} \{E_{x \sim P_{data}} [D(x)] - E_{x \sim P_G} [D(x)]\} \quad (5.4)$$

The  $W(P_{data}, P_G)$  in equation (5.4) represents the evaluation of Wasserstein distance [36] between the real data distribution  $P_{data}$  and the generated data distribution  $P_G$ . Meanwhile, the learning procedure of WGAN can be regarded as a regression task. In equation (5.4), based on its mathematical expression, the discriminator obtains a larger output values when its sample arises from a real data distribution. In contrast, the discriminator obtains a smaller

value when its sample arises from the generated data distribution. In equation (5.4), the 1-lipschitz term [37] is a constraint term to control the discriminator at a certain range of infimum in order to keep it converge (keep the discriminator “smoother” without intractable infimum). Therefore, the distribution  $x$  value should belong to the region from the 1-Lipschitz term function as shown below:

$$\|f(x_1) - f(x_2)\| \leq \|x_1 - x_2\| \quad (5.5)$$

where in equation (5.5), the  $x_1$  and  $x_2$  represent two different inputs, and  $f(x_1)$  and  $f(x_2)$  represent two different outputs. So  $\|x_1 - x_2\|$  and  $\|f(x_1) - f(x_2)\|$  represent the absolute change of the input and output. Therefore, the equation (5.5) means the output change must smaller or equal to the input change for the discriminator training in the WGAN (to keep the discriminator smooth enough).

Based on the structure as shown in Figure 5.5, the WGAN has a similar structure to that of the DCGAN. First, the discriminator, with a  $376 \times 376$ -pixels image samples (raw or generated images) as the input, then projected feature will go through a certain number of convolutional layers; batch normalization; leaky rectified linear unit (LRelu). The generator has a nearly symmetrical structure with the discriminator (autoencoder format), which first inputs a certain dimension of the noise vector from a normal distribution, then generates the feature vector feeding it through certain deconvolutional layers to restructure itself to be the original size ( $376 \times 376$  pixels).

The overall training workflow of the regular GAN and WGAN networks is illustrated in Figure 5.6 (a) and Figure 5.6 (b), respectively. The main reason for introducing WGAN is to reduce the training difficulty and potential issue of model collapse. As the regular GAN in Figure 5.6 (a) indicated, the loss function is a binary classification (judge), but in Figure 5.6 (b), the loss function is an EM distance, which is a regression task rather than a classification problem during the training process.

The WGAN modifying strategies can effectively improve the GAN to obtain higher efficient performance during the training. As a result, WGAN can successfully overcome the issues caused by the unstable training due to an imbalance between the discriminator and generator. In addition, the WGAN has a reasonable and effective way to estimate the distance of the Earth-Mover (EM), which means the minimum cost of the transporting plan to convert the data distribution  $q$  to the data distribution  $p$ . In other words, during the data generation procedure of synthetic thermogram, WGAN has the capability of furtherly improving the training and then providing a high quality of generated thermal images [38].

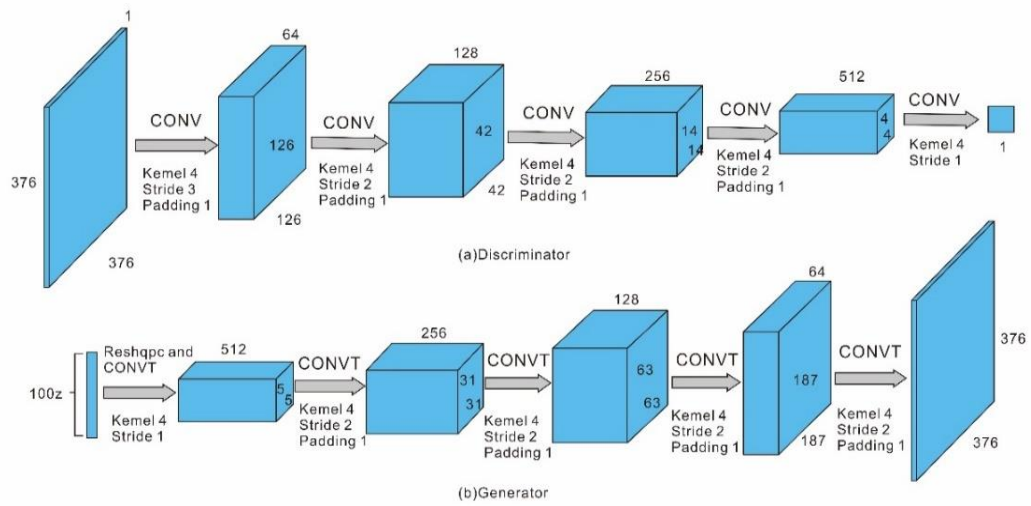
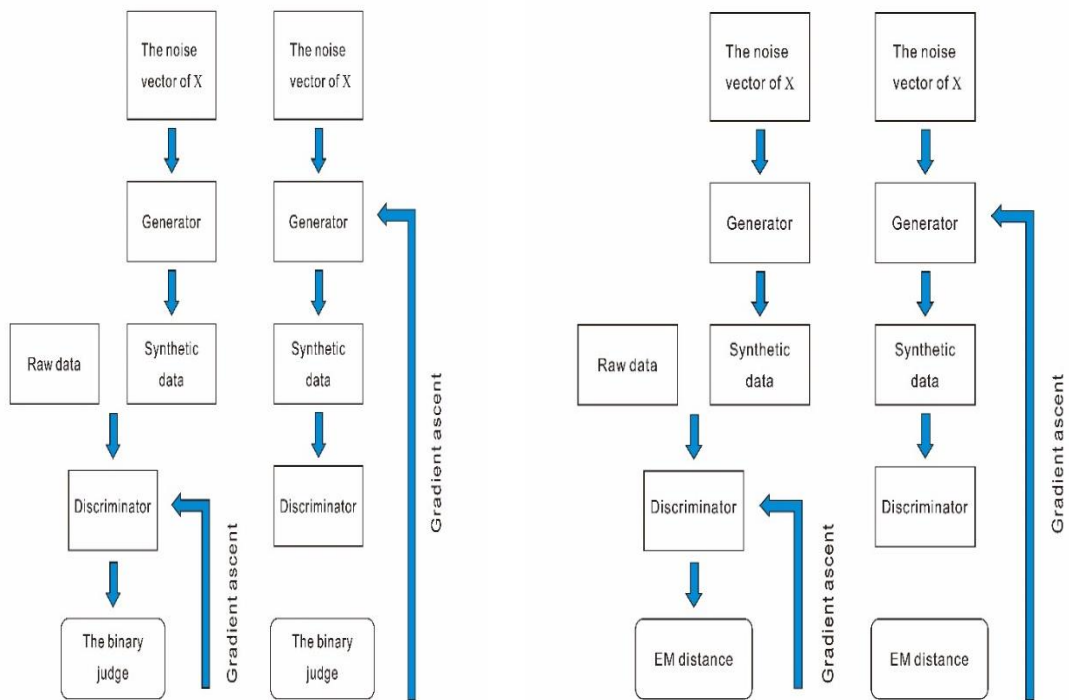


Figure 5. 5 WGAN generator architecture



(a) The training path of regular GAN(DCGAN) network; (b) The training path of regular WGAN network;

Figure 5. 6 The training procedure flowchart of WGAN

### 5.5.2 The general GAN-thermographic implementation steps

After the deployment of the two types of GAN network: DCGAN and WGAN, our basic training, implementation steps can be described as the following procedure (Part A, B, C). During this procedure,  $D(x'_i)$  is defined as a sigmoid function (0,1),  $\beta$  as a gradient ascent,  $P_{prior}(z)$  as the distribution, Initialize Parameter  $\theta_d$  of Discriminator  $D(x)$  and  $\theta_g$  of Generator  $G(x)$ ,

Part A. In each training iteration (learning Discriminator):

Step 1: Sample m examples  $\{X^1, X^2, \dots, X^n\}$  from data Distribution  $P_{data}(x)$ ;

Step 2: Sample m noise samples  $\{Z_1, Z_2, \dots, Z_n\}$  from the prior  $P_{prior}(x)$ ;

Step 3: Obtaining generated data  $\{X'_1, X'_2, \dots, X'_n\}$ ,  $G(X'_i) = G(Z_i)$ ;

Step 4: Update discriminator parameter  $\theta_d$  to maximiz

$$V' = \frac{1}{m} \sum_{i=1}^m \log D(x^i) + \frac{1}{m} \sum_{i=1}^m \log(1 - D(x'_i)); \quad (5.6)$$

$$\theta_d \leftarrow \theta_d + \beta \nabla V'(\theta_d); \quad (5.7)$$

Part B. In each training iteration (learning Generator):

Step 1: Sample another m noise samples  $\{Z_1, Z_2, \dots, Z_m\}$  from the prior  $P_{prior}(z)$ ;

Step 2: Update generator parameter  $\theta_g$  to maximize;

$$V' = \frac{1}{m} \sum_{i=1}^m \log(D(G(z^i))); \quad (5.8)$$

$$\theta_g \leftarrow \theta_g + \beta \nabla V'(\theta_g); \quad (5.9)$$

Part C. loop operation of Part A (Step 1-Step 3) and Part B (Step 1-Step 3) until the  $D(x)$  and  $D(G(z))$  values are equal to 0.5, then the two networks achieve equilibrium, and the whole training process is completed.

Note that the Wasserstein GAN has several points on its implementation steps from the algorithm which are different from regular GAN network implementation steps (DCGAN) as described below:

1. The discriminator of WGAN removed the sigmoid function in its last layer;

2. A new cost function equation (5.10) and equation (5.11) from WGAN corresponding to equation (5.6) in Part A (Step 4) and Part B (Step 2) equation (5.8) does not take  $\log$  in its loss function.

$$V' = \frac{1}{m} \sum_{i=1}^m D(x^i) - \frac{1}{m} \sum_{i=1}^m D(x'_i); \quad (5.10)$$

$$V' = -\frac{1}{m} \sum_{i=1}^m D(G(z^i)); \quad (5.11)$$

3. The absolute values of the discriminator parameters being truncated no more than a fixed constant  $c$  after each update (weight clip) [39].

### 5.5.3 PLST-GANs Thermographic Framework

The PLS-GANs thermography framework consists of the PLST thermography data analysis algorithm and WGAN/DCGAN-based data augmentation method. In this work, we chose the state-of-the-art data analysis method PLST [40] for comparison of the performance with PLST-GAN data augmentation method.

In addition, the overall PLST-GAN-based thermography framework is described in Figure 5.7. A series of fake thermography are generated from the GAN, then merged with raw experimental data from the pulsed thermography. The restructuring thermal sequences combined from the fake thermograph and raw data being normalized by standard deviation data preprocessing. Then, the restructuring thermography is analyzed by the PLST algorithm. The final comprehensive feature is visualized in the real thermal images from the extraction. Further, to automatically detect the defects and achieve the defect visibility enhancement, the detailed implementation steps (data merging; data normalization; data visualization) of PLST-GAN can be illustrated in the steps below:



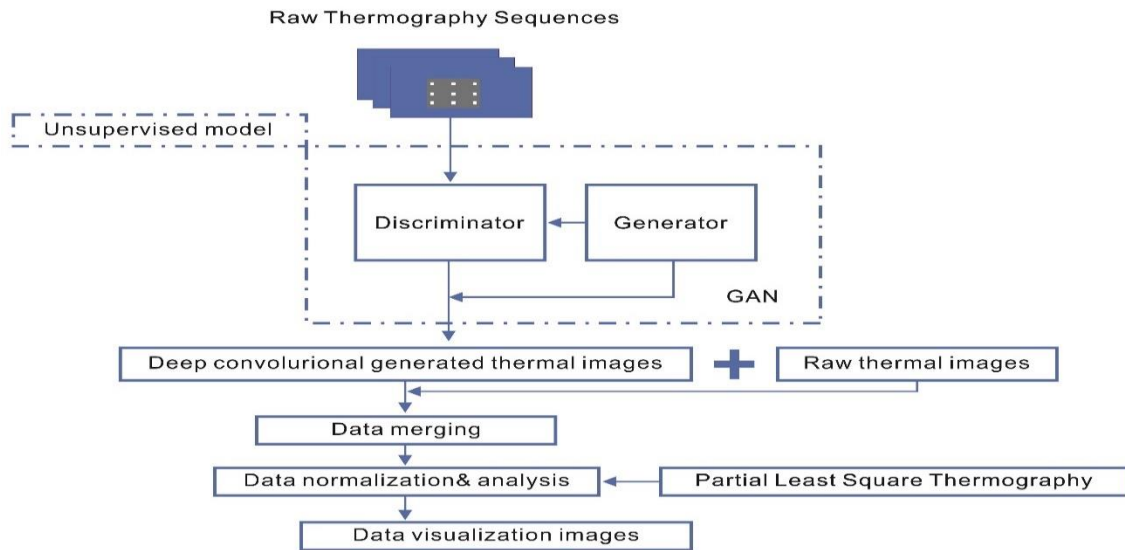


Figure 5. 7 Defect enhancement strategy with GAN model

#### A. Data merging

In this step, the thermal images outputted from the GAN network (WGAN/DCGAN) are merged with the original raw thermal images from the experimental pulsed thermography. In more detail, the original thermogram is inserted in front of the generated thermogram, then restructured into a new 3D matrix. Therefore, the pixels of the restructured thermograph temporal transient response and information have no specific physical meaning. The only key information is the spatial feature for identifying the defect regions. On the other hand, the PLST-GAN data analysis algorithm processes with spatial information only and does not consider temporal information. Although merging with different sequences takes place (generated GAN images; experimental thermal sequences) and changing the order of the data matrix, this will not influence the results from defect detection.

In pulsed thermography (PT), the use of PLSR to PT data is achieved by decomposing the raw thermal data into multi-independent components of PLS. The dataset obtained during the PT inspection from the IRT is arranged as a 3d data matrix as indicated in Figure 5.8 ( $n_x \times n_y \times n_T$ ).  $n_T$  is the total thermal frame obtained from the cooling process of PT, where each frame is represented as  $n_x$  ( $x$  - axis)  $\times$   $n_y$  ( $y$  - axis) pixels. The value of each pixel represents the surface temperature at the corresponding position and also corresponds to a point of the temperature decay variation curve at a specific time instant. Therefore, the meaningful feature can be extracted based on the temperature decay difference between the defect and non-defect region.

Therefore, in order to perform thermal data under PLSR algorithms to decompose these data into PLS components, the whole 3D matrix (experimental thermograph) is reshaped into a two-dimensional matrix with  $(n_x \times n_y, n_g)$ , referred to as the folding process as shown in Figure 5.8. Simultaneously, the generated GAN images reshape into another two-dimensional matrix with  $(n_x \times n_y, n_t + n_g)$ , which has the same data structures with experimental thermograph. The two 3D matrixes reconstructed as a new 2-D unfolded matrix  $X$  with a size of  $(n_x \times n_y, n_t + n_g)$ , where the  $n_t$  is the frame number of the experimental thermograph, and the  $n_g$  is the frame number of the generated thermograph.

This 2D matrix may produce a high level of representative feature of thermal contrast to analysis the defect structure in composites and also improve the capability of defect visibility enhancement compared to other regular thermography procedures. In addition, in this algorithm, the predicted matrix  $H$  from session 5.4.3 is a column vector  $(n_t, 1)$  based on time series for the implementation of PLSR for Pulsed thermography.

#### B. Data normalization& Dimensionality reduction procedure

In order to further analysis, a merging matrix  $\mathbf{X} = [X_1, X_2, \dots, X_K]^T$  has been created, where the temperature response of a single thermogram is represented by the row vector from  $\mathbf{X}$ , while the temporal temperature evaluation of a single pixel represented by the column vector. A standard deviation normalization method is applied to the merging matrix  $X$  as the pre-processing method to reduce the influence of the uneven heating and noise background which would be beneficial for the extraction of meaningful features of the defect information. The normalization is based on each column, which focuses on the information of each pixel to subtract the mean value and divided by the standard deviation.

After the matrix  $X$  is normalized by the standard deviation, the PLST data analysis method will be applied to the matrices. Through the GAN-PLS model, the feature of defects will be extracted and analysed. The predicted matrix  $H$  obtained from the covariance matrix of  $X$  will contain the information for the visualization of defects and the visibility of enhancement for detects.

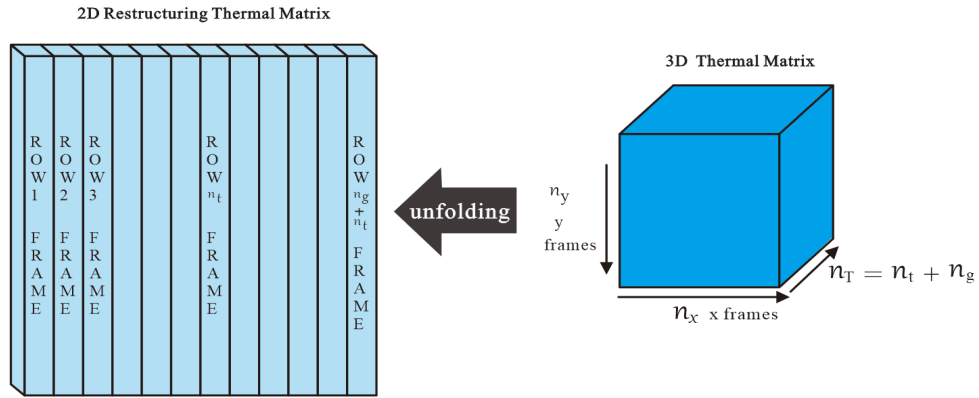


Figure 5. 8 Schematic representation of the transformation of the 3D thermal data into a 2D raster-like matrix.

### C. Data visualization

As mentioned in section 5.4.3, in PLST feature extraction equation (1)-(2), the PLS algorithm will find the compromised components:

- The value of dot product  $TP^T$  must illustrate the unfolding matrix  $W$  well;
- The score vector  $T$  should be related with the predicted matrix  $H$ ;

Statistically, unfolding matrix scores  $T$  will be the highest possible covariance with the predicted matrix  $H$  through the score vector  $U$ . Therefore, in order to analysis the results obtained from pulsed thermography (PT), the merged 2D matrix (raw data; generated data) is applied to the PLST through decomposition. Then the 2D unfolding matrix  $W$  will be converted into a series of PLS component combination  $P$ (predictor loading matrix) as shown in Figure 5.9.

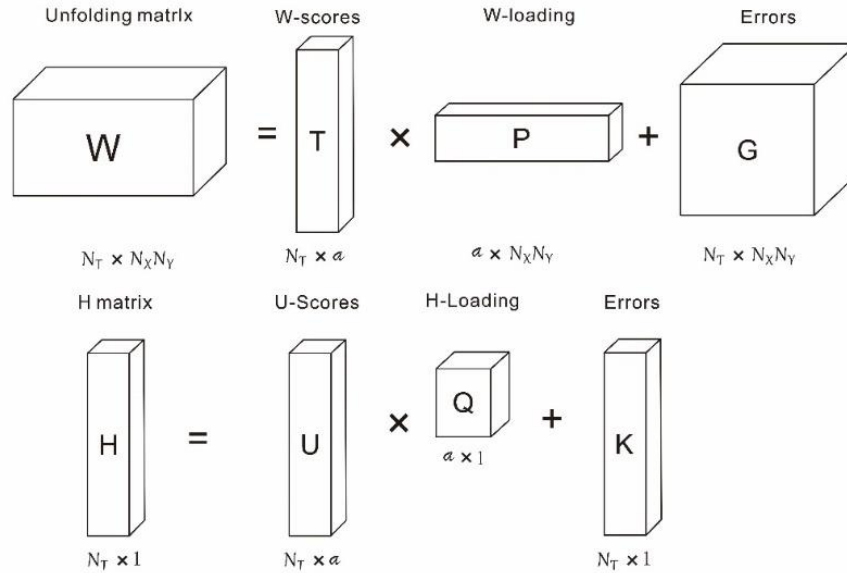


Figure 5. 9 Graphical representation and decomposition of pulsed thermography by PLST

Simultaneously, the whole procedure can be described as the PLS algorithm implementation is described below:

The data are adopted to apply PLS is a 3D matrix (thermal sequence) obtained from PT inspection of the carbon-fiber-reinforced-polymer (CFRP). This 3D sequence ( $n_x, n_y, n_T$ ) is normalized and converted into the  $W$  ( $n_T, n_x \times n_y$ ). (as mentioned in section B (3.3));

The decomposition of the predictor  $W$  is implemented by the nonlinear iterative partial least squares (NIPALS) [41];

The number of components from PLS is extracted based on the application of RMSE (root mean square error) [42] in the equation (5.12), where  $n$  is the number of samples,  $H_i$  is defined as the predictive value, the  $H_{i,ref}$  is the reference value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (H_i - H_{i,ref})^2}{n}} \quad (5.12)$$

After the application of the PLS algorithm, the matrix  $H$  is generated as a column vector ( $n_T, 1$ ) based on the time series. The predictor loading matrix  $P$  (see Fig. 5.9) is transformed into a 3D matrix, composed of ( $n_x, n_y, n_{loading\ vector}$ ).

In order to analyse the PLS-GANs results for the defect visibility enhancement, we selected and visualized certain corresponding PLS component images (2D images) based on this 3D converted matrix for the visualization. The obtained PLS loading vectors from the predictor loading matrix P need to be reconstructed into the thermal images in order to identify the defects. The PLS-GANs generate many PLS component images. But not every single image contains useful information and may still include noise.

In Figure 5.10, the overall PLS-GANs framework is illustrated. As shown in the right side of Figure 5.10, the DCGAN structure has been indicated which is an auto-encoder model [43] to achieve dimension reduction. There are two pathways in total to extract the features in this structure. The contract path extracts the input images into a lower dimension feature map (reducing the length and width of the original images). In contrast, an expansive path reconstructed the lower dimension features back into the larger size. Specially, DCGAN has three convolutional layers reducing the size of images to half compared to the previous layer (corresponding to the discriminator). Simultaneously three deconvolutional layers expand the size of feature maps twice compared with previous layers (corresponding to the generator). Besides, as it is discussed in section 3.1 (in Figure 5.5), the WGAN has a similar structure to that of DCGAN, but updates its unique loss function to keep its high feature extraction performance.

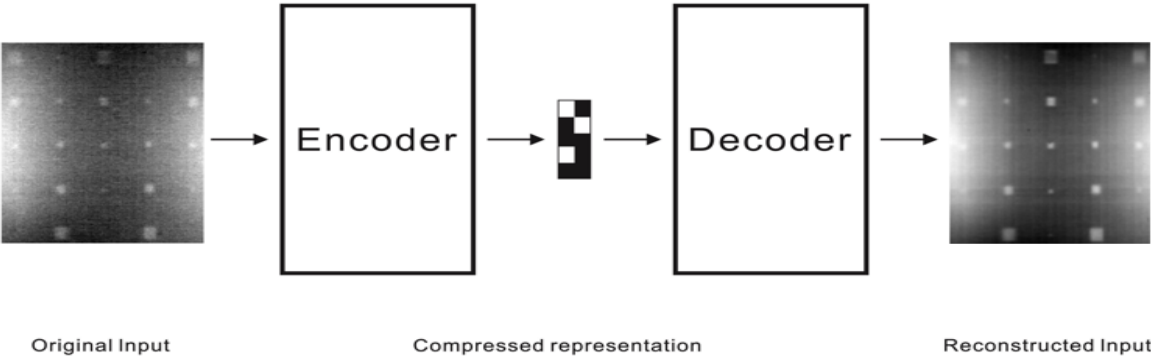


Figure 5. 10 feature extractor of encoder and decoder format convolutional neural network

Therefore, the structure of DCGAN\WGAN is a type of neural network designed to replicate the input feature(images) to the reconstructed feature. (DCGAN\WGAN) are implemented by compressing the input images into a latent-space representation [44] and then reconstructing to the output based on this representation feature. Such networks regularly both consist of two-part network as shown in Figure 5.10: 1. an encoder network: the input being compressed into a latent-space representation; 2. a decoder network: aiming to reconstruct the input from the hidden-space representation (latent-space representation) for feature extraction.

In detail, the encoder and decoder format network are fully represented by the convolutional layers from the WGAN/DCGAN structure. During the learning stage of the GANs network, assuming a noisy image is fed into the input of the GANs network through the training, then a generated image could be able to be outputted from the GANs model. If the output image is noisy, then GAN will be "punished" by its own loss function. As a result, during this whole process, the GANs model gradually and eventually learns to remove the noise and reconstructs less-noise images from the training.

Once the feature is obtained, the generated images will be merged with original data (data merging), then be normalized and processed by dimension reduction method (PLST) to generate final enhanced images as indicated in Figure 5.11 in the PLS-GANs framework.

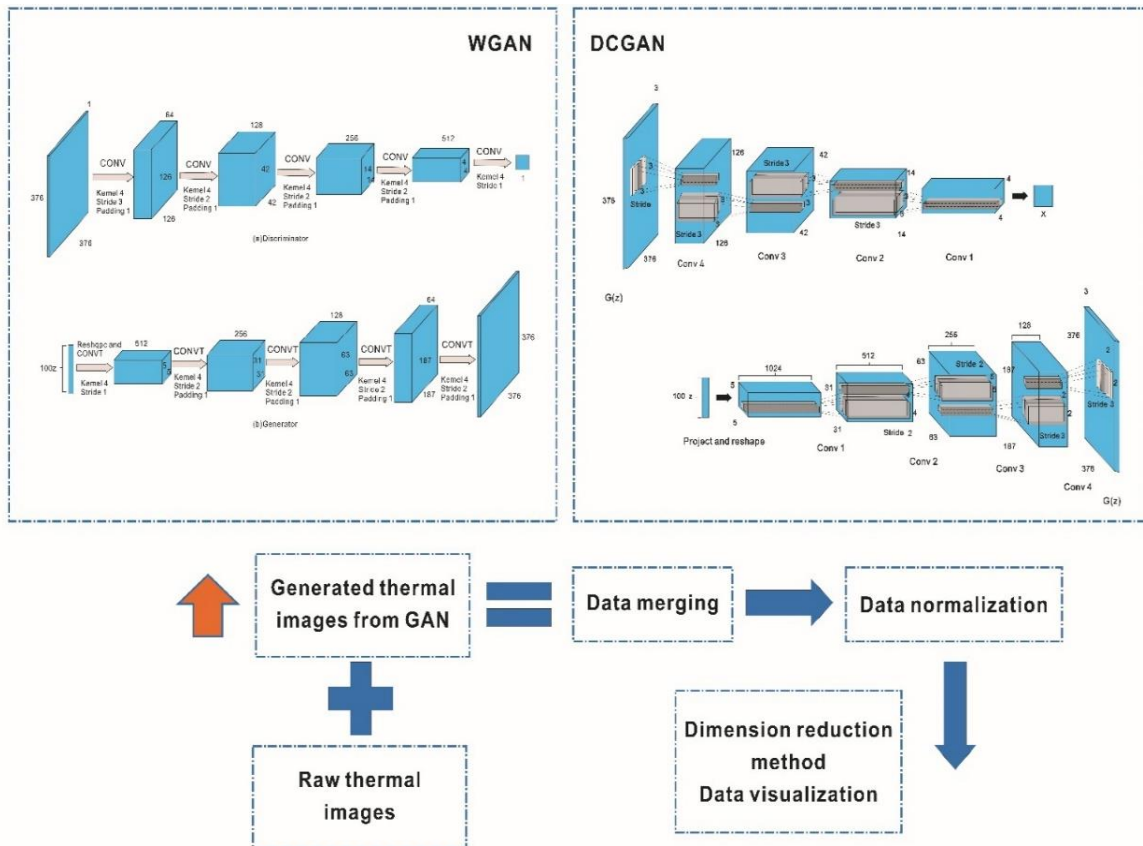


Figure 5. 11 Overall PLS-GANs framework

### 5.5.4 Evaluation metric and implementation details for defect visibility enhancement

#### a. GANs Training

The procedure for the GANs model training was set according to the following principles: 1 For parameter adjustment, it's based on the Pytorch deep learning framework; 2. The training

processing was conducted on a GeForce GTX2080TI for about 60 min; 3. The operating system is set as Ubuntu 16.04; 4. The framework of the learning model is set as Darknet. CPU: i9-9900k. Memory: 64GB, GPU: NVIDIA Ge-Force GTX1080TI.

The training process consists of a simultaneous Stochastic gradient descent (SGD); The learning rate is 0.002; The number of training epochs is 1000.

The optimization method of the WGAN network being adopted in this work is RMSProp. The optimization method of the DCGAN network being adapted is the Adam optimization with the parameter beta2=0.999.

In the DGAN/WGAN learning process, the whole process was trained with mini-batch stochastic gradient descent (SGD) with the size of 64 mini-batches/128 mini-batches.

#### b. Analysis metrics

In this section, the PLS-GANs model is investigated with three different geometric distribution CFRP specimens for the detect visibility enhancement. The evaluation metrics that are adopted here are the common Signal-to-Noise-Ratio (SNR) [45] and the Peak Signal-to-Noise Ratio (PSNR) [46] which can be expressed as follows:

##### (1) Signal-to-Noise ratio (SNR)

$$SNR = \frac{M_{def} - M_{in}}{\sigma_{in}} \quad (5.13)$$

SNR can effectively reflect the thermal contrast and defect feature based on the difference between the defect regions and non-defect regions. The higher SNR values indicate that the image contains more defect information and indicate a better performance of the learning model for the defect visibility enhancement. In equation (5.13),  $M_{def}$  represents the average pixel of the defect' regions and the  $M_{in}$  represents the average pixel of the boundary non-defect regions.  $\sigma_{in}$  represents the standard deviation of the pixel value in the non-defect regions.

##### (2) Peak Signal-to-Noise Ratio (PSNR)

In order to discuss the SNR at maximum signal contrast, the PSNR is also defined through the mean square error(MSE) (equation(14)),assuming that the image has the size of  $(m \times n)$  from the denoising image  $I$  (output images from PLS-GANs) and the noise image  $k$  (input images from PLS-GANs) as follows: (the coordinate value in each image is noted by  $i, j$  in these images)

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - k(i, j)]^2 \quad (5.14)$$

The PSNR (in dB) is defined as:

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) \quad (5.15)$$

The  $MAX_I$  is the maximum possible pixel value of the image (this is 255 when pixels using 8 bits represent the sample).

c. Research results and analysis- Learning curves and training procedure

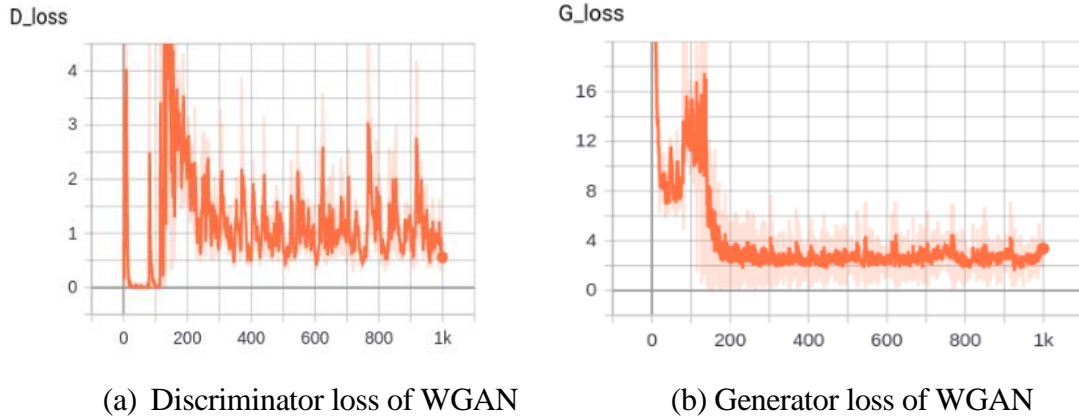


Figure 5. 12 The learning loss from Discriminator (a) and Generator (b)

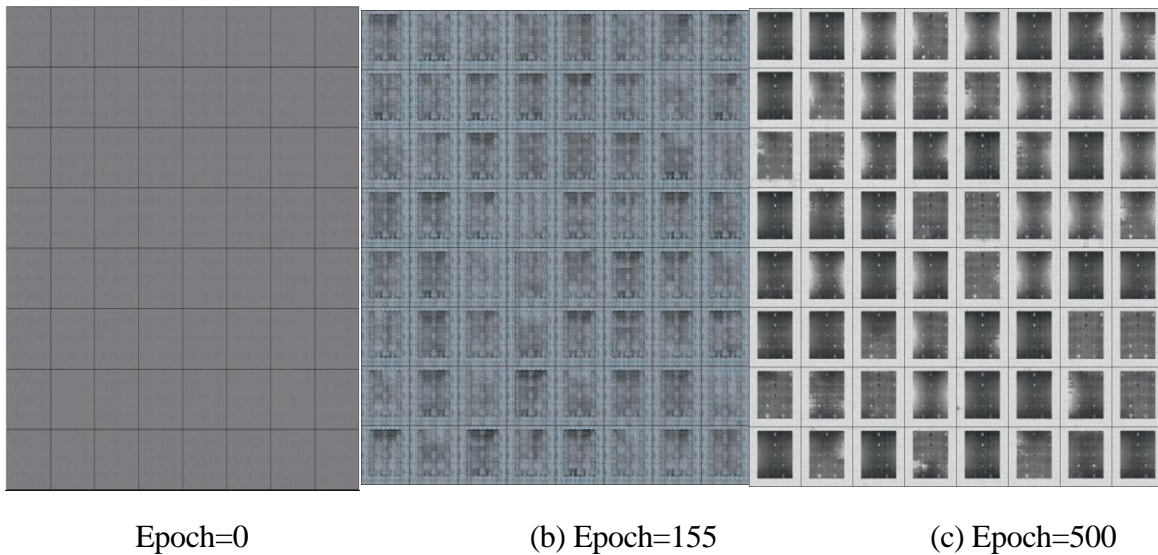


Figure 5. 13 The generated images 64\*64 from different epoch (a) epoch=0; (b)epoch=155;(c) epoch=500



The learning loss of the generator and the learning loss of the discriminator remain relatively balance as shown in Figure 5.12. The loss from the Discriminator increase during oscillation, then becomes stable at a value of around 0.5. The loss from the Generator oscillates at the beginning during the training, then stabilizes its value around 3.5. It is furtherly indicated the stable performance during the whole training procedure.

As shown in Figure 5.13, the generator generates 64\*64 images by random noise  $z$  at epoch = 0 (a); at epoch =155 (b); at epoch=500 (c). As the epochs increasing, the clearer thermal images are gradually generated. The key aspect of this algorithm is that after DCGAN achieves defect enhancement, the generated fake data is superimposed on the original image in order to suppress the noise. If the WGAN defect enhancement fails, then all subsequent analysis methods will fail. Therefore, after normalizing, by superimposing the WGAN-generated image on the original image at epoch=1000, the following image is generated in Figure 5.14, where one notes that the defect is indicated and enhanced from the images to some extent.

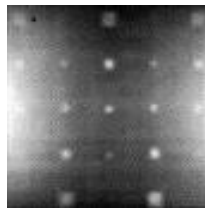


Figure 5. 14 The generated images at epoch=1000 (64\*64)

#### d. Feature extraction and defect enhancement from GANs

Notice that the DCGAN generates the defective feature graph by learning from the noise vector  $z$ . The images are generated do not have a temporal relationship as stated in experimental infrared thermography, where it is theoretically impossible to obtain a temporal connection. Even if it is a temporal association in the data, the final training is a graph-by-graph training and generation. Therefore, there is no temporal relationship.

As shown in Figure 5.9 and Figure 5.10, we displayed generated thermograms- the size of 64\*64 (during the initial stage of training). This indicated that the extracted thermal images from GAN contain less noise to some extent. After the GAN generation procedure, the DCGAN-generated thermal images (on the right side of Figure 5.9) and the WGAN-generated thermal images (on the right side of Figure 5.10) can be visualized. The defect information in both generated images is clearly highlighted, facilitating further defect recognition. In the specific frames from 17<sup>th</sup> – 48<sup>th</sup>, the generated thermal sequences have a more obvious clarity due to the training of DCGAN in Figure 5.15 and WGAN in Figure 5.16, compared with the raw thermal sequences. Simultaneously, the generated thermograms

from WGAN are slightly clearer and lighter than the thermograms from DCGANs especially between the 17<sup>th</sup> – 48<sup>th</sup> frames there are 64 frames in total).

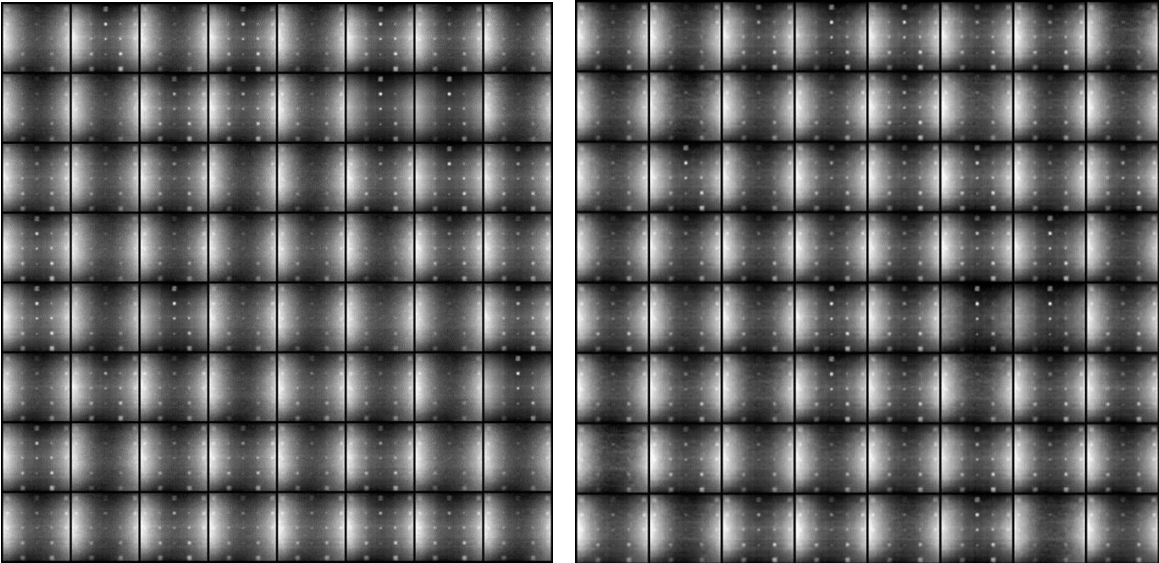


Figure 5. 15 Raw CFRP-A (left side) and Generated CFRP-A (right side) sequence thermal images from DCGAN

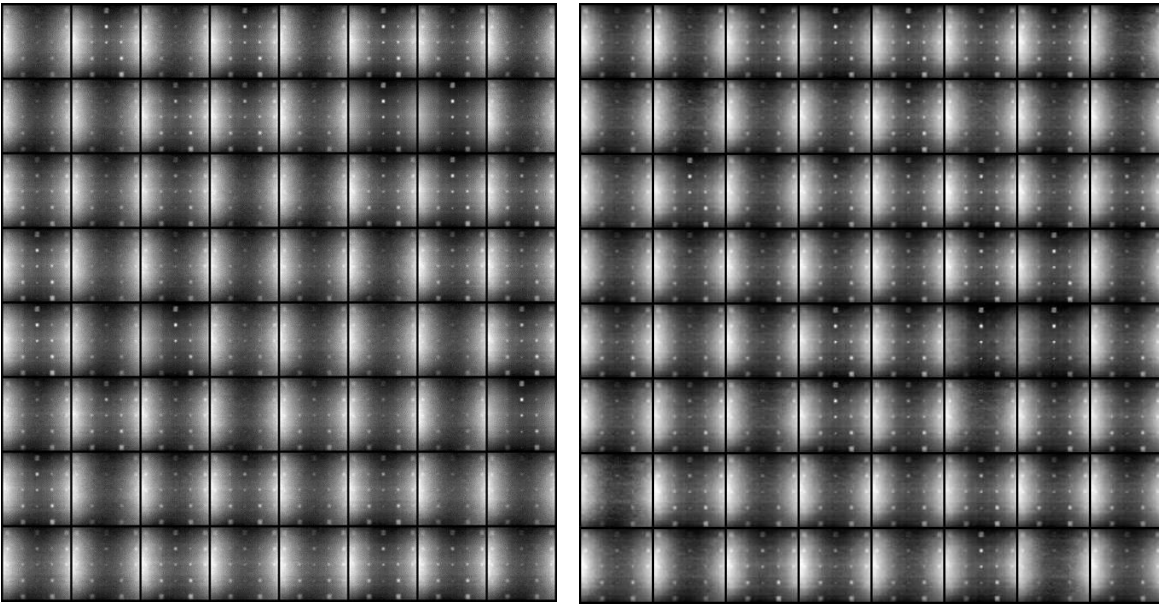
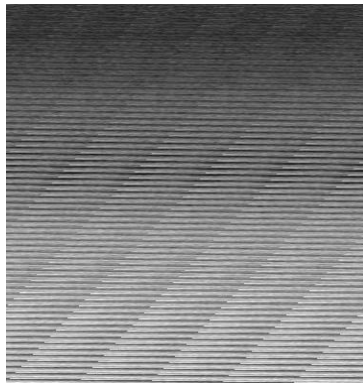


Figure 5. 16 Raw CFRP-A (left side) and Generated CFRP-A (right side) sequence thermal images from WGAN

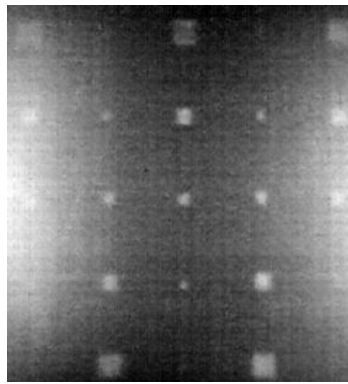
e. Visual explanations from GANs training

The deep models from Convolutional neural networks (including GANs) can be regarded as the "black box" which may involve a lack understanding of the internal functioning of the learning procedure. As a result, previous researchers introduced a proposed method referred to as Grad-CAM++ [47] to investigate the training process to provide an improvement of the visual explanations for the deep convolutional network. Specifically, the Grad-CAM++ adapted a weighted combination regarding specific scores from the positive partial derivatives in the feature map of the last convolutional layer of the GAN network to reconstruct a visual explanation for specific class labels.

(a) Pure noise



(b) Visible defect



(c) The Grad-CAM result from noise after GAN training; (d) The Grad-CAM result defect after GAN training

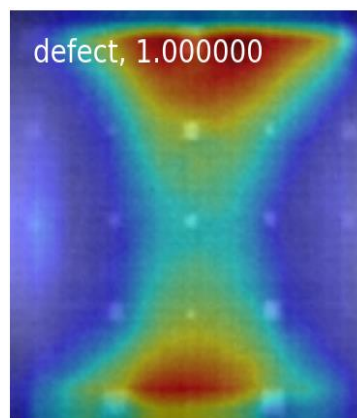
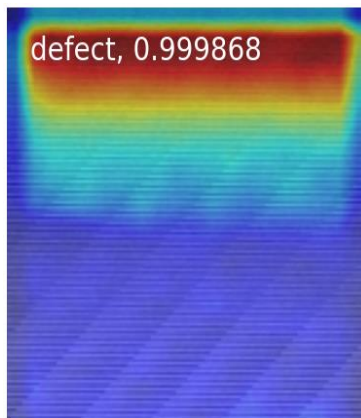
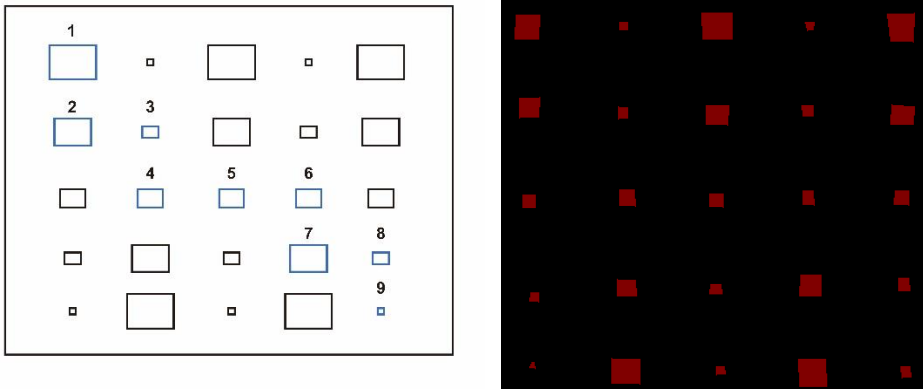


Figure 5. 17 Gradient-weighted Class Activation Mapping; without defect (a)(c); with defect (b)(d)

Generally, the main goal of using the Global Average Pooling [48] is to visualize what the GAN is “looking for” and how GAN shifts its attention over time. The Grad-CAM++ can be used as the generalized method to visualize the output feature map of CNN and evaluate the generator from the GAN to confirm that the Convolutional neural networks (CNN) whether it take the defect regions as the key learning feature to generate with or not.

As shown in Figure 5.17, two parts of the databases (one labeled as 1; another labeled as 0) have been prepared in order to train a binary classified GAN. With Grad CAM++, the heat map of the original image is visualized through the classification results to see which parts of the image are important for the features(defects). As indicated from the pure noise image in Figure 5.17(a), due to the region of the image activated by random noise (picture on the top left side), the CNN classifies the characteristics of the region, which has a greater gray value (assuming the possibility of defect = 0.99) as the actual defect in the Grad-CAM result (picture on the bottom left side). In Figure 5.17(b)(d) on the right side, based on the heat map generated, the most important identification of features that have been classified (red region in Figure 5.17(d)) does not include the boundaries of region at the right and left boundaries (containing a lot of noise). Therefore, it could further be confirmed that the GANs encoding process is to remove non-critical information (i.e. on both sides) and generate meaningful information to reconstruct the original maps.



(a) (b)  
 Figure 5. 18 Descriptions of the analysed defect regions for visibiliy in the CFRP specimen geography(a); The labelling coordinate values of each defect region matrix (b)

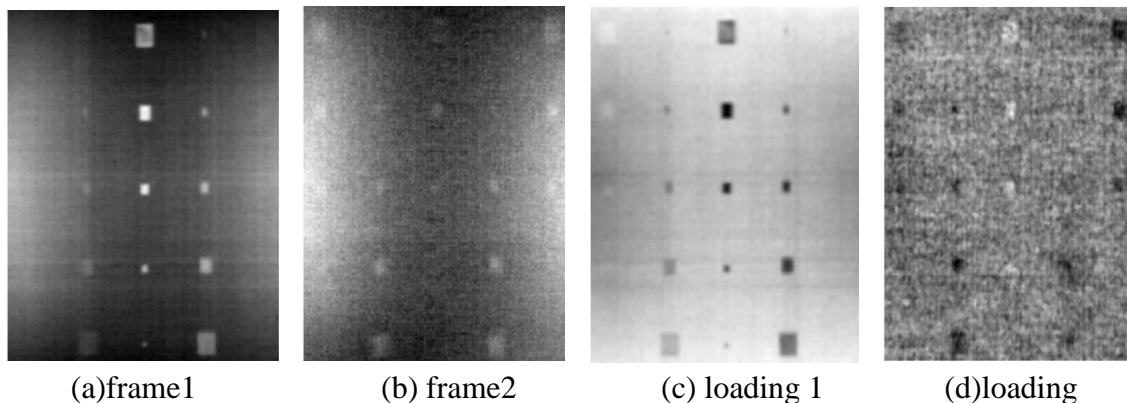
g. Defects analysis and enhancement in PLS-GANs

Once the merged thermal sequence (raw data; generative data) has been processed by PLS, a new set of thermal images which contain less noise and redundant information are generated. Specific defects ( nine representative locations) are selected on the specimens for further

analysis of defect visibility from SNR values. The detailed information of the location and sizes of defects has been illustrated in Figure 5.18(a). In Figure 5.18(b), a labeling coordinate graph for the GANs network training has been indicated for each defect region in the image matrix.

Nine defects of rectangular shapes, located at specific positions which ranged from large to small, have been selected. These defects represent different depths and distributions of defects from the three CFRP specimens in section 5.4.2 in order to evaluate the ability of the model to detect and provide quantitative analysis.

A total of 64 informative raw images from each thermal sequence in the specimen (a)-(c) with the region of interest ( $376 \times 376$ ) were selected during the cooling period from the pulsed thermography experiment (less than 2s). The PLS model directly processed the raw thermal database ( $3 \times 64 = 192$  thermal images) from thermal sequence (Specimen(a)-(c)). Simultaneously, the PLS-GANs model adapted the 64 informative raw images from each sequence as the real data generating a series of fake thermograms (64 generated thermal images) in each sequence from the specimen (a)-(c). As shown in (a)(b) from Figure 5.19- Figure 5.21, the representative original thermogram at different specific time frames has been revealed. The (c)(d) from Figure 5.19 - Figure 5.21 described the loading images from the regular PLS model. The (e)(f) and (g)(h) from Figure 5.19 - Figure 5.21 are the processed results from the DCGAN-PLS model and WGAN-PLS model respectively. Note that the components from Figure 5.19-Figure 5.21 are not based on the time order but selected by the feature indication (the most representative defects frames). As shown in Figure 5.19-Figure 5.21 (e)-(h), the PLS-GANs (PLS-WGAN; PLS-DCGAN) noise images contain less noise and enhanced defects.



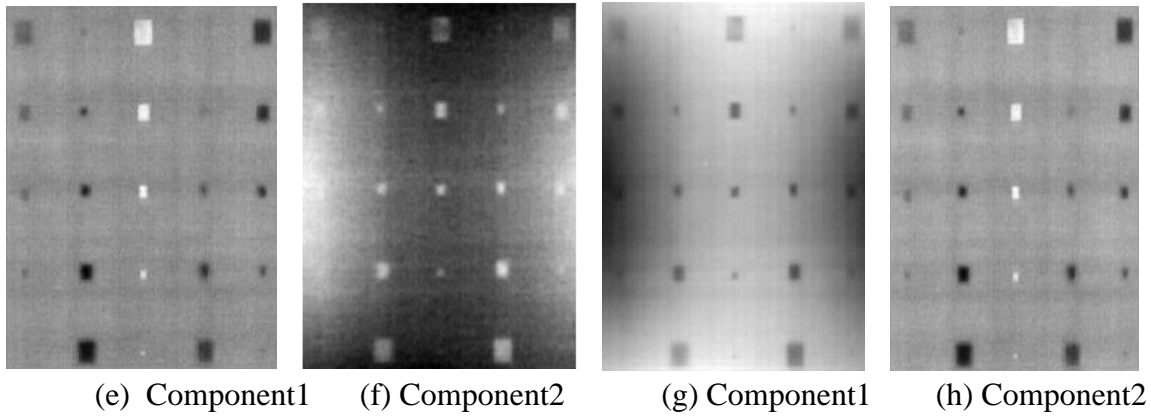


Figure 5. 19 Enhanced thermal images from Specimen 1 (376\*376): 64 raw representative thermal images (a)-(b); PLST(c)-(d); PLST-DCGAN(e)-(f); PLST-WGAN (g)-(h) (with limited 64 raw representative thermal images merging)

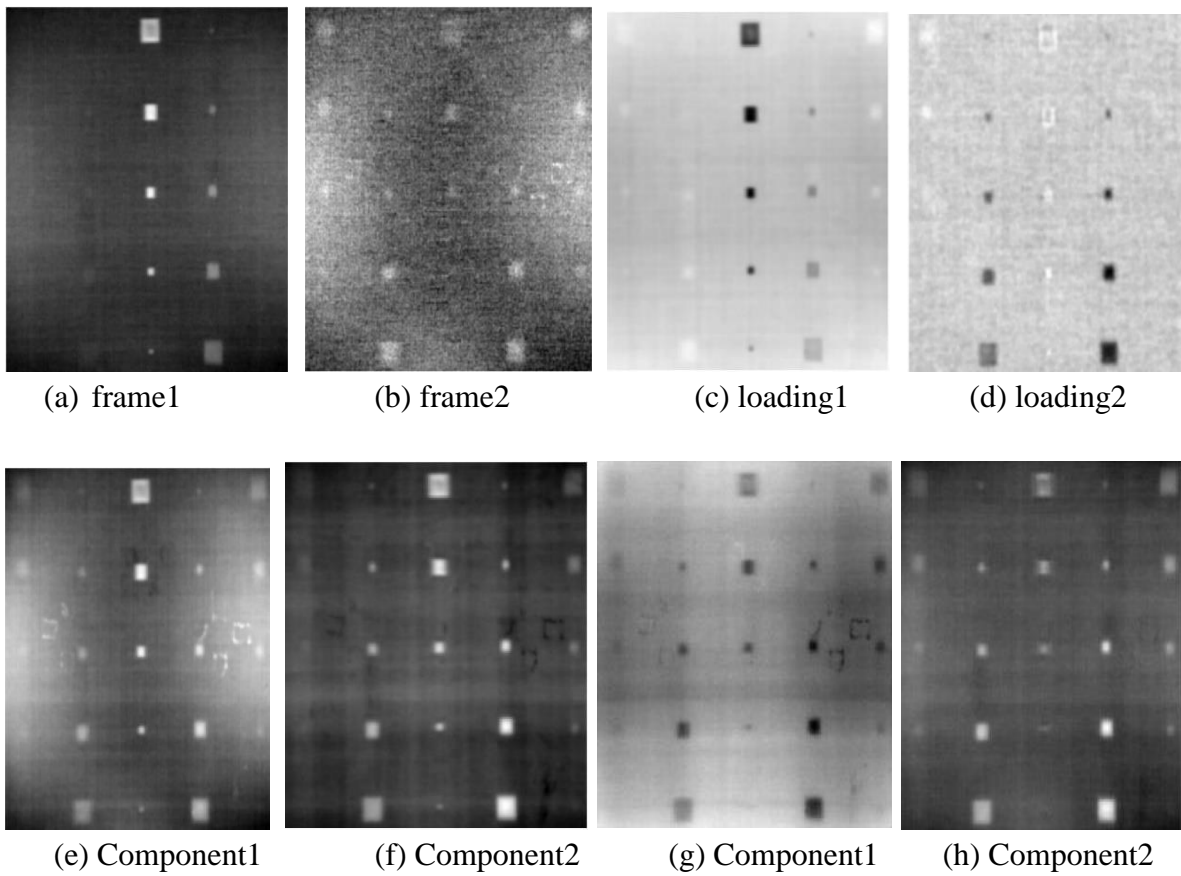


Figure 5. 20 Enhanced thermal images from Specimen 2 (376\*376): 64 representative

thermal raw images (a)-(b); PLST(c)-(d); PLST-DCGAN(e)-(f); PLST-WGAN (g)-(h) (with limited 64 representative raw thermal images merging)

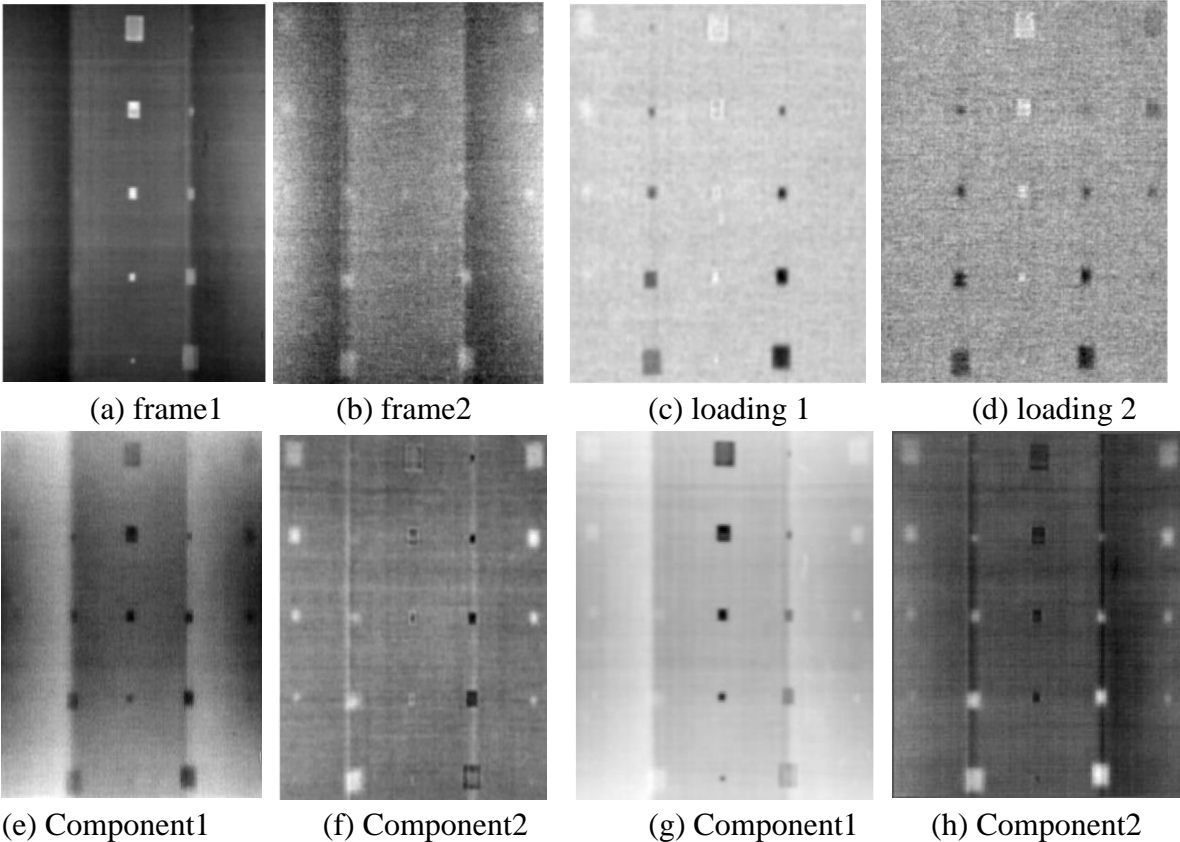


Figure 5. 21 Enhanced thermal images from Specimen 3 (376\*376): 64 representative thermal raw images (a)-(b); PLST(c)-(d); PLST-DCGAN(e)-(f); PLST-WGAN (g)-(h) (with limited 64 representative raw thermal images merging)

As shown in Figure 5.19-Figure 5.21, among the results from the PLS ((c)(d)) and PLS-GANs (DGAN:(e) (f); WGAN:(g) (h)), the PLS results (c)(d) obtained two notable loading components after the visualization. In comparison to the original raw thermogram, the related defect information has been highlighted. In addition, the contrast between defect-region and sound-region has become more significant. However, only the first two loading components of the experiment that extracted the most useful defect features, the rest of the components extracted less defect features and contain much noise making it more difficult to recognize defects.

It can be noticed in Figure 5.19(a),that the 1<sup>st</sup> component from the raw image indicated the most clearly detected defects at a depth level of 0.2mm. In contrast, the results from the PLS-GANs model are somehow different and indicated more defective features compared with

the regular PLS model, which clearly indicated the defects at a deeper depth level at 1mm. Therefore, the PLS-GANs model indicated the more high quality defects features basing on the extracted images.

For signal-to-noise analysis(SNR), the following principle could be applied to the grayscale thermal images: the pixel value of the brightest defective region is closer to 1. In contrast, the pixel value of the darkest non-defective part is closer to 0. Assuming that the noise from the thermal images is removed, then the prediction could be indicated as follows: the value of the numerator in the formula for SNR should become larger; the value of the denominator of SNR should become smaller; eventually, the signal-to-noise ratio will become larger. Therefore, in table 1, an indication of SNR values from a different number of generated images merging with the original thermograms has been illustrated. The SNR values increased gradually when the different number of generated thermograms was added on the original thermograms(64 frames). For example on Specimen(a), for the DCGAN model, as the generated thermal images (different number) were stacked on the raw thermograms, the SNR values obtained a significant increase from 0.78 (0 generative thermograms) to 1.23 (64 generative thermograms). However, after 40 generative thermograms were stacked on the original thermograms, the SNR keeps similar values regardless of how many images were adding on the databases after that.

In contrast, for the WGAN model, as the generated thermal images (different number) were stacked on the raw thermogram, the SNR values obtained a significant increase from 0.78 (0 generative thermal images) to 1.35 (64 generative thermal images). After 48 generative images were stacked on the original thermogram, the SNR keeps similar values regardless of how many images were added on the databases after that. Based on the experimental results, it further proven that the WGANs network obtained the highest SNR values when it is merged with the raw thermogram in this experiment. Similarly, SNR values in the generated image by DCGAN and WGAN from Specimen (b)-(c) have a similar scenarios as indicated in Table 2.



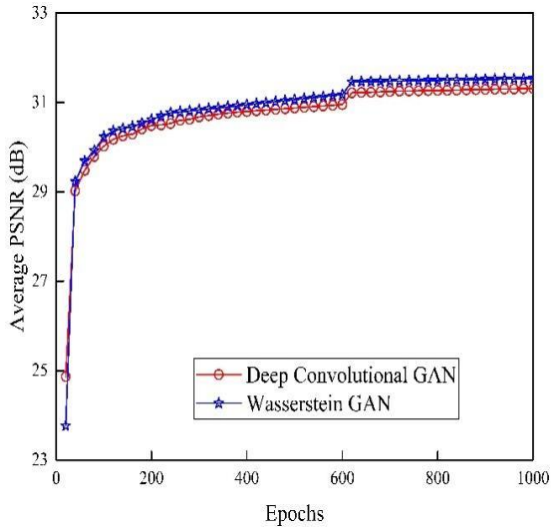
Table 5. 2 SNR values indication with different numbers of generated thermal images integrations

Raw thermograms that merging	Generate thermograms that from GANs	The Total thermograms	SNR of all defects								
			Specimen a		Specimen b		Specimen c				
64	The number of the thermograms		DCGA	WGAN	DCGA	WGAN	DCGA	WG			
			N		N		N	AN			
				0	64	0.78	0.78	0.81	0.81	0.77	0.77
				8	72	0.86	0.88	0.82	0.85	0.79	0.81
				16	80	0.95	0.99	0.94	0.96	0.92	0.94
				24	88	1.01	1.04	0.98	0.99	0.98	0.99
				32	96	1.19	1.22	1.06	1.09	1.04	1.08
				40	104	1.25	1.26	1.09	1.12	1.07	1.09
				48	112	1.23	1.35	1.07	1.08	1.10	1.09
				56	120	1.21	1.31	1.06	1.09	1.08	1.12
	64	128	1.23	1.34	1.09	1.11	1.10	1.12			

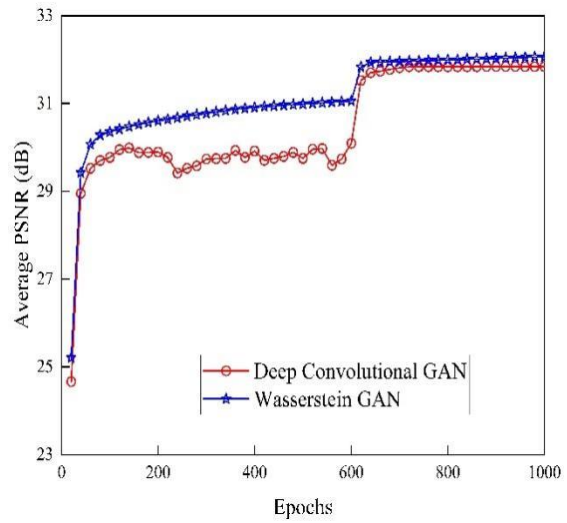
In order to specifically evaluate noise reduction performance from different feature extraction methods, the SNR values from representative defects (9 positions on the specimen) on three different specimens (a)-(c) from four methods have been listed in Table 2 to describe the SNR value comparison. As shown in table 2, it indicated that the SNR values of thermal images obtained from different dimension reduction methods: PLS+DCGAN (64 generated images + 64 raw images); PLS+WGAN (64 generated images+ 64 raw images); PLS (64 raw images); 64 Raw limited images. The SNR values obtained from the PLS+WGAN became the largest SNR values observed from nine specific representative locations of defects, compared with the other three methods (PLS+DCGAN; PLS; Raw). The average SNR values of defects from PLS+WGAN and PLS-DCGAN from three specimens are both larger than the regular PLS method and raw data. In total, the SNR values rate of the PLS+WGAN increased 20 % compared with the PLS extracted thermal images. As a result, the PLS+WGAN models are the effective method for the defect enhancement in the Non-destructive testing (NDT) for CFRP specimen' evaluation.

Table 5. 3 The comparison of SNR values from CFRP specimens with 9 selected representative defects

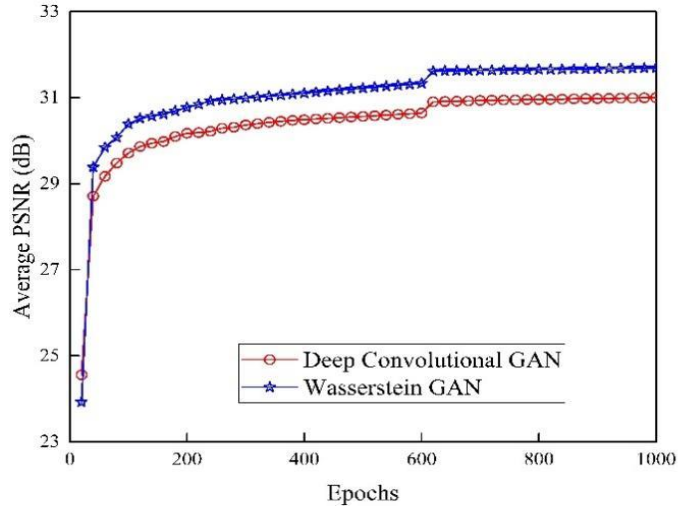
Feature extraction methods		Signal Noise Ratio (SNR)									
		Far left		Middle left		Middle	Middle right		Far right		All defects
		1	2	3	4	5	6	7	8	9	average
Raw	a	0.54	0.71	0.77	0.68	1.76	1.61	1.57	1.12	0.79	0.91
	b	0.55	0.78	0.64	0.69	1.67	1.47	1.38	1.12	0.65	0.90
	c	0.52	0.69	0.78	0.86	1.62	1.32	1.26	1.15	0.54	0.88
PLS	a	0.71	0.92	0.98	1.01	1.57	1.39	1.27	1.01	0.72	1.07
	b	0.67	0.82	0.88	0.97	1.68	1.41	1.36	1.02	0.73	1.03
	c	0.62	0.81	0.72	1.19	1.72	1.20	1.39	1.05	0.76	1.015
PLS+DC GAN	a	0.71	0.91	1.06	1.71	2.12	2.01	1.62	1.27	0.85	1.23
	b	0.73	0.87	1.08	1.74	2.01	1.92	1.51	1.16	0.72	1.19
	c	0.74	0.85	1.03	1.76	2.04	1.91	1.49	1.08	0.71	1.098
PLS+WG AN	a	0.81	0.95	1.08	1.87	2.15	2.06	1.67	1.29	0.93	1.32
	b	0.73	0.93	1.10	1.76	2.02	1.93	1.51	1.15	0.69	1.24
	c	0.84	0.92	1.08	1.88	2.07	2.24	1.52	1.09	0.79	1.23



(a) Average PSNR values in Specimen (a)



(b) Average PSNR values in Specimen (b)



(c) Average PSNR values in Specimen (c)

Figure 5. 22 Gaussian denoising results from two specific GAN models (WGAN; DCGAN) from Specimen (a)-(c)

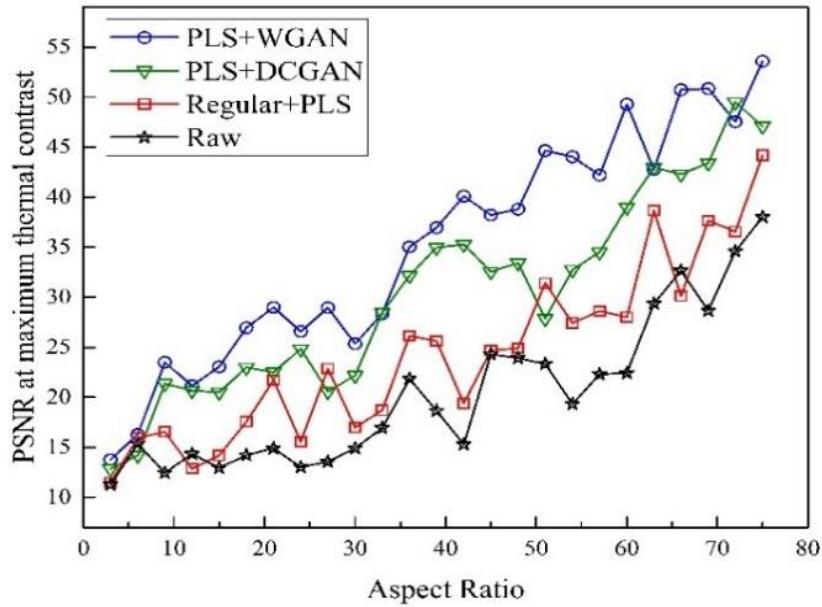
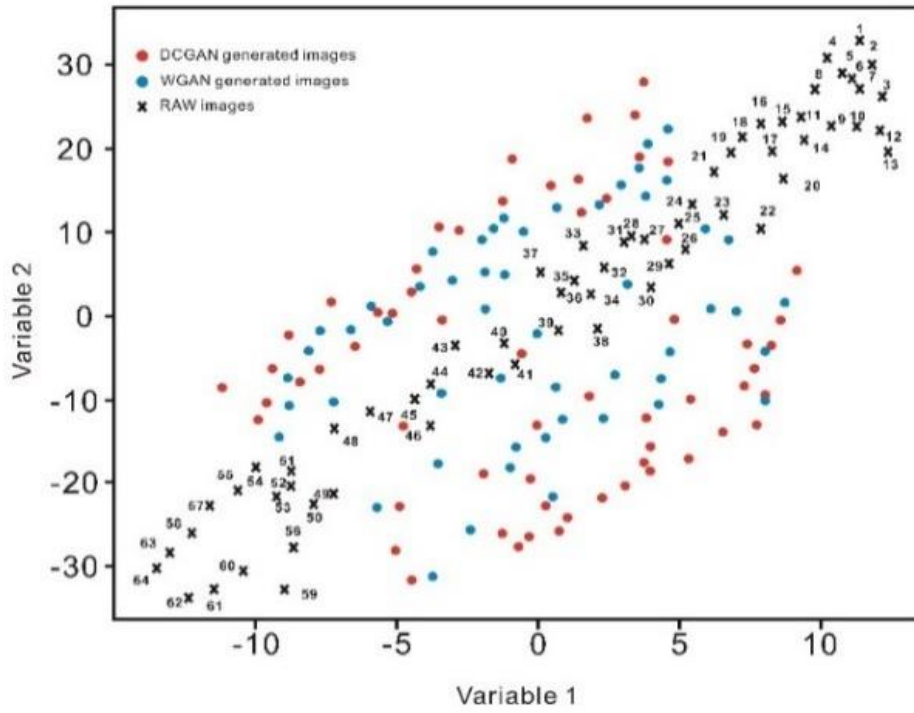
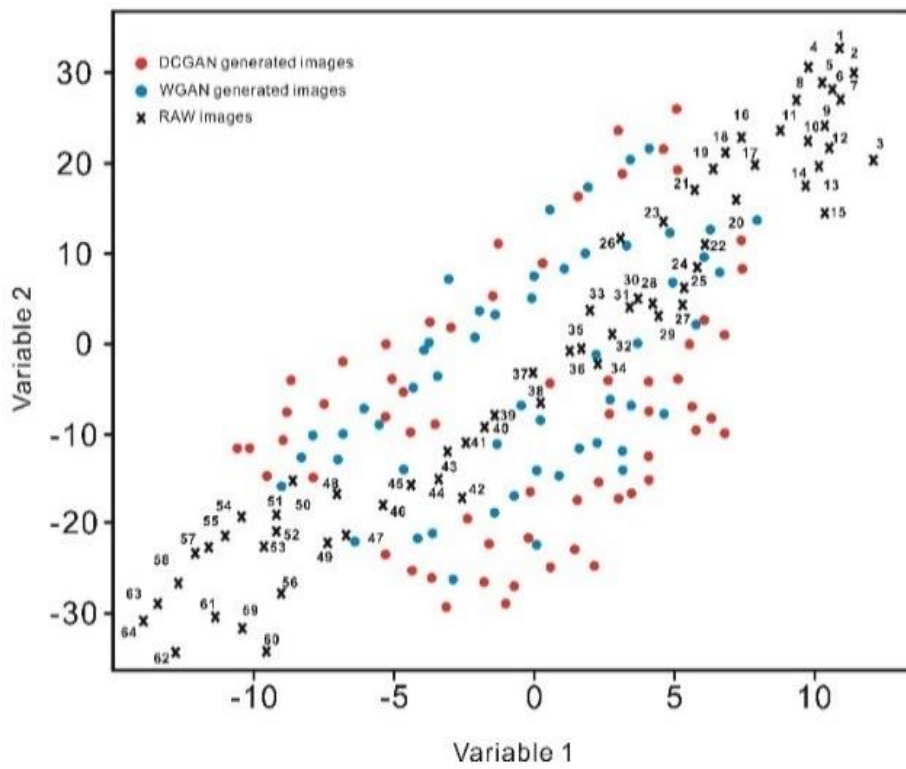


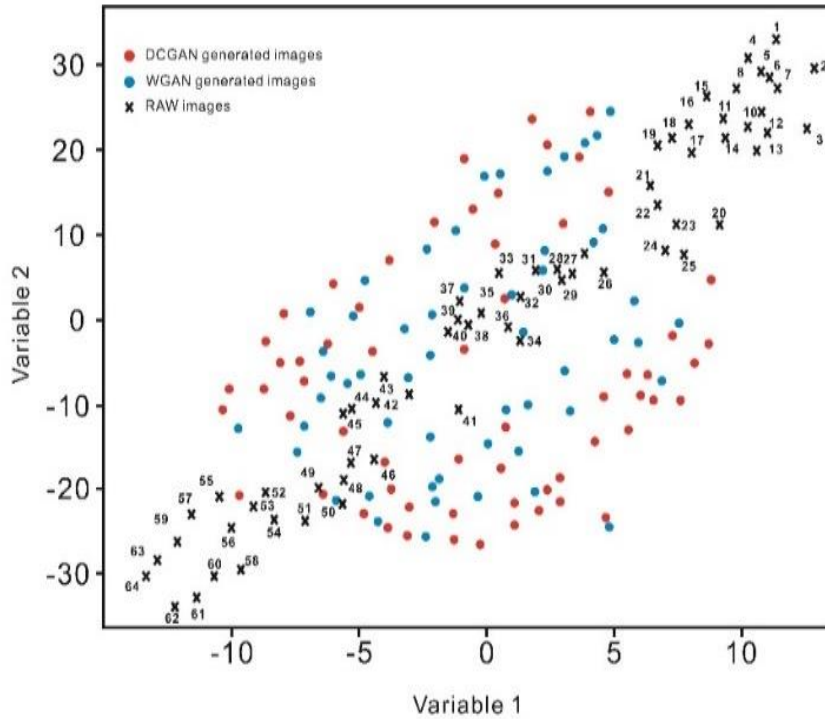
Figure 5. 23 Comparison of signal-to-noise ratio at maximum signal contrast (PSNR) for raw data; PLS; PLS-WGAN; PLS-DCGAN(three specimens(a)(b)(c) in total caculation) with 64 raw thermal images being processed



(a) Perplexity=50, iteration=500 in Specimen(a);



(b) Perplexity=50, iteration=500 in Specimen(b);



(c) Perplexity=50, iteration=500 in Specimen(c)

Figure 5. 24 t-SNE visualization of raw images and DCGAN/WGAN-generated images from Specimens (a)(b)(c)

Simultaneously, to evaluate the pixel values based on the coordinate of defect locations from the raw image and the generated images by GANs network (DCGAN/WGAN), the average resolution values for the whole thermogram (from three specimens) have been obtained in the following: for the raw image, 1. the mean pixel value of the defective part of the generative map is calculated as 0.778; 2 the mean pixel value of the non-defective part is 0.466; 3.the variance is 0.153; 4.the standard deviation is 0.391. As a result, the average signal-to-noise ratio for the entire raw image dataset is 0.796.

For the WGAN generated image, 1. The mean pixel value of the defective part of the generative map is calculated as 1.012; 2. The mean pixel value of the non-defective part is 0.382; 3. The variance is 0.299; 4. The standard deviation is 0.547. In contrast, for the DCGAN generated image, 1. The mean pixel value of the defective part of the generative map is calculated as 1.008; 2. The mean pixel value of the non-defective part is 0.399; 3. The variance is 0.364; 4. The standard deviation is 0.603. The average signal-to-noise ratio of the generated images in WGAN and DCGAN are 1.14 and 1.01 respectively. Therefore, the

WGAN network obtained the highest signal-to-noise value further testifying to its efficient noise reduction capability.

Figure 5.22 (a)-(c), it shows the performance of the noise removal from the generated thermal images in the WGAN and DCGAN based on the specimen(a)-(c) during the training. Two generated thermal image databases(64\*3=192 images) generated from the WGAN and DCGAN model for the specimen (a)-(c) being evaluated during the epochs increasing of training respectively. The average PSNR values were obtained during the entire training process. Although each graph has a certain degree of variation in the PSNR values curve from WGAN and DCGAN, it can be seen in Figure 5.22 (a)-(c) that the performance of WGAN is better than DCGAN for image denoising. The average PSNR values from DCGAN are slightly lower than WGAN's value on the average regardless of the graph in Figure 5.22 for the three thermogram databases of the specimen (a)-(c) during the 1000 epochs training. This further proves that WGAN has a more stable capability for thermal data generation and noise reduction in this experiment in comparison to DCGAN.

In Figure 5.23, it is shown that the PSNR values improved after the processing using the models of PLS-WGAN, PLS- DCGAN, PLST in comparison with original raw thermal sequences (64 thermal image processing). The PSNR was classified as a function of an aspect ratio  $r = \text{diameter (D)}/\text{depth (d)}$  in the Figure. The detection capability relies on the size and depth of defects. The PSNR values based on the average pixels of defect regions and sound regions which are calculated at the maximum thermal contrast signal.

As shown in Figure 5.23, four polylines were computed respectively for different types of data: the unprocessed raw data(red-polyline); the preprocessed data from PLST(green-polyline); the preprocessed data from the PLS-WGAN (blue-polyline); the preprocessed data from the PLS-DCGAN(yellow-polyline). As shown, the blue polyline from PLS-WGAN obtained the highest values in the PSNR compared with the three other polylines (PLS-DCGAN; PLST; Raw) during the entire aspect ratio (diameter D/depth d).

More specifically, with the significant improvement based on the PSNR values from raw data to three feature extraction methods: regular PLST method, PLS-DCGAN, and PLS-WGAN, the average Peak SNR values increase based on each defect specifically from 53% (40/75) in the raw images, to 77% (58/75) in PLST, 85% (63/75) in PLS-DCGAN, 90% (68/75) in PLS-WGAN of all the defects (75 defects) in three different specimens (a)-(c) in total. The PLS-WGAN obtained the highest SNR values, in comparison with the PLS-DCGAN, the PLST, the raw data based on Figure 5.23. The total performance of SNR values has been improved from WGAN/DCGAN due to its noise reduction of the functional feature extraction structure (autoencoder structural format).

In Figure 5.24 (a)-(c), the t-distributed stochastic neighbor embedding (t-SNE) [49] plot in three specimens (a)-(c) has been indicated with 2 dimensional embedding feature points from 3 types of thermograms: the raw thermogram (black forks), the WGAN thermogram (red points), the DCGAN thermogram (blue points). The frame order for the time sequence of the original image is being marked based on the feature points in the Figures. In this case, the t-SNE finds a way to project thermal data into low dimensional space (in this case, the 2d plane dimension) so that the clustering in the high dimensional space (original thermal data) is preserved. Besides, the dissimilar feature points will be modeled as far apart as possible in the t-SNE visualized figure.

Due to the noise reduction from the GANs network (autoencoder format), as indicated in Figure 5.24, the generated image (WGAN; DCGAN) thermograms (64 images from each model) keep a similar pattern and features with the middle parts of raw images but have a wider scattering. The most informative and featuring part of the original sequence is from the middle part of the sequence (17<sup>th</sup> – 48<sup>th</sup>), simultaneously the GANs (WGAN; DCGAN) network tends to focus on analysing the feature from this part and highlight the useful information from it (17<sup>th</sup> – 48<sup>th</sup>). Based on the results, the points from the generated images in the DCGAN have the widest distribution in three specimen cases (Specimen (a) - Specimen (c)) rather than the generated images from WGAN, but feature points from WGAN keep the close neighborhood connections with the informative feature points from the 17<sup>th</sup> – 48<sup>th</sup> frames in the original thermal sequence to mimic its patterns .

## 5.6 Results discussion and experimental analysis

With the results obtained as discussed above, the following points could be concluded:

1. For the DCGAN model, the principle is that the generator produces a defective graph of noise vector  $z$  to fool the discriminator. Once it has failed, the discriminator is punished, then the training continues until two networks achieve equilibrium. That is once, the graph generated after training is very close to the original graph with the raw data.. However, if the original graph contains a lot of noise, the generator learned it also so that it produces a graph containing noise along with defect enhancement.. As a result, the discriminator should be given the thermal images contain heavily feature of defects with less noise (Figure 5.15; Figure 5.16) so that what the generator learned by training during this procedure generates less noisy thermal images. The model does not work well when noisy information from the input images are being processed. This affects the model's performance. Based on the results, when the input images are noisy and each image features large, the model may will not be able to learn a uniform data distribution to generate enhanced image as shown in Section 5.5.4 (part c).

2. The noise from thermogram influences the generation of the GANs network. When original thermogram contains noise and present uneven heating background (in Figure 5.24), this influences the GANs and prevents it from learning features from the raw data to generate suitable thermal images. Particularly, the initial frames of original thermogram always contain a significant amount of noise and less meaning features related to defect indication. Similarly, thermogram at the end of each sequence has the exhibit much noise as well as shown on the t-SNE visualized in Figure 5.23 (a)-(c).

3. GAN network training complexity issues: GANs training difficulty problem. Although GAN itself has been shown to be effective, it has been found that changing the batch normalization for some layers or changing the activation function during the training experiment can result in noisy images. When reproducing a GAN network, it is important to consider not only the network structure is correct, but also whether the loss function optimization process is correct (such as: WGAN). Therefore, it is better to build a simple network and use small size images at the beginning to test the correctness of the loss function optimization process, and then change the network structure to implement the function of generating large size images like the learning section we described in Section 5.5.4 (part a and part c).

4. The non-uniform heating influence the results from raw images and the PLS model. Therefore, the amount of noise still appears on the processed images from the PLS model. After the PLS-GANs model generates the noise-reduction images, the quality of the processed images improves and contains less noise. The generation of PLS-GANs is similar to the process from non-uniform heating noise removal from the regular method as shown in the (e)-(h) from Figure 5.19-Figure 5.21 due to its noise reduction performance of the autoencoder structure.

5. Data augmentation for feature enhancement: the most functional point that GANs achieve is based on data augmentation to generate useful defect features for dimensional reduction methods such as PLST (partial least-square thermography), the PCT (principal component thermography), TSR (thermographic signal reconstruction) further enhance and extract defect features. Therefore, the generative data from GANs network could be a useful alternative for feature enhancement and extraction in the scenario where data is noisy and limited (Table 5.2).

6. The synthetic data (sequences) generated from GANs could be beneficial for feature extraction and defect enhancement. As shown in Figure 5.17-5.19 (a)-(d), the raw thermal sequence (which includes 64 frames) clearly indicates the defects. After merging with an amount of synthetic images generated from GANs (DCGAN; WGAN) network, the detection results have been significantly improved based on the original PLS model. Therefore, by



using synthetic data from GANs network, it could deal with the situation when only limited data is obtained from the experimental thermography acquisition procedure in order to increase the features lacking due to the data limitation.

7. The increasing expansion of the limited thermal dataset could be the subject of further discussion and research. Note that, only limited representative 64 raw images are taken due to time and training complexity from the multiple GANs network in this experiment. After the expansion of informatic thermal images being fed into the network, the results might lead to further significant improvement.

## 5.7 Conclusions

This study introduced the deep learning algorithm into the field of thermographic-data analysis and developed the GANs-based data-augmentation thermography model named PLS-GANs (PLS-WGAN; PLS-DCGAN). Compared with the classical feature extraction model (PLS), the proposed deep-learning thermography model utilizes the data-augmentation strategy to improve the accuracy and reliability of the defect detection results in composite materials. Generally, the GAN-based data augmentation thermography modeling process includes data generation, data merging, data preprocessing, data analysis, data visualization, and result evaluation. Using the unified framework, other thermography methods available in the literature will also be combined with the GANs-based data augmentation approaches to further enhance their defect detection performance and reduce image noise. Other composite materials available in the literature will be also combined with the GANs-based data augmentation approaches to further enhance their results with defect detection performance. The modified structured-convolutional neural networks will be utilized and be proposed to enhance the defect visibility as well.

## References

- [1] C. Ibarra-Castanedo, J. R. Tarpani, and X. Maldague, "Nondestructive testing with thermography," *Eur. J. Phys.*, vol. 34, pp. S91–S109, Oct. 2013.
- [2] B. Gao, L. Bai, W. L. Woo, G. Y. Tian, and Y. Cheng, "Automatic defect identification of eddy current pulsed thermography using single channel blind source separation," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 913-922, Apr. 2013.
- [3] Shepard S M. *Advances in pulsed thermography[C]//Thermosense XXIII. International Society for Optics and Photonics, 2001, 4360: 511-515.*

- [4] Nixon M, Aguado A. *Feature extraction and image processing for computer vision*[M]. Academic press, 2019.
- [5] N. Rajic, "Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures," *Compos. Struct.*, vol. 58(4), pp. 521–528, Dec. 2002.
- [6] Maldague X, Galmiche F, Ziadi A. *Advances in pulsed phase thermography*[J]. *Infrared physics & technology*, 2002, 43(3-5): 175-181.
- [7] Balageas D L, Roche J M, Leroy F H, et al. *The thermographic signal reconstruction method: a powerful tool for the enhancement of transient thermographic images*[J]. *Biocybernetics and biomedical engineering*, 2015, 35(1): 1-9.
- [8] Deng L, Yu D. *Deep learning: methods and applications*[J]. *Foundations and trends in signal processing*, 2014, 7(3–4): 197-387.
- [9] Fang Q, Nguyen B D, Castanedo C I, et al. *Automatic defect detection in infrared thermography by deep learning algorithm*[C]//*Thermosense: Thermal Infrared Applications XLII. International Society for Optics and Photonics*, 2020, 11409: 114090T.
- [10] Fang Q, Ibarra-Castanedo C, Maldague X. *Automatic defects segmentation and identification by deep learning algorithm with pulsed thermography: Synthetic and experimental data*[J]. *Big Data and Cognitive Computing*, 2021, 5(1): 9.
- [11] Hu J, Xu W, Gao B, et al. *Pattern deep region learning for crack detection in thermography diagnosis system*[J]. *Metals*, 2018, 8(8): 612.
- [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., "Generative adversarial nets," *In Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [13] Wang J, Yu L, Zhang W, et al. *Irgan: A minimax game for unifying generative and discriminative information retrieval models*[C]//*Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. 2017: 515-524.

- [14] Liu K, Li Y, Yang J, et al. *Generative principal component thermography for enhanced defect detection and analysis*[J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(10): 8261-8269.
- [15] Lopez F, Ibarra-Castanedo C, de Paulo Nicolau V, et al. *Optimization of pulsed thermography inspection by partial least-squares regression*[J]. *Ndt & E International*, 2014, 66: 128-138.
- [16] Liu B, Zhang H, Fernandes H, et al. *Experimental evaluation of pulsed thermography, lock-in thermography and vibrothermography on foreign object defect (FOD) in CFRP*[J]. *Sensors*, 2016, 16(5): 743.
- [17] Adler J, Lunz S. *Banach wasserstein gan*[J]. *arXiv preprint arXiv:1806.06621*, 2018.
- [18] Radford A, Metz L, Chintala S. *Unsupervised representation learning with deep convolutional generative adversarial networks*[J]. *arXiv preprint arXiv:1511.06434*, 2015.
- [19] Tavolga W N. *Signal/noise ratio and the critical band in fishes*[J]. *The Journal of the Acoustical Society of America*, 1974, 55(6): 1323-1333.
- [20] Nielsen C, Okoniewski M M. *GAN Data Augmentation Through Active Learning Inspired Sample Acquisition*[C]//*CVPR Workshops*. 2019: 109-112.
- [21] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.
- [22] Ibarra-Castanedo C, Maldague X P V. *Infrared thermography*[M]//*Handbook of technical diagnostics*. Springer, Berlin, Heidelberg, 2013: 175-220.
- [23] C. Ibarra-Castanedo, S. Sfarra, M. Klein, and X. Maldague, "Solar loading thermography: Time-lapsed thermographic survey and advanced thermographic signal processing for the inspection of civil engineering and cultural heritage structures," *Infrared Phys. Technol.*, vol. 82 pp. 56–74, May. 2017.
- [24] Vinzi V E, Chin W W, Henseler J, et al. *Handbook of partial least squares*[M]. Berlin: Springer, 2010.

- [25] Larsen D R, Speckman P L. *Multivariate regression trees for analysis of abundance data*[J]. *Biometrics*, 2004, 60(2): 543-549.
- [26] S. Wold, H. Martens and H. Wold, "The Multivariate Calibration Problem in Chemistry Solved by the PLS Method," in *Conference Matrix Pencils, Heidelberg, Germany, 1984*.
- [27] H. Martens and T. Naes, *Multivariate Calibration*, Chichester, UK: John Wiley & Sons, 1989.
- [28] Alma Ö G. *Comparison of robust regression methods in linear regression*[J]. *Int. J. Contemp. Math. Sciences*, 2011, 6(9): 409-421.
- [29] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*. Nov. 2015.
- [30] Q. Xuan, Z. Chen, Y. Liu, H. Huang, G. Bao, D. Zhang, "Multiview generative adversarial network and its application in pearl classification," *IEEE Trans. Ind. Electron.*, vol. 66(10), pp. 8244–8252, Oct. 2019.
- [31] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," *arXiv preprint arXiv:1711.04340*. 2017.
- [32] Su J. *GAN-QP: A novel GAN framework without gradient vanishing and lipschitz constraint*[J]. *arXiv preprint arXiv:1811.07296*, 2018.
- [33] Gao F, Yang Y, Wang J, et al. *A deep convolutional generative adversarial network (DCGANs)-based semi-supervised method for object recognition in synthetic aperture radar (SAR) images*[J]. *Remote Sensing*, 2018, 10(6): 846.
- [34] Wang Q, Zhou X, Wang C, et al. *WGAN-based synthetic minority over-sampling technique: Improving semantic fine-grained classification for lung nodules in CT images*[J]. *IEEE Access*, 2019, 7: 18450-18463.
- [35] Andoni A, Indyk P, Krauthgamer R. *Earth mover distance over high-dimensional spaces*[C]//SODA. 2008, 8: 343-352.
- [36] Rippl T, Munk A, Sturm A. *Limit laws of the empirical Wasserstein distance: Gaussian distributions*[J]. *Journal of Multivariate Analysis*, 2016, 151: 90-109.

- [37] Ioffe A D. Regular points of Lipschitz functions[J]. *Transactions of the American Mathematical Society*, 1979, 251: 61-69
- [38] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *Proc. Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 214–223.
- [39] Han S, Sung Y. Dimension-wise importance sampling weight clipping for sample-efficient reinforcement learning[C]//*International Conference on Machine Learning*. PMLR, 2019: 2586-2595.
- [40] Lopez F, Nicolau V, Maldague X, et al. Multivariate infrared signal processing by partial least-squares thermography[C]//*ISEM Conference*. 2013.
- [41] Wold H. Nonlinear iterative partial least squares (NIPALS) modelling: some current developments[M]//*Multivariate analysis—III*. Academic Press, 1973: 383-407.
- [42] Chai T, Draxler R R. Root mean square error (RMSE) or mean absolute error (MAE)—Arguments against avoiding RMSE in the literature[J]. *Geoscientific model development*, 2014, 7(3): 1247-1250.
- [43] Meng Q, Catchpole D, Skillicom D, et al. Relational autoencoder for feature extraction[C]//*2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017: 364-371.
- [44] Bojanowski P, Joulin A, Lopez-Paz D, et al. Optimizing the latent space of generative networks[J]. *arXiv preprint arXiv:1707.05776*, 2017.
- [45] C. Ibarra-Castanedo, J.-M. Piau, S. Guilbert, N. P. Avdelidis, M. Genest, A. Bendada, et al., “Comparative study of active thermography techniques for the nondestructive evaluation of honeycomb structures,” *Res. Nondestruct. Eval.*, vol. 20(1), pp. 1–31, Jan. 2009.
- [46] Matworks E. Compute peak signal-to-noise ratio (PSNR) between images[J].
- [47] Chattopadhyay A, Sarkar A, Howlader P, et al. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks[C]//*2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018: 839-847.

[48] Hsiao T Y, Chang Y C, Chou H H, et al. Filter-based deep-compression with global average pooling for convolutional networks[J]. *Journal of Systems Architecture*, 2019, 95: 9-18.

[49] L. van der Maaten, and G. E. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

## **Part IV. Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning**

This following one chapter will present one article published on Applied science Journal 2020 – a method of defect depth estimation for simulated infrared thermography data with deep learning which concerning the exploration of deep learning approach in infrared thermography specifically for Quantitative analysis involved in Non-Destructive Testing & Evaluation.

### **Generation explanation:**

Quantitative analysis is playing an important role in the modern industrial field of non-destructive testing (NDT). Defect characterization is one of the current topics of interest in quantitative data analysis of active thermography.

The traditional approach in defect characterization is usually taking regular thermal contrast-based techniques used for the quantitative analysis which is based on obtaining thermal contrast values at a specific time involving mathematical calculation and definitions.

In this study, we mainly focused on a defect characterization issue by using simulated IR data combining with the deep learning algorithm to achieve automatic defect characterization. The recurrent neural network (Gated Recurrent Units) trained the simulated infrared data can automatic outputted the specific depth quantification for composite material samples via a supervised regression learning task based on the training and evaluation.

## Chapter 6 Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning methods

*The results of this study were firstly presented at 2020 Structural Health Monitoring & Nondestructive Testing conference, then it was published on-line in the Applied Sciences Journal, in September 2020. Cited 7 times up to now.*

### 6.1 Résumé

La thermographie infrarouge s'est déjà révélée être une méthode importante dans l'évaluation non destructive, car elle fournit des informations de manière immédiate, rapide et à faible coût. Cependant, le problème le plus épineux pour une application plus large de l'IRT est la quantification. Dans ce travail, nous avons proposé une technique spécifique de quantification de la profondeur en utilisant les unités récurrentes à grille (GRU) dans des échantillons de matériaux composites par thermographie pulsée (PT). La modélisation par la méthode des éléments finis (FEM) permet l'examen économique de la réponse de la thermographie pulsée. Dans ce travail, des échantillons de polymères renforcés de fibres de carbone (CFRP) incorporés avec des trous à fond plat sont stimulés par une modélisation FEM (COMSOL) avec une profondeur et une géométrie des défauts contrôlées avec précision. Le modèle GRU a quantifié automatiquement la profondeur des défauts présentés dans le matériau CFRP stimulé. La méthode proposée a évalué la précision et la performance des données synthétiques de PRFC provenant de la FEM pour les prédictions de la profondeur des défauts.

#### Contributing authors:

**Qiang Fang (Ph.D. Candidate):** a part of the experiment planning, data collection, data analysis, designing and implementing the algorithm. Moreover, testing their accuracy and robustness throughout the process and writing the manuscript.

**Xavier Maldague (The research director):** supervision, revision and correction of the manuscript.

#### Other contributors:

**Farima Abdolahi. Mamoudan:** COMSOL thermogram simulation

**Annette Schwerdtfeger** (research officer): manuscript preparation.



# A Method of Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning

Qiang Fang 1\*, and Xavier. Maldague 1\*

1 Department of Electrical and Computer Engineering, Computer Vision and Systems Laboratory, Université Laval, 1065, av. de la Médecine, QC G1V 0A6, Canada;

\* Correspondence: qiang.fang.1@ulaval.ca (Q.F); Xavier.Maldague@gel.ulaval.ca (X.M);

Received: 25 August 2020; Accepted: 21 September 2020; Published: 25 September 2020

## 6.2 Abstract

Infrared thermography has already been proven to be a significant method in non-destructive evaluation since it gives information with immediacy, rapidity, and low cost. However, the thorniest issue for the wider application of IRT is quantification. In this work, we proposed a specific depth quantifying technique by employing the Gated Recurrent Units (GRUs) in composite material samples via pulsed thermography (PT). Finite Element Method (FEM) modeling provides the economic examination of the response pulsed thermography. In this work, Carbon Fiber Reinforced Polymer (CFRP) specimens embedded with flat bottom holes are stimulated by a FEM modeling (COMSOL) with precisely controlled depth and geometrics of the defects. The GRU model automatically quantified the depth of defects presented in the stimulated CFRP material. The proposed method evaluated the accuracy and performance of synthetic CFRP data from FEM for defect depth predictions.

**Keywords:** NDT Methods; Defects depth estimation; Pulsed thermography; Gated Recurrent Units

## 6.3 Introduction

Non-destructive evaluation (NDE) has emerged as an important method for the evaluation of the properties of components or systems without damaging their structure. Several state of the arts methodologies such as Pulsed Phase Thermography (PPT) [1], Principal Component Thermography (PCT) [2], Differential of Absolute Contrast (DAC) [3], Thermographic Signal Reconstruction (TSR) [4], as well as Candid Covariance Free Incremental Principal Component Thermography [5], have been implemented to process thermographic sequences and improve the defect visibility. These techniques can be beneficial for qualitative analysis of composite materials. However, to introduce a method for conducting quantitative study (defect depth estimation) with deep learning is a novel topic to be explored.

Quantitative analysis is playing an important role in the modern industrial field of non-destructive testing (NDT). Defect characterization is one of the current topics of interest in quantitative data analysis of active thermography. In this topic, thermographic information involves extracting quantitative subsurface properties from defects such as defect depth, lateral size, thermal resistivity, from an experimental Thermal Non-destructive Testing

(TNDT) dataset utilized to characterize defects. Several approaches have already been proposed to analyze the depth in the region of defects in pulsed thermography. In general, these approaches evaluate the defects depth by using the maximum thermal contrast  $C_{max}$ , the instant of maximum thermal contrast  $Tc_{max}$  or artificial neural networks to analyze the depth of defects based on mathematical equations. The Peak Temperature Contrast Method [6] estimated the depth based on the characteristic time from peak contrast. The peak contrast corresponding to the maximum contrast has a proportional correlation with the square of the defect depth. Daribi et al. [7] proposed neural networks for defect characterization of defect depth. The results demonstrated that the networks should be trained by representative and non-redundant data in order to obtain a high degree of classification accuracy.

In this work, there is an attempt to detect the depth of defects in Carbon Fiber Reinforced Polymer (CFRP) via Gated Recurrent Units (GRUs) [8]. GRU is an updated recurrent neural network (RNN) particularly designed for time series prediction. GRU can be considered as a variation of Long Short-Term Memory (LSTM) [9]. Compared with the LSTM and RNN temporal model, the GRU has adapted a few learning parameters which could save computational expenses for training and obtained an excellent performance. According to our knowledge, this is the first time that the thermal temporal characteristic model (GRU; RNN; LSTM etc.) is used to qualify the depth of defects.

We modeled a 3D version of CFRP specimen stimulation from COMSOL. Then, it was further tested on the systemic data by the GRU model to validate its accuracy. The remainder of this paper is structured as follows: Section 2 provides the pulsed thermography theory and conception indication, as well as the detailed characteristics of FEM simulations. Section 3 proposed a GRU model-based defects depth estimation strategy and introduces the GRU deep learning model architectures. Section 4 provides the experimental results analysis. Section 5 concludes this paper.

## **6.4 Thermal Consideration and FEM Stimulation**

### **6.4.1. Pulsed Thermography**

In pulsed thermography (PT), a high-power exponential heating impulse is applied to the samples, and a thermal response is measured during a period of time. Due to the heat conduction, a surface region which has an internal defect underneath the surface perturbs the thermal waves propagation on the surface of specimens in comparison to the sound (non-defective) region. We can then see the changes of the temperature variation, since the internal defects possess different thermo-physical (conductivity, density and heat capacity) that have an impact on heat flow. These thermal differences can be observed as surface features and recorded with an infrared camera as indicated in Figure 6.1 [10].

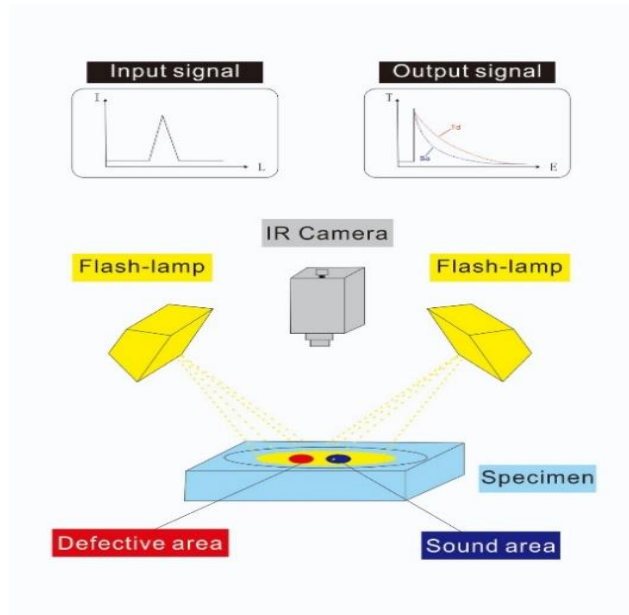


Figure 6. 1 Pulsed thermographic testing using optical excitation.

Temporal evolutions can be observed from the defective regions and subsurface sound regions. A thermal contrast is acquired as a feature vector which is obtained distinctly via the thermal value from the defective region subtracted from the corresponding value from the surrounding sound region [11] as indicated in Equation (6.1), where  $T_d(t)$  is the temperature value on the pixel point of the defect area. The temperature value on the reference point of the sound area is  $T_s(t)$  Then the  $\Delta T(t)$  is the absolute thermal contrast extracted from the defect and sound region. The thermal contrast is an excellent technique to distinguish the temperature difference to learn the depth of the defects, as shown in Figure 6.1.

$$\Delta T(t) = T_d(t) - T_s(t) \quad (6.1)$$

#### 6.4.2. Finite Element Modeling with Transient Heat Transfer

Finite Element Modelling (FEM: COMSOL, etc.) [12] has become an important and economical platform to evaluate the thermal response of pulsed thermography, which builds up models for the platform of components that allow us to flexibly examine all specific physical aspects of thermal data such as geometries and properties of materials. Simulated thermograms matched well with experimental data.

In this work, COMSOL will be utilized as a 3D based simulation to build up models for the synthetic data to provide for depth estimation of artificial defects. A Carbon Fiber Reinforced Polymer (CFRP) structure-based specimen containing artificial defects of different shapes (flat-bottom holes) is modeled as shown in Figure 6.2. A high thermal pulse is projected on

the surface of specimens. Due to the existence of the temperature gradients in the sample, a thermal front propagates from the high temperature region on the surface to the region underneath. A delamination or discontinuities create a lower thermal diffusion rate to the heat flow and then reflect the unnormal thermal patterns on the surface. The COMSOL software is utilized as the heat transfer simulation model for obtaining the temperature evaluations of the surface of each sample, as indicated below [13]:

$$\rho C_P \left( \frac{\partial T}{\partial t} \right) - \nabla \cdot (k \nabla T) = 0 \quad (6.2)$$

$$\frac{\partial}{\partial x} \left( w_x \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left( w_y \frac{\partial T}{\partial y} \right) + \frac{\partial}{\partial z} \left( w_z \frac{\partial T}{\partial z} \right) = \rho C_P \frac{\partial T}{\partial t} \quad (6.3)$$

$$n \cdot (w \cdot \nabla T) = h_{conv}(T_{amb} - T) + \sigma_{Tc} \varepsilon (T_{amb}^4 - T^4) \quad (6.4)$$

where in Equations (6.2)–(6.4), the density is  $\rho \left( \frac{kg}{m^3} \right)$ , constant specific heat is  $c_p \left( \frac{J}{kg \cdot m} \right)$ , and the absolute temperature is T(K). Time variable is set as t(s). A rectangular coordinate system (x, y, z) as in Equation (3) in anisotropic media can lead to numerous possible solutions.  $w, w_x, w_y, w_z \left( \frac{w}{m} \cdot K \right)$  are the main conductivity rates and conductivity rates which respectively for three coordinates (x, y, z) in 3D thermal modeling. The convection heat transfer is given by  $h_{conv} \left( \frac{w}{m^2 k} \right)$ . In Equation (4), a boundary condition that regarding to the thermal heat transfer from radiation and convection between the all the specimen surfaces and the ambient temperature has been indicated, where  $T_{amb}$  is an initial environment temperature which consider all the external environment as the same temperature;  $\sigma_{Tc} \left( \frac{w}{m^2 k^4} \right)$  is the Stefan–Boltzmann physical constant which links the temperature with energy;  $\varepsilon$  represents surface emissivity.

COMSOL's heat transfer module are adapted in this work which uses the 3D transient heat equation and provides the specimen temperature distribution, to simulate the specimen's thermal response after a heat pulse is applied to the specimen surface. A Gaussian power density distribution was employed to model the heat source, which closely approximates a real heat pulse from two flashes, i.e., a slightly larger heat accumulation at the center of the specimen. The heat pulse is applied as a radiative source that is partly absorbed by the specimen surface and transferred inside the specimen by heat conduction.

Table 6.1 briefly illustrates the physical properties of CFRP specimens and pulsed thermography parameters involved in the COMSOL simulation in this experiment. All properties and parameters in Table 6.1 are cited from the literature [14]. In this table,  $T_i$  is the initial temperature of an experimental specimen that was texted by PT.  $t_{end}$  means the whole cooling processing to observe the temperature evaluation after the heating pulse

application, also means the process time to finish computations. The time step is 0.0063 (the framerate: 158 frame/second). In this FEM stimulation, three types of heat flux transformation (radiation; convection; conduction) occurring during the thermal excitation period and cooling process are illustrated in Figure 6.2.

Table 6.2 and Table 6.3 provide the description of defect characteristics including the depth, size, shape of each defects, which characterized respectively two different group samples for training and testing. The training group has six CFRP samples (30 cm × 30 cm). Each sample has a different geometric distribution, depth, and size of defects.

Each row in every CFRP samples has the same depth of defect, which included 22 different constant values from 0.5 mm to 2.2 mm to be set from the first row in training sample 1 to the last row in training sample 6. We extracted five vectors from each defect region for GRU model training with depth estimation. The testing group has four CFRP samples (30 cm × 30 cm).

In order to differ from the training samples, the testing group consists of samples A, B, C, D. The depth of defects in the test samples ranges from [0.5 mm, 2.2 mm] but has a size that differs from that of the training samples. The CFRP geometrics of the structure of training sample 1 is indicated in Figure 6.3.

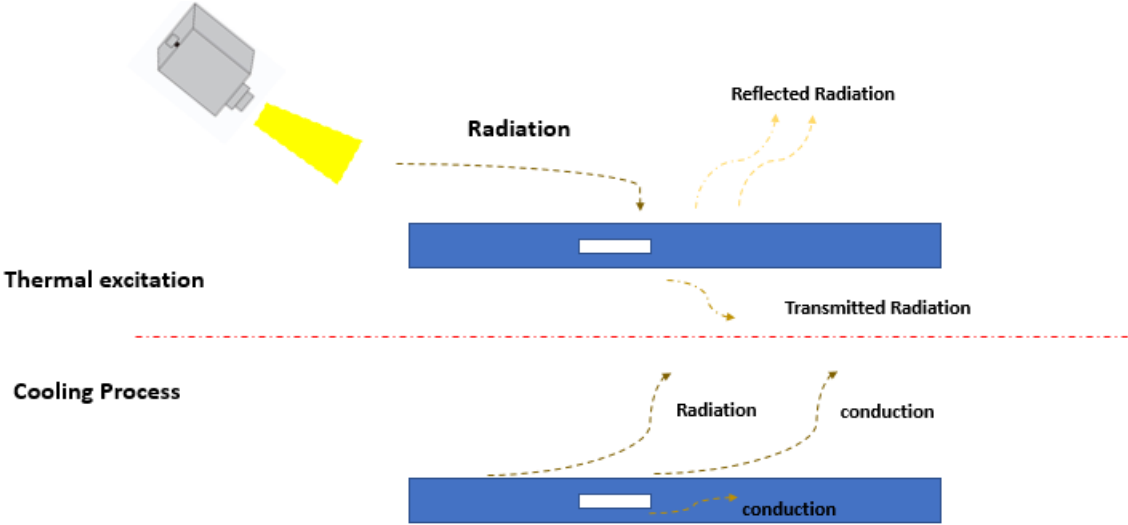


Figure 6. 2 Heat transfer during the thermal excitation and cooling process during FEM stimulation.

Table 6. 1 Physical properties and parameters of stimulation.

<b>Sign</b>	<b>Parameters in Experimental Simulation</b>	<b>Real Value</b>
$\rho$	Material density	$1500 \frac{kg}{m^3}$
$\varepsilon$	Emissivity	0.90
$C_p$	Constant specific heating	$1000 \frac{J}{kg.m}$
L	Specimen length	300 mm
W	Specimen width	300 mm
H	Specimen height	5 mm
$t_{end}$	Processing time to finish computations	9 s
$T_i$	The temperature from initialization	293.15 k
$T_{amb}$	The temperature from ambient	293.15 k
$W_{source}$	Heat source wattage	600 Watts
$H_{center}$	Heat source center	(15 cm, 15 cm)
P	Heat pulse density	$100,000 \frac{W}{m.m}$
h	The distance between the lamp and specimen	80 cm
S	Time step	0.0063 s
$t_{pulse}$	Heating time	0.0126 s

Table 6. 2 Defect characteristics of training samples.

<b>Sample</b>	<b>Row</b>	<b>Depth(mm)</b>	<b>Shape</b>	<b>Defect Size (mm)</b>
1	1	0.5 mm	Quadrangle	Size = 10; 15; 18
	2	0.6 mm	Round	Diameter = 18; 15; 5
	3	0.7 mm	Quadrangle	Size = 5; 10; 18
2	1	0.8 mm	Quadrangle	Size = 5; 10; 15
	2	0.9 mm	Round	Diameter = 18; 15; 10
	3	1.0 mm	Quadrangle	Size = 5; 15; 18
3	1	1.1 mm	Quadrangle	Size = 5; 10; 18
	2	1.2 mm	Round	Diameter = 15; 10; 5
	3	1.3 mm	Quadrangle	Size = 10; 15; 18
4	1	1.4 mm	Round	Size = 5; 15; 18
	2	1.5 mm	Quadrangle	Diameter = 18;10; 5
	3	1.6 mm	Round	Size = 5; 10; 15
5	1	1.7 mm	Round	Diameter = 10; 15; 18
	2	1.8 mm	Quadrangle	Size = 18; 15; 5
	3	1.9 mm	Round	Diameter = 5; 10; 18
6	1	2.0 mm	Round	Diameter = 5; 10 ;15
	2	2.1 mm	Quadrangle	Size = 18; 15;10
	3	2.2 mm	Round	Diameter = 5 ;15 ;18

Table 6. 3 Defect characteristics of testing samples

Sample	row#	Depth (Left; Middle; Right)	Shape	Defect Size (Left; Middle; Right) (mm)
A	1	Depth = 0.5; 0.8; 1.1	Quadrangle	Size = 3 ;16 ;13
	2	Depth = 0.6; 0.9; 1.2	Round	Diameter = 8; 3; 16
	3	Depth = 0.7; 1.0; 1.3	Quadrangle	Size = 13; 8; 16
B	1	Depth = 1.4; 1.7; 2.0	Quadrangle	Size = 8; 3; 16
	2	Depth = 1.5; 1.8; 2.1	Round	Diameter = 13; 8; 3
	3	Depth = 1.6; 1.9; 2.2	Quadrangle	Size = 16; 13; 8
C	1	Depth = 0.5; 0.8; 1.1	Round	Size = 4; 17; 14
	2	Depth = 0.6; 0.9; 1.2	Quadrangle	Diameter = 9; 4; 17
	3	Depth = 0.7; 1.0; 1.3	Round	Size = 14; 9; 4
D	1	Depth = 1.4; 1.7; 2.0	Round	Size = 9; 4; 17
	2	Depth = 1.5; 1.8; 2.1	Quadrangle	Diameter = 14;9;4
	3	Depth = 1.6; 1.9; 2.2	Round	Size = 17; 14; 9

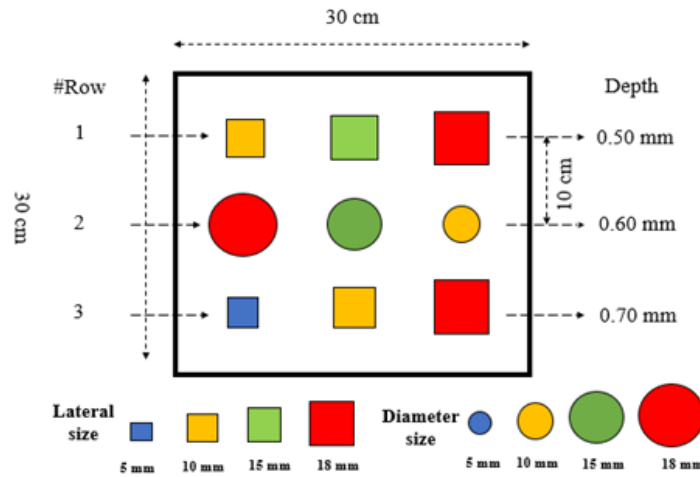


Figure 6. 3 Representative CFRP training sample 1 configuration.

### 6.5 Temperature and Thermal Contrast Curves

In transient thermography, previous researchers [6] concluded that the time  $t_c$  that corresponds to the maximum temperature contrast  $\Delta T$  has an approximately proportional relationship with the square of the depth of the defect ( $d^2$ ). Simultaneously, the proportionality coefficient of this relationship rested with the lateral size of the depth: the smaller the defects, the lower the maximum contrast  $\Delta T$  and earlier peak time. As indicated in Figure 6.4, an interesting case is highlighted. We observed the temperature evaluation of three defects of the same size (18 mm  $\times$  18 mm) but different depth such as (0.5 mm, 1.0

mm, 1.5 mm). Although the three defects have the same shape, it can be demonstrated that the shallower the defect, the higher the peak temperature value which can be obtained.

In this work, five contrast vectors have been extracted above each defect region on the surface from the different points in the defects to reduce the inaccurate influences caused by the small temperature variation. In Figure 6.5d as indicated in the below, in each defect region, the red cube is a defect point which localized on the corner of the upper left (C1); upper right (C2); center (C3); lower left (C4); lower right (C5). The blue cube is the reference point with respect to each defect point that we extracted. In Figure 6.5a–c, the data distribution of the thermal contrast curves is illustrated. C1-C5 are thermal contrasts that extracted from each defect point (in Figure 6.5d) based on the temperature difference between the defect point and reference sound point.

Notice that each training sequence to be processed in this work are extracted from the three quarters frames in the total frames of the one thermal sequence which already include the information of peak thermal contrast  $\Delta T_{max}$  and corresponding  $t_{max}$ . As a result, the last quarter frames of thermal curves were not extracted in this work which would show a dramatic decrease of the thermal contrasts in the graphs. To be notice that, all the thermal sequence stimulated from the COMSOL software based on the parameters as indicated on Table 6.2. All the extraction work was implemented in a program in MATLAB software. This partial extraction in thermal data could save the computational expense for the training in GRU to some extent.

A corresponding synthetic thermogram frame for training sample 2 in  $t = 3$  s generated from COMSOL is indicated at in Figure 6.6 a; nine artificial flat bottom Figure 6.6 b hole defects were embedded with different depths in the shapes of either circles or squares.

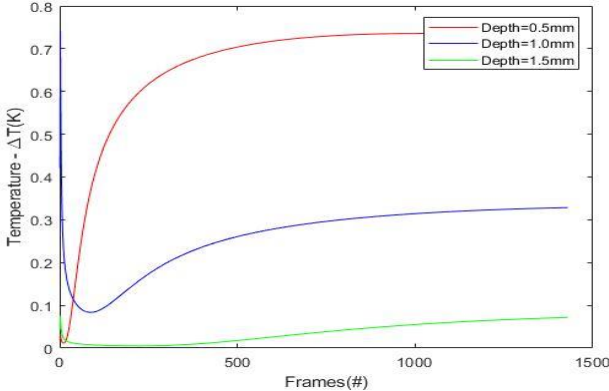
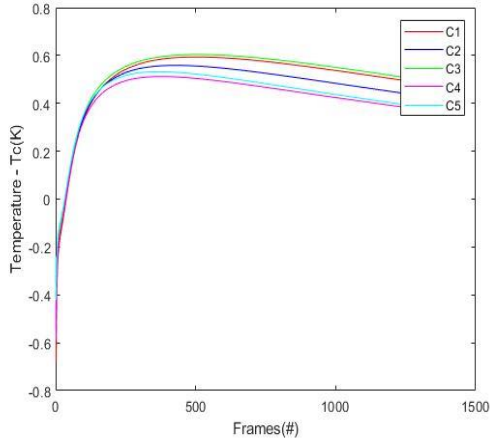
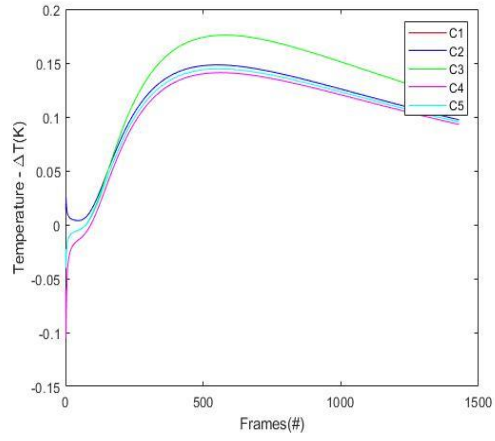


Figure 6. 4 Temperature contrast evaluation for defects located at different depths with the same size (0.5 mm, 1.0 mm, 1.5 mm) in training samples (the framerate: 158 frame/second)

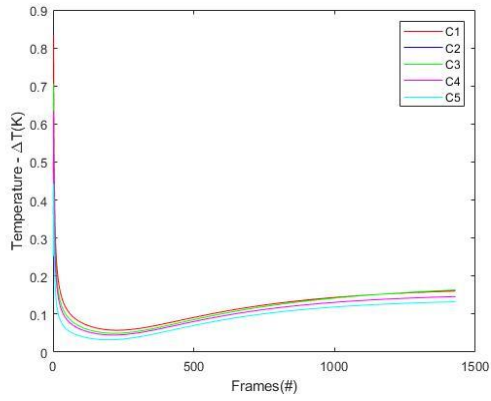




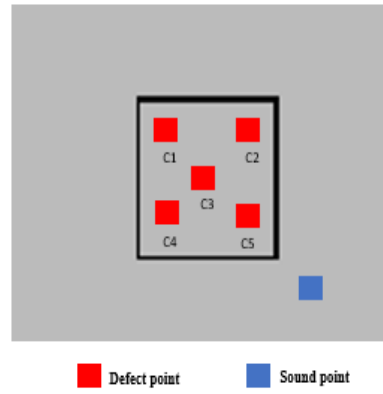
(a)



(b)

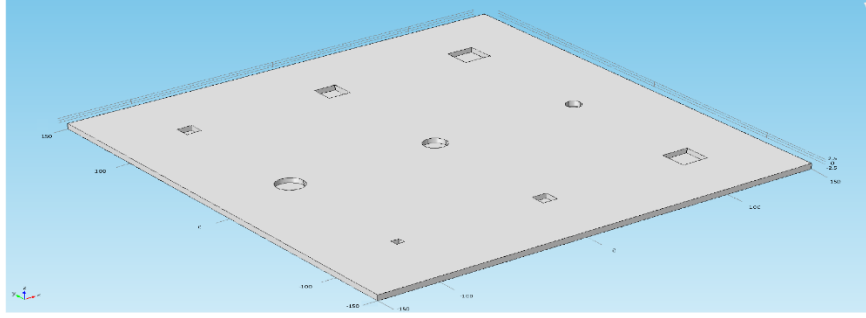


(c)

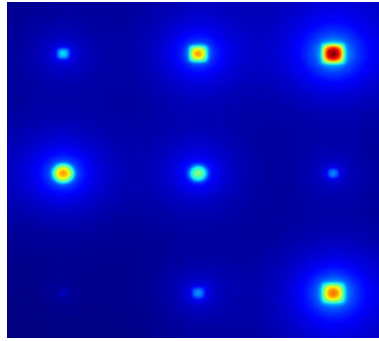


(d)

Figure 6. 5 Data distribution of temperature contrast evaluation for different depths in training samples (a) 0.5 mm; (b) 1.0 mm; (c) 1.5 mm (the framerate: 158 frame/second); (d) five defective point with corresponding reference point on a defect region



(a)



(b)

Figure 6. 6 (a) 3D printed defects geometrics are highlighted in training samples; (b) the corresponding synthetic colorful thermograms in  $t = 3$  s.

### 6.6 Proposed Strategy for Defect Depth Estimation

In this section, a defect depth estimation strategy has been proposed to detect and identify the depth of each defect in thermal images as indicated in Figure 6.7. This design of detection system originated from the GRU neural network. The infrared thermal module and simulations are provided by COMSOL (FEM simulation) to examine the depth of defects in pulsed thermography (PT). First, the synthetic thermal sequences are acquired from FEM (COMSOL) based on the heat transfer modeling in FEM. Then, several thermal contrast feature vectors are extracted from each defect region to feed into the input of GRU. The output of GRU consists of a unique node which estimated the depth during the training. The mean absolute error is chosen as the loss function with the GRU model in the equation as below Equation (5). In the end, the predicted depth output from the GRU is based on the feature extraction of thermal contrast vectors.

$$L_{cost} = \frac{\sum_{i=1}^m (y - \hat{y})^2}{m} \quad (6.5)$$

The GRU takes each vector at a time point in the input. This learning model was trained for 2500 epochs of each process. The training loss converged to the optimistic value and then flattened. In addition, the number of training curves (batch size) is set  $m$ .  $y, \hat{y}$  are the ground truth and estimated depth respectively.

### 6.6.1. Gated Recurrent Unit Model with Depth Estimator

Due to the time continuity of the thermal sequence, each frame collected from the experiment is related to the recent historical frame, therefore the time series memory deep learning model can be applied to the thermography data based on this feature. The GRU model is originally from RNN which is a time series model that can handle the continuous information such as thermal sequences [15]. During the cooling period of the thermal data, the temperature evaluation curves over time are acquired from the given infrared frames. The learned GRU model is able to distinguish whether the pixel is from a defective region or a sound region due to the training period. Therefore, the multiple units of GRU could be applied to extract the features of the temperature evaluations from the samples based on physical properties.

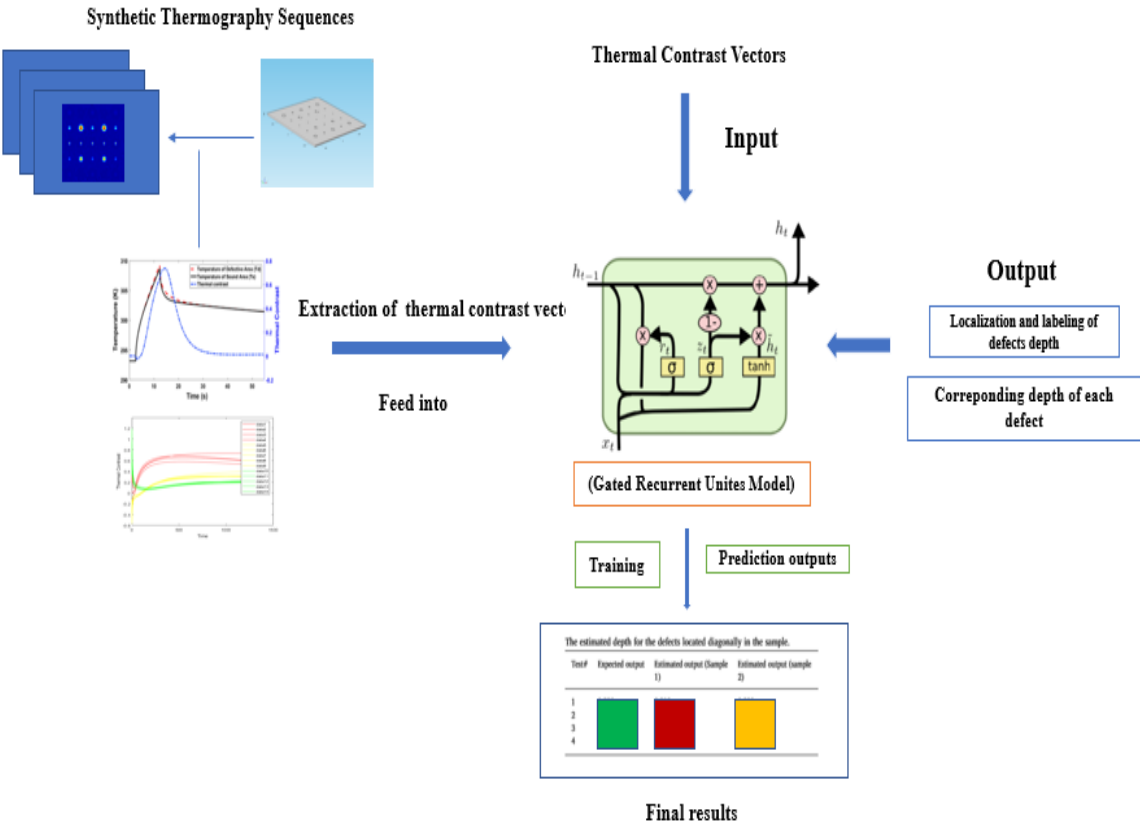


Figure 6. 7 GRU model-based defects depth estimation strategy.

The GRU neural network in Figure 6.8 are structured as follows [16]. In Equations (6.6)–(6.9),  $R_t, Z_t$  represent the update, reset gate units respectively.  $S^{(z)}, S^{(r)}, S$  are the weight factors for each gate unit.  $x_t$  is the current input.  $h_{t-1}$  is the input from the previous time step in the hidden state.  $h'_t$  is the input of current memory from the hidden state.  $U^{(z)}, U^{(r)}, U_o$  are the weight factors for previous time information  $h_{t-1}$ .  $\sigma(\cdot)$  represents the sigmoid function.  $\circ$  is denoted as the Hadamard product [17].

$$Z_t = \sigma(S^{(z)}(x_t) + U^{(z)}(h_{t-1})) \quad (6.6)$$

$$R_t = \sigma(S^{(r)}(x_t) + U^{(r)}(h_{t-1})) \quad (6.7)$$

$$h'_t = \tanh(S(z_t) + R_t \circ U_o(h_{t-1})) \quad (6.8)$$

$$h_t = Z_t \circ \tanh + (1 - Z_t) \circ h'_t \quad (6.9)$$

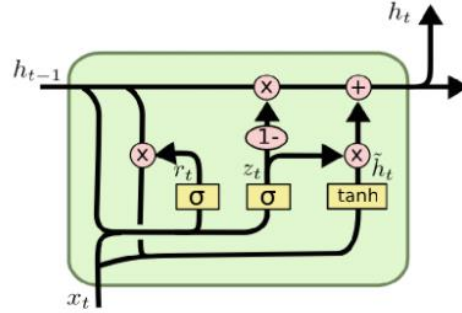


Figure 6. 8 Gated Recurrent Unit.

As shown in Figure 6.9, in this work, the original thermal sequences were reshaped into vectors. The particular thermal contrast vectors are directly fed into the input of the GRU network structure. Each thermal contrast vector in the defect region is decoded with the depth value of corresponding defect at the output of GRU based on the thermal properties from the training sequences. In order to select the points for the simulated thermal sequences to extract temperature curves vectors, 5 different locations inside each defect surface in these defective areas were selected. Since the temperature of the defect region is not even, these selected points accounted for small temperature variations and change above each defect surface region. Each thermal contrast vector is vectorized with the same length in the time of the thermal sequence. Therefore, the input values of GRU are fed into the particular thermal contrast vectors. The output from the decode section (to be connected with the fully connected layer) are set with the corresponding defect depth of each vector. The estimated

depth values in the defect region output from GRU are based on the thermal properties from the training sequences.

### 6.7. Experimental Validation and Results

#### 6.7.1. Inference and Training

In this research, the training processing on GPU (NVIDIA GeForce GTX 1080Ti) took about 30 min. The operating system was set as: Ubuntu 14. 04. CPU: i7-5930k. Memory: 64 GB. Adam was introduced as optimizer.

The whole training process was also conducted using the Adam optimization. Some main hyper parameters and training parameters are set as below: weight decay 0.0001, the learning rate 0.001 and learning momentum 0.9. In this work, the temperature variety and contrast reflect on each vector. Therefore, the LSTM time step was set to 1429 (1429 frames as a time step as input).

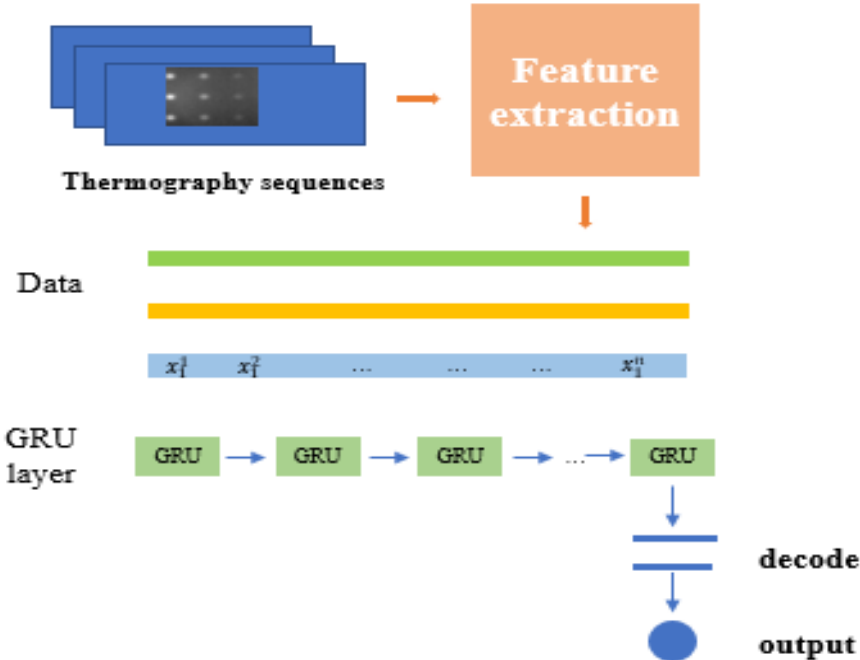


Figure 6. 9 The process of Gated Recurrent Unit (GRU) depth estimation.

#### 6.7.2. Data Processing

To reduce the overfitting issue, the cross-validation verification method is proposed in this study. First, we collected 270 thermal contrast curves from the simulation data (5 thermal contrast curves extracted from each defect averagely). Then, we shuffled all of the data and

spilt 80% of the data to train with the GRU model, while 20% were utilized to validate the performance of the GRU.

All obtained training data (the thermal contrast evaluation curves) were normalized and truncated as a fixed length of duration. Simultaneously, the input data were normalized by subtracting from the mean value of the thermal curves ( $\mu_{T_c}$ ) and dividing by the standard deviation of thermal contrast ( $\sigma_{T_c}$ ) using Equation (10).

$$\widehat{T}_{c,n}^i(t) = \frac{\widehat{T}_c(t) - \mu_{T_c}}{\sigma_{T_c}} \quad (6.10)$$

## 6.8 Depth Estimation Results and Validation

### 6.8.1. Result Analysis—Mean Absolute Error (MAE)

In statistics, the mean absolute error (MAE) [18] is one of the metrics to evaluate how close the forecasts are to the eventual outcomes. In the machine learning field, it can indirectly reflect the accuracy and performance of the machine learning model (GRU). In this work, we adapted the MAE to assess the performance of the GRU for depth estimation with infrared thermography. As we can see from Figure 10a, we trained the GRU to estimate the data from testing before the data normalization. The obtained training loss was 0.055. The MAE converged to 0.0165 mm. The error between the predictive value and actual value was within the range of  $[-0.17 \text{ mm}, 0.17 \text{ mm}]$ . After the standard deviation normalization for all the distributed data in Figure 10b, the predictive value tended to approach the actual depth for the defect. The MAE error shrank to the range of  $[-0.11 \text{ mm}, 0.11 \text{ mm}]$  and the training loss converged to 0.0295. This showed an acceptable performance with an improved estimate of the depth by the GRU model with standard deviation normalization.

In Table 6.4, the estimated output 1 and the MAE 1 were obtained from raw data without normalization (average values over the two estimated outputs from CFRP specimen groups: (A, C); (B, D)). Then, the estimated output 2 and the MAE 2 resulted from raw data of the same procedure but with normalization.

Based on Table 6.2, the calculated accuracy in the GRU model for the depth estimation reached 90% before data normalization (standard deviation). After normalization, the results provided an accuracy greater than 95%. This performance demonstrated that the GRU enabled a high performance for accurate depth estimation. This estimation is attributed to the ideal environment without experimental issues (noise; defective pixels). As shown in the Figure 6.11 below, the thermal data distribution from training (before; after) normalization has been indicated. In the data distribution (a), each group of color data curves (yellow; red;

green) represents a different specific depth from defects. The thermal data were normalized by Equation (6.10) in Figure 6.10 b. The distinguishable features of difference between the depths were recognized by the following principle: shallow defects had greater maximum thermal contrasts that occur earlier than deep defects. These results outperformed the previous works obtained from [7] for depth estimation in automated infrared thermography with regular neural networks.

Table 6. 4 The results of depth estimation of defects located in the designated specimen.

<b>Sample</b>	<b>Expected Output (mm)</b>	<b>Estimated Output 1 (mm)</b>	<b>MAE1 *</b> <b>(mm)</b>	<b>Estimated Output 2 (mm)</b>	<b>MAE2*</b> <b>(mm)</b>
A C	0.5	0.522	0.022	0.507	0.007
A C	0.6	0.604	0.004	0.603	0.003
A C	0.7	0.708	0.008	0.706	0.006
A C	0.8	0.814	0.014	0.807	0.007
A C	0.9	0.912	0.012	0.913	0.013
A C	1.0	1.041	0.041	1.025	0.025
A C	1.1	1.109	0.009	1.011	0.011
A C	1.2	1.222	0.022	1.218	0.018
A C	1.3	1.318	0.018	1.314	0.014
B D	1.4	1.420	0.020	1.418	0.018
B D	1.5	1.514	0.014	1.509	0.009
B D	1.6	1.630	0.030	1.619	0.019
B D	1.7	1.718	0.018	1.715	0.015
B D	1.8	1.820	0.020	1.817	0.017
B D	1.9	1.918	0.018	1.920	0.020
B D	2.0	2.013	0.013	2.005	0.005
B D	2.1	2.112	0.012	2.010	0.010
B D	2.2	2.225	0.025	2.222	0.022

\* MAE means mean square error. An average over sample for absolute different between actual and predicted observation.

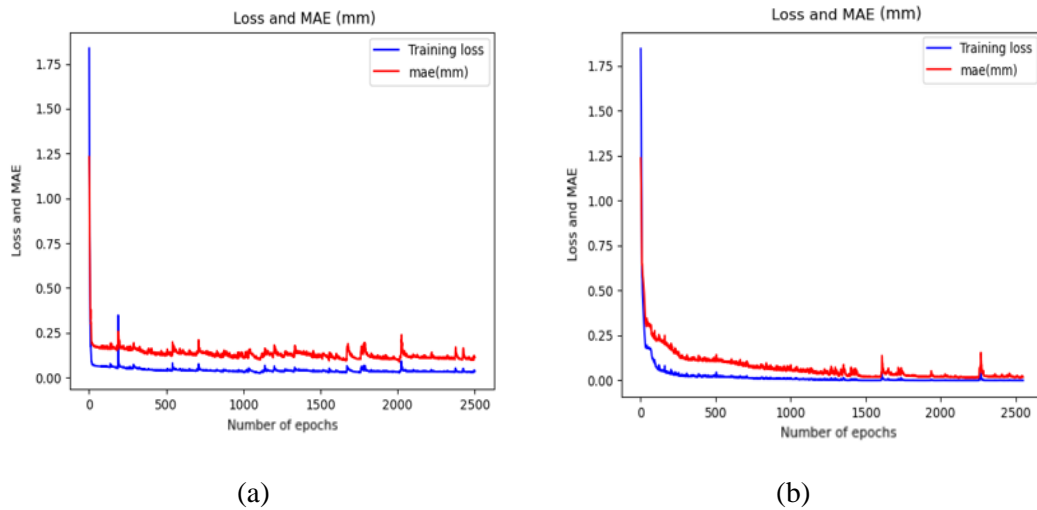
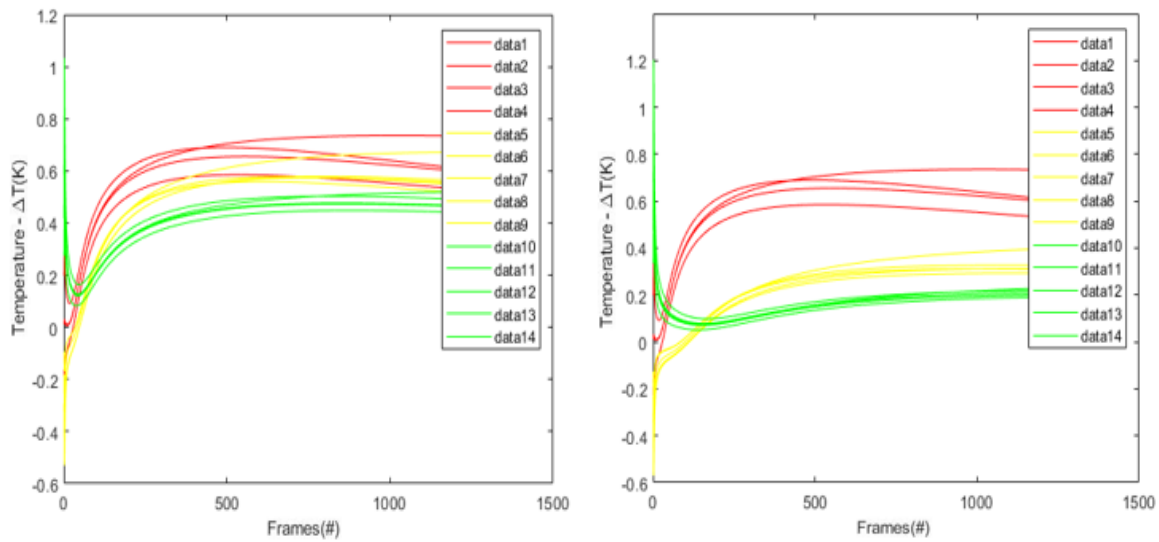


Figure 6. 10 The mean absolute error and the training loss with GRU before (a) and after (b) standard deviation normalization.

## 6.9. Conclusions

This work elaborated the complicated and non-linear issues of evaluating defect depths in composite materials via infrared thermography with a GRU learning model. The methodology proposed here employed a GRU model combined with pulsed thermography to analyze the depth of defects. The simulated samples provide an economical platform for GRU training and depth estimation. Quantitative analysis of defect depth (subsurface features) has been evaluated by a GRU based statistical method through developed neural network modeling and cross validation experimental verification. It has been proven that the GRU modeling can produce an advanced depth detection. For future work, the experimental data have to be evaluated for the robustness of the GRU model. Further, other types of deep learning models and modified versions of the GRU model have to be applied to increase the depth estimation ability in this topic.





(a)

(b)

Figure 6. 11 The data distribution before (a) and after (b) standard deviation normalization from all the selected locations for training (the framerate: 158frame/second).

## References

- [1] D'Accardi, E.; Palano, F.; Tamborrino, R.; Palumbo, D.; Tati, A.; Terzi, R.; Galietti, U. Pulsed phase thermography approach for the characterization of delaminations in cfrp and comparison to phased array ultrasonic testing. *J. Nondestruct. Eval.* **2019**, *38*, 20.
- [2] Rajic, N. Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures. *Compos. Struct.* **2002**, *58*, 521–528.
- [3] Martin, R.E.; Gyekenyesi, A.L.; Shepard, S.M. Interpreting the results of pulsed thermography data. *Mater. Eval.* **2003**, *61*, 611–616.
- [4] Shepard, S.M.; Lhota, J.R.; Rubadeux, B.A.; Wang, D.; Ahmed, T. Reconstruction and enhancement of active thermographic image sequences. *Opt. Eng.* **2003**, *42*, 1337–1342.
- [5] Weng, J.; Zhang, Y.; Hwang, W.S. Candid covariance-free incremental principal component analysis. *IEEE Trans. Pattern Anal. Machine Intell.* **2003**, *25*, 1034–1040.
- [6] SUN J G. Analysis of pulsed thermography methods for defect depth prediction[J]. *Journal of heat transfer*, 2006, 128(4): 329-338.

- [7] Darabi, A.; Maldague, X. *Neural network-based defect detection and depth estimation in TNDE. Ndt E Int.* **2002**, *35*, 165–175.
- [8] Chung, J.; Gulcehre, C.; Cho, K.H.; Bengio, Y. *Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv* **2014**, arXiv:1412.3555.
- [9] Hochreiter, S.; Schmidhuber, J. *Long short-term memory. Neural Comput.* **1997**, *9*, 1735–1780.
- [10] Zissis, G.J. *Infrared technology fundamentals. Opt. Eng.* **1976**, *15*, 156484.
- [11] Saeed, N.; Abdulrahman, Y.; Amer, S.; Omar, M.A. *Experimentally validated defect depth estimation using artificial neural network in pulsed thermography. Infrared Phys. Technol.* **2019**, *98*, 192–200.
- [12] Peeters, J.; Steenackers, G.; Ribbens, B.; Arroud, G.; Dirckx, J. *Finite element optimization by pulsed thermography with adaptive response surfaces. In Proceedings of the 12th International Conference on Quantitative Infrared Thermography (QIRT), Bordeaux, France, 7–11 July 2014.*
- [13] Lopez, F.; de Paulo Nicolau, V.; Ibarra-Castanedo, C.; Maldague, X. *Thermal–numerical model and computational simulation of pulsed thermography inspection of carbon fiber-reinforced composites. Int. J. Therm. Sci.* **2014**, *86*, 325–340.
- [14] Pilling, M.W.; Yates, B.; Black, M.A.; Tattersall, P. *The thermal conductivity of carbon fibre-reinforced composites. J. Mater. Sci.* **1979**, *14*, 1326–1338.
- [15] Chung J, Gulcehre C, Cho K, et al. *Gated feedback recurrent neural networks[C]//International conference on machine learning. 2015: 2067-2075.*
- [16] Gulcehre C, Cho K, Pascanu R, et al. *Learned-norm pooling for deep feedforward and recurrent neural networks[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2014: 530-546.*
- [17] Horn, R.A. *The hadamard product. Proc. Symp. Appl. Math.* **1990**, *40*, 87–169.
- [18] Kassam, S. *Quantization based on the mean-absolute-error criterion. IEEE Trans. Commun.* **1978**, *26*, 267–270.

# General conclusions & perspectives

## General conclusion

The objective of the present thesis was implementing an automatic defect detection system via the deep learning algorithms. The main chapters in this thesis elaborated the routine of using deep learning approaches to achieve the goals of defect detection which include defect visibility enhancement, automatic defect detection and segmentation, defect depth estimation in infrared thermography.

Overall, the proposal and implementation of deep learning with Infrared non-destructive testing was developed a more intelligent and automated for defect detection and analysis system. In this thesis, we systematically introduced the principle, the methodology, the experimental set up, the research results, discussion, summaries and fully exploit the capability of each proposed deep learning models' methods for infrared thermography of defect detection.

The methods proposed during this Ph.D. project illustrated several innovative points and contributions for detection principles:

To improve the visibility of defects in pulsed thermography by introducing Generative adversarial network. We proposed the GANs method to reconstruct conspicuous images through the underlying features extraction in temperature evolutions of Pulse Thermography. Due to the reduced feature dimension, the deep autoencoder architecture from the GANs network (DCGANs; WGANs) combined with the original Partial Least Square Thermography has further reduced the noise from the original infrared data and enhanced the accuracy of defect visibility. The deep learning algorithm was combined with the dimension reduction method in the field of thermographic-defect detection developed the GANs-based data-augmentation strategy model named PLS-GANs (PLS-WGAN; PLS-DCGAN) to further upgrade their results with defect enhancement analysis and detection performance.

In order to acquire enough training data for identifying defects, we adapted a synthetic data generation pipeline to generate a series of simulated thermal data. We used transfer learning strategy to extract meaningful features, and fine-tuned the features with specific thermal sequences for the purpose of offsetting limited thermal data and acquiring enough training data for identifying defect. This gave a rise to a method where the deep learning architecture is combined with a small amount of synthetic thermal data for defect segmentation. The proposed architecture is based on a Mask-RCNN that has been applied in natural image segmentation. We collected our thermal database from the experimental results and synthetic data in the Finite Element Model (COMSOL). In the end, the network yields a better performance when it is fused with synthetic data for training.

We provided three types of deep learning detection models (instance segmentation; objective detection; semantic segmentation) to attempt to perform automatic detection and segmentation applications in infrared thermography due to data augmentation with elastic deformations. Six spatial deep learning models: instance segmentation (Mask R-CNN; Center-mask); the autoencoder format semantic segmentation (U-net; Res-u-net); the object localization model (YOLO-V3; Faster-RCNN) are applied for defect detection in infrared thermography. The key evaluation metrics are analyzed based on the detection results from DL models: Probability of Detection (POD); Mean-average Precision(mAP); Time speeding complexity. The results also show that DL models have the reliability of automatic defect detection assessment. Comparing the experimental results among these methods, the Center-Mask is the most promising method for accuracy whereas YOLO-V3 has the faster time frame speed based on the infrared testing database evaluation.

The methodology proposed here employed a GRU model combined with pulsed thermography to analyze the depth of defects. An algorithm based on the recurrent neural network has been proposed in order to estimate the defect depth of testing samples from the active thermography data. We explicitly illustrated the architecture of the fully Gated Recurrent Units (GRUs) spatial and temporal sequence model for thermography data to be trained end to end. A cross-validation routine was introduced to improve the efficacy of the algorithm in the case of over-fit. The regression training from GRUs elaborated the complicated and non-linear issues of evaluating defect depths in composite materials via infrared thermography with a GRU learning model. Based on the analyses of the predicted results from the GRUs model for depth estimation. The predictive value tended to precisely approach the actual depth for each defect. The MAE error shrank to the absolute value of 0.11mm with a normalization process. It further indicated the acceptable performance with the estimation of depth from Recurrent neural network (RNN)-GRU model.

Moreover, the scientific challenge of the successful application in the projects discussed in the introduction section has been solved to a certain degree in this project:

1. The limited training data of thermography for deep learning. Due to issues such as the presence of noise, limited resolution, and probing of depths in TINDE, it is more difficult to precisely and fully extract defect information. We introduced the generation pipeline from the synthetic data and the GANs network which could further compensate for the databases limitation and enhance the capability of defect detection.
2. For the defect characterization issue, it is complex to combine the information of thermal contrast and pixels of images with the depth in each pixel of defect to train the neural network. We introduced the Gated recurrent Units (GRUs) which was trained with thermal contrast

curves from different detect depths to extract different thermal behavior and output the depth of defects automatically.

3. In order to analyze the cases of samples which contained noise and non-uniform heating, weak detection signal, limited IR data; etc, we introduced the defects enhancement methods: Partial least square-GANs network (PLS-GANs) , which allow us to remove the noise and obtain the defect feature based on the autoencoder architecture (feature extraction mechanism) and dimension reduction strategy from GANs (to further decompose the original thermal data into meaningful feature components) for defect detection and improve the performance in comparison to the original PLST methods.

### **Future Perspectives**

For the future work in this project, more detailed experimental results and discussion could demonstrate that deep learning algorithms could be effective for defect detection in infrared thermography.

#### a) New spatial- temporal combination modeling

In this project, we have already proposed and implemented diverse types of spatial and temporal deep learning modelling respectively for the analysis of the features of infrared databases for defect detection tasks. It would be more efficient in the future to deal with complex conditions. This could lead to the use of the proposed methods to design new spatial-temporal combination modeling which would be more flexible and compatible with the characteristics of infrared thermal sequences.

#### b) Application of more advanced data fusion methods.

Data fusion involves integrating multiple types of databases to extract the meaningful, consistent, accurate resource and features, directly obtain from each individual database resource. This can help the deep learning algorithms to further train the objective feature without high training costs from the original databases.

#### c) Transfer learning Strategy

Transfer learning is a technology focused on obtaining knowledge while solving a problem by applying one to another issue. It can be used for defect representation in thermal images through a pre-trained network and learning framework. Using a transfer learning strategy with a pre-trained Convolutional Neural Network Res-net as an unsupervised feature extractor for analyzing defects in specimens and fine-tuning with specific thermal sequences could further improve training and evaluation. A pre-trained CNN could be used for the extraction of the vectorized features along with prediction scores from the output of models

for defects, then a deep residual learning framework (Resnet) could be used to segment each defect at each region and depth of defects correspondingly

d) Possible modification of loss function of deep Learning model

The loss function could be a crucial factor that influences the training capability of training models from deep learning. This could directly indicate the difference between the predicted values and objective values. Different types of loss functions could be optimized based on the problem that we target. Due to the limitation and low visibility (semantic) information from the infrared databases, it might be difficult to obtain convergence or obtain a satisfactory performance from the results when the inner defects or the complex and irregular surface is evaluated from the algorithms. Therefore, to modify the loss function from the deep learning model may benefit and overcome the challenge when the defects are mixed with complex samples and significantly noises.

e) Experimental conditions for the feature acquisition

As we predicted, the background noises from defect pixels or non-uniform emissivity influence the possibility to acquire suitable results during the infrared thermography experiment. This influences the model that we introduced from deep learning to learn and extract the features based on this foundation.

Meanwhile, the experimental setup for the acquisition procedure from infrared thermography involved an infrared camera, heating source, macro lens. The optimized physical experimental conditions are a significant part of the system attempting to obtain pure samples for the post- images processing especially for real world applications.

An important factor to be considered in future efforts is to create a setup that has optimized experimental conditions to further eliminate the physical factors interfering in order to clearly obtain the features in different thermal evaluation experiments for deeper subsurface defect detections. It would not only dramatically benefit the acquisitions of the training database for the model in the process but also reduce (or even eliminate) the learning difficulty and improve the quality of training and accuracy of predictions from deep learning.

## Appendix A List of publications

Journal articles and conference papers related to but not limited to the work in this thesis are listed as follows:

Journal Articles:

— **Qiang Fang** and Xavier Maldague, A Method of Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning, MDPI **Appl. Sci.** 2020, 10, 6819; doi:10.3390/app10196819.

— **Qiang Fang**, Ibarra-Castanedo C, Maldague X. Automatic defects segmentation and identification by deep learning algorithm with pulsed thermography: Synthetic and experimental data[J]. **Big Data and Cognitive Computing**, 2021, 5(1): 9.

— **Qiang Fang**, Ibarra-Castanedo C, Duan Yuxia, Jorge Erazo-Aux, Iván Garrido, Maldague X. Defect enhancement and image noise reduction analysis using Partial Least Square-Generative Adversarial Networks (PLS-GANs) in Pulsed Thermography, was accepted by the **Journal of Nondestructive Evaluation**, in August, 2021.

— **Qiang Fang**, Ibarra-Castanedo C, Maldague X. Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data and is going to submit **Journal of Applied Remote Sensing**, in August, 2021.

— Garrido González I, Lagüela López S, **Qiang Fang**, Introduction of the combination of thermal fundamentals and Deep Learning for the automatic thermographic inspection of thermal bridges and water-related problems in infrastructures, was submitted for **Quantitative InfraRed Thermography Journal**, on July 27<sup>th</sup>, 2021.

— Duan Y, Liu S, Hu C, Hu J, Zhang X, Yan Y, Tan N, Zhang C, Maldague X, **Fang Q**, Ibarra-Castanedo, Chen D, Li X, Meng J. Automated defect classification in infrared thermography based on a neural network[J]. **NDT & E International**, 2019, 107: 102147.

Journal Articles in Preparation:

— **Qiang Fang**, Maldague,X, “Defect visibility enhancement of step-wage CFRP by Generative Adversarial Network with Dynamic Lin Scan Thermography: Experimental and Syntenic data.”

- **Qiang Fang**, Maldague X, “Defect depth estimation in infrared thermography by deep learning algorithm with experimental pulsed thermography data.”
- **Qiang Fang**, Maldague X, “Defect visibility enhancement and analysis using Convolutional Auto-Encoders in pulsed thermography data.”

Conference Papers:

- **Qiang Fang**, Ba Diep Nguyen, Clemente Ibarra Castanedo, Yuxia Duan, Xavier Maldague “Automatic defect detection in infrared thermography by deep learning algorithm”, **SPIE Defense+Commercial Sensing**, Proceedings Volume 10214, Thermosense: Thermal Infrared Applications XXXIX; 102140T (2020).
- **Qiang Fang**, Farima. Abdolahi. Mamoudan, Celment. Ibarra. Castanedo, Xavier. Maldague “Defect depth estimation in infrared thermography by deep learning algorithm”. **2020 Structural Health Monitoring & Nondestructive Testing conference, 22th to 23th of November 2020, Quebec City, QC.**
- **Qiang Fang**, Xavier Maldague Iván Garrido, Jorge Erazo Aux, Clemente Ibarra-Castanedo “Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data” **15th Quantitative InfraRed Thermography Conference, 21st to 30th of September 2020, University of Porto, Portugal.**
- **Qiang Fang**, Clemente Ibarra Castanedo, Xavier Maldague” Automatic segmentation and Identification of defects by deep learning Algorithms from Pulsed thermography Data” was submitted to **20th World Conference on Non-Destructive Testing, Songdo Convensia, Incheon, South korea.**
- Garrido González I, Lagüela López S, **Fang Q**, Combination of thermal fundamentals and Deep Learning for infrastructure inspections from thermographic images. Preliminary results. **15th Quantitative InfraRed Thermography Conference, 21st to 30th of September 2020, University of Porto, Portugal.**

## First page of published papers issued in thesis

The following are the first pages of published papers which are issued in the thesis:



Article

# Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data

Qiang Fang <sup>\*</sup>, Clemente Ibarra-Castanedo  and Xavier Maldague <sup>\*</sup>

Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Université Laval, 1065, av. de la Médecine, Québec, QC G1V 0A6, Canada; clemente.ibarra-castanedo@gel.ulaval.ca

\* Correspondence: qiang.fang.1@ulaval.ca (Q.F.); Xavier.Maldague@gel.ulaval.ca (X.M.)

**Abstract:** In quality evaluation (QE) of the industrial production field, infrared thermography (IRT) is one of the most crucial techniques used for evaluating composite materials due to the properties of low cost, fast inspection of large surfaces, and safety. The application of deep neural networks tends to be a prominent direction in IRT Non-Destructive Testing (NDT). During the training of the neural network, the Achilles heel is the necessity of a large database. The collection of huge amounts of training data is the high expense task. In NDT with deep learning, synthetic data contributing to training in infrared thermography remains relatively unexplored. In this paper, synthetic data from the standard Finite Element Models are combined with experimental data to build repositories with Mask Region based Convolutional Neural Networks (Mask-RCNN) to strengthen the neural network, learning the essential features of objects of interest and achieving defect segmentation automatically. These results indicate the possibility of adapting inexpensive synthetic data merging with a certain amount of the experimental database for training the neural networks in order to achieve the compelling performance from a limited collection of the annotated experimental data of a real-world practical thermography experiment.

**Keywords:** automatic defect detection; infrared thermography; deep learning algorithms; Non-Destructive Evaluation (NDE); supervised learning; image processing; data augmentation; Finite Element Models (FEM)



**Citation:** Fang, Q.; Ibarra-Castanedo, C.; Maldague, X. Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data. *Big Data Cogn. Comput.* **2021**, *5*, 9. <https://doi.org/10.3390/bdcc5010009>

Received: 29 December 2020

Accepted: 19 February 2021

Published: 26 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).


## 1. Introduction

Quality evaluation is playing a fundamental role for modern industrial production and manufacturing processing. The demand for the inspection of materials with respect to the possible presence of defects, damage, and flaws has increased due to the wide use of composite materials and metals in industries such as aerospace. The method used to inspect defects via manual, visual evaluation involving humans can be hampered by the human fatigue and subjectivity. In order to meet the need of high-quality production and maintain the stringent high-quality level, an advanced inspection system is becoming more and more essential for structural health monitoring and production lines application. Automated quality control [1] can be applied in the industrial field to facilitate the consistent and efficient inspection. The high inspection rate and the inspection procedure without human involvement are the main advantages of automated inspection systems [2].

Non-Destructive Testing (NDT) [3] encompasses a group of inspection techniques that are intended to assess the integrity of an object without causing any kind of modification or permanent damage. Infrared testing (IRT) is an NDT technique in which the differences in thermo-physical properties between the inspected object sound material and possible surface and/or subsurface defects are exploited through the use of an infrared camera. However, automatically identifying defects in materials via IRT still remains an ongoing and challenging task in the image processing domains.

Article

# A Method of Defect Depth Estimation for Simulated Infrared Thermography Data with Deep Learning

Qiang Fang \*  and Xavier Maldague \*

Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Université Laval, 1065, av. de la Médecine, Québec, QC G1V 0A6, Canada

\* Correspondence: qiang.fang.1@ulaval.ca (Q.F.); Xavier.Maldague@gel.ulaval.ca (X.M.);  
Tel.: +1-581-922-0301 (Q.F.); +1-418-656-2962 (X.M.)

Received: 25 August 2020; Accepted: 25 September 2020; Published: 29 September 2020



**Abstract:** Infrared thermography has already been proven to be a significant method in non-destructive evaluation since it gives information with immediacy, rapidity, and low cost. However, the thorniest issue for the wider application of IRT is quantification. In this work, we proposed a specific depth quantifying technique by employing the Gated Recurrent Units (GRUs) in composite material samples via pulsed thermography (PT). Finite Element Method (FEM) modeling provides the economic examination of the response pulsed thermography. In this work, Carbon Fiber Reinforced Polymer (CFRP) specimens embedded with flat bottom holes are stimulated by a FEM modeling (COMSOL) with precisely controlled depth and geometrics of the defects. The GRU model automatically quantified the depth of defects presented in the stimulated CFRP material. The proposed method evaluated the accuracy and performance of synthetic CFRP data from FEM for defect depth predictions.

**Keywords:** NDT Methods; defects depth estimation; pulsed thermography; Gated Recurrent Units

## 1. Introduction

Non-destructive evaluation (NDE) has emerged as an important method for the evaluation of the properties of components or systems without damaging their structure. Several state of the arts methodologies such as Pulsed Phase Thermography (PPT) [1], Principal Component Thermography (PCT) [2], Differential of Absolute Contrast (DAC) [3], Thermographic Signal Reconstruction (TSR) [4], as well as Candid Covariance Free Incremental Principal Component Thermography [5], have been implemented to process thermographic sequences and improve the defect visibility. These techniques can be beneficial for qualitative analysis of composite materials. However, to introduce a method for conducting quantitative study (defect depth estimation) with deep learning is a novel topic to be explored.

Quantitative analysis is playing an important role in the modern industrial field of non-destructive testing (NDT). Defect characterization is one of the current topics of interest in quantitative data analysis of active thermography. In this topic, thermographic information involves extracting quantitative subsurface properties from defects such as defect depth, lateral size, thermal resistivity, from an experimental Thermal Non-destructive Testing (TNDT) dataset utilized to characterize defects. Several approaches have already been proposed to analyze the depth in the region of defects in pulsed thermography. In general, these approaches evaluate the defects depth by using the maximum thermal contrast  $C_{max}$ , the instant of maximum thermal contrast  $Tc_{max}$  or artificial neural networks to analyze the depth of defects based on mathematical equations. The Peak Temperature Contrast Method [6] estimated the depth based on the characteristic time from peak contrast. The peak contrast corresponding to the maximum contrast has a proportional correlation with the square of the defect depth. Daribi et al. [7] proposed neural networks for defect characterization of defect depth. The results

# AUTOMATIC DEFECT DETECTION IN INFRARED THERMOGRAPHY BY DEEP LEARNING ALGORITHM

Qiang. Fang <sup>\*a</sup>, Ba Diep Nguyen <sup>a</sup>, Clemente. Ibarra. Castanedo, <sup>a</sup> Yuxia. Duan <sup>b</sup>,Xavier. Maldague <sup>a</sup>

<sup>a</sup> Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Universite Laval, 1065, av. de la Médecine, Québec (QC), Canada, G1V 0A6;

<sup>b</sup> School of Physics and Electronics, Central South University, 932 Lushan South Road, Changsha, Hunan 410083, China Abstract

Non-destructive Evaluation (NDE) is a field that is used to identify all kinds of structural damage in an object of interest without resulting in any permanent damage or modification to the object. This field has been intensively investigated for many years. Among several research topics in this field, the supervised defect detection methods are among the most innovative and challenging. In recent years, the deep learning field of artificial intelligence has made remarkable progress in image processing applications. Deep learning has shown its ability to overcome most of the disadvantages suffered by previous existing approaches in a great number of applications. In this paper, we propose a deep learning architecture based on infrared thermography inspection intended to automatically identify defects (including internal and invisible cracks, delamination, etc.) efficiently and accurately. We studied the proposed deep learning algorithms to achieve automatic defect detection and precise localization (subsurface defects case) from different thermal image sequences. To evaluate the efficiency and robustness of the proposed methodology, specimens containing artificial defects were selected for experimental configuration.

**Keywords:** Automatic Defect detection, Infrared Thermography, Deep learning algorithms, Non-destructive Evaluation (NDE), Supervised learning, Image processing.

## 1. INTRODUCTION

Quality management (QM) is playing a crucial role in modern industrial production fields. A qualified quality management and controlling system can be a significant technology advancement for industry and manufacturing applications. In order to preserve the health structural monitoring, the application of visual inspection systems becomes more and more essential in the production lines. However, the inspection services involving manual inspection during the stages of quality control can be hampered by the fatigue of inspectors. In this case, automatic quality control and defect detection become more vital so as to improve the inspection rates and achieve cost-effective [1] condition monitoring.

Nondestructive Evaluation (NDE) is a wide group of analysis techniques used in industry to evaluate the properties of material without jeopardizing the quality of the products. Infrared thermography, which is one of many NDT techniques, evaluates subsurface defects and hidden structures for the quality control of materials such as metals and composites among others. Due to the thermal wave propagation and attenuation, each thermal non-homogeneity perturbs the thermal wave propagation through the specimen in comparison to the surrounding sound non-defect area, which results in differences in thermal signatures on the object surface. An infrared camera can record this thermal perturbation, and the results can be analyzed to obtain further information. For defect identification and detection, several state of the art techniques such as Pulsed Phase Thermography (PPT) [2], Principal Component Thermography (PCT) [3], Difference of Absolute Contrast (DAC) [4], Thermographic Signal Reconstruction [5], as well as Candid Covariance Free Incremental Principal Component Thermography [6], have demonstrated to be powerful tools that have been well-documented and provide noticeable results in Infrared Thermography. However, these methods are pattern-based unsupervised methods [7] [26-27] which extract more distinctly defect information from time and frequency domains. Compared to unsupervised learning methods, a supervised learning-based algorithm which automatically identifies and detects defects is an interesting and unexplored topic in the IRT field with automated inspection.

In this work, we exploit a proposed supervised deep learning algorithm -YOLO-v3 during the infrared nondestructive evaluation with an automatic defect detection task (subsurface defects case). To the best of our knowledge, there are

Thermosense: Thermal Infrared Applications XLII, edited by Beate Oswald-Tranta,  
Joseph N. Zalameda, Proc. of SPIE Vol. 11409, 114090T · © 2020 SPIE  
CCC code: 0277-786X/20/\$21 · doi: 10.1117/12.2555553

Proc. of SPIE Vol. 11409 114090T-1

Downloaded From: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie> on 23 Apr 2020  
Terms of Use: <https://www.spiedigitallibrary.org/terms-of-use>



## Automatic Defects Segmentation and Identification by Deep Learning Algorithm with Pulsed Thermography: Synthetic and Experimental Data

by Q. Fang\*, X. Maldague\*, I. Garrido\*\*, J. H. Erazo\*\*\*, C. I. Castanedo\*,

\* Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Université Laval, 1065, av. de la Médecine, Québec (QC), Canada, G1V 0A6;

\*\* Applied Geotechnologies Research Group, Centro de Investigación en Tecnologías, Energía e Procesos Industriales (CINTECX), Universidade de Vigo, Campus Universitario Lagoas-Marcosende, 36310 Vigo, Spain, [ivgarrido@uvigo.es](mailto:ivgarrido@uvigo.es);

\*\*\* Escuela de Ingeniería Eléctrica y Electrónica, Universidad del Valle, Cali, VA, 760032, Colombia

### Abstract

Infrared thermography is used for evaluating composite materials due to the properties of low cost, fast inspection of large surfaces. The application of deep neural networks tends to be a prominent direction in the Infrared Non-Destructive Testing. During the training of the neural network, the Achilles heel is the database. The collection of huge amounts of training data is the high expense task. In Non-Destructive Testing with deep learning, the synthetic data contributing to training in infrared thermography remains unexplored. In this paper, synthetic data from the standard Finite Element Models is combined with experimental data to build repositories with Mask-RCNN to achieve defect segmentation.

**Keywords:** Automatic Defect detection, Infrared Thermography, Deep learning algorithms, Non-destructive Evaluation (NDE), Supervised learning, Image processing, Data augmentation; Finite Element Models (FEM)

### 1. Introduction

Quality evaluation is playing a fundamental role for modern industrial production and manufacturing processing. The demand for inspection of materials with respect to the possible presence of defects, damage and flaws has increased due to the wide use of composite materials and metals in industry. The method used to inspect defects via manual, visual evaluation involving humans can be hampered by the human fatigue. In order to meet the need of high-quality production and maintain the stringent high-quality level, an advanced visual inspection system is becoming more and more essential for structural health monitoring and production lines application. Automated quality control [1] can be applied in the industrial field to facilitate the consistent and efficient inspection. The high inspection rate and the inspection procedure without human involvement are the main advantages of automated inspection system [2].

Non-Destructive Testing (NDT) [3] encompasses a group of inspection techniques that are intended to assess the integrity of an object's interest without causing any kind of modification or permanent damage. Infrared Testing (IRT) is an NDT technique in which the differences in thermo-physical properties between the inspected object material and possible surface and/or subsurface defects are exploited through the use of an infrared camera. However, automatically identifying defects in materials via infrared thermography still remains a fresh and challenging task in the image processing domains.

In this research, an automatic instance segmentation and identification algorithm in deep learning (Mask-RCNN) [4] is introduced for defect segmentation in an automatic infrared detection system. The defect detection procedure can be regarded as either an object detection task [5] or object segmentation task [6]. In the object detection task, the objective is to fit the bounding box localized around each defect in the image. In the object segmentation task, the objective is based on pixel level classification to distinguish each pixel if it is detected. Compared with the earlier segmentation strategies [7-8], the Instance segmentation associated each pixel of an image with an instance label [9]. It can predict a whole segmentation mask for each of those objects and predict which pixel in the input image corresponds to each object instance. It also reduces the restriction concerning the position of defects rather than predicting a group of bounding boxes for the defects.

Meanwhile, like any other learning approach, deep learning's Achilles heel is the training dataset. It must be accurate enough and contain enough images and exceptions (e.g. occlusions) so as to allow the algorithm to learn reliable features of interest. If the dataset does not contain enough images and cases of possible situations or if the data contains too much noise or artifacts, the algorithm may not be accurate enough or even may learn to identify the wrong features. In recent years, the exploitation of synthetic data [10] during the training and validation of deep neural networks has emerged as a popular topic. In this work, synthetic data generated with Finite Element Models (FEM) is used during the training process to greatly reduce the high expense from real experiments in infrared thermography. The synthetic data generated



License: <https://creativecommons.org/licenses/by/4.0/deed.en>

## AUTOMATIC SEGMENTATION AND IDENTIFICATION OF DEFECTS BY DEEP LEARNING ALGORITHMS FROM PULSED THERMOGRAPHY DATA

Qiang FANG<sup>1\*</sup>, Clemente. Ibarra. CASTANEDO <sup>2</sup>, Xavier. MALDAGUE <sup>3</sup>

<sup>1</sup>Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Université Laval, 1065, av. de la Médecine, Québec (QC), Canada, G1V 0A6, qiang.fang.1@ulaval.ca

<sup>2</sup>Computer Vision and Systems Laboratory Electrical and Computing Engineering Department, Pavillon Adrien Pouliot, Laval University, Québec (PQ) Canada, G1K 7P4, clemente.ibarra-castanedo@gel.ulaval.ca

<sup>3</sup>Computer Vision and Systems Laboratory, Department of Electrical Engineering and Computer Engineering, Faculty of Science and Engineering, Université Laval, Québec (QC), Canada, Xavier.Maldague@gel.ulaval.ca

\*Contact: qiang.fang.1@ulaval.ca

### Abstract

Infrared thermography (IRT), is one of the most interesting techniques to identify different kinds of defects such as delamination and damage existing in materials. Raw thermal signatures, however, are usually noisy and require further processing to improve contrast and improve defect detection. Image segmentation algorithms have been widely applied in the image processing field, although very rarely in the IRT field. In this paper, the spatial deep learning algorithms for defects segmentation and identification are investigated. The obtained results are evaluated with Probability of Detection (POD) from deep learning structures such as Mask Region based Convolutional Neural Networks (Mask-RCNN), U-net and Res-U-net etc. The application of the state-of-the-art deep Convolutional Neural Networks for automating detection and identification of defects by instance and semantic segmentation are presented using thermal data from Pulsed Thermography (PT). A series of academic samples composed of different materials and containing artificial defects of different shapes and nature (flat-bottom holes, Teflon inserts) have been tested by PT and the results were studied in order to evaluate the efficacy and performance of the proposed algorithms.

**Keywords:** Deep learning NDE, Automatic defect segmentation, Infrared Thermography, Adapted visible image processing.

### 1. Introduction

Non-destructive Evaluation (NDE) techniques [1] is a group of techniques used in industry for analyzing and evaluating the properties of a material without causing damage. NDE methods rely upon use of electromagnetic radiation, sound and other signal conversions to examine a wide variety of articles. Infrared inspection techniques have been applied frequently to evaluate subsurface defects and hidden structures etc. for the quality control of materials such as metals, composites and so on. Due to the analysis of thermal front propagation, each thermal non-homogeneity perturbs the thermal waves propagation on the surface of specimen in comparison to the surrounding sound without defect region. We can then see the changes of the temperature variation. The infrared camera and corresponding equipment can record this thermal perturbation, and the results can be analyzed to obtain further information [2].

Since the beginning of the twenty-first century, deep learning (deep neural networks), a field of artificial intelligence, has made remarkable progress. Initially developed to make artificial neural networks more efficient, deep learning, neural networks have shown their ability to outdo most of the other approaches existing previously in a great number of applications, which has contributed to making them very popular in many scientific communities. Deep learning is a field of artificial intelligence. Initially developed to make the neural network more efficient,

## Appendix B Programming codes and Implementations

This appendix summarized the PYTHON and MATLAB codes which used for implementation of the proposed algorithms. The documentation of these codes is based on presented chapters and articles. For partial chapters the relevant code mentioned a long with a link corresponds to the code that can be downloaded online.

### B.1 Automatic defect detection from deep learning models (Chapter4)

#### B.1.1 YOLO-V3 for defect localization

```
importtkinter as tk
fromtkinter import *
import datetime
fromtkinter.filedialog import askdirectory, askopenfilename
fromtkinter.messagebox import showinfo
import cv2
fromPIL import Image, ImageTk
import threading
import numpy as np
import os
import matplotlib.pyplot as plt
```

Indication of the predication  
results on the screen

```
class goods_counting_gui:
    def __init__(self):
        super().__init__()
        self.create_gui()
    def create_gui(self):
        self.window = tk.Tk()
        self.window.title(' Defect detection and identification system ')
        self.label = tk.Label(self.window, text='Current time: ', bg='green', font=30)
        self.label.grid(row=0, column=0)
        self.cur_time = tk.Label(self.window,
text='%s%d'%(datetime.datetime.now().strftime('%Y-%m-%d %H:%M:%S'),
datetime.datetime.now().microsecond//100000), font=30)
        self.cur_time.grid(row=0, column=1)
        self.window.after(100, self.update_time)

        self.img_path = StringVar()
        self.label_yolo_obj = tk.Label(self.window, text='choose image: ')
        self.label_yolo_obj.grid(row=1, column=0, pady=6)
        self.entry1 = tk.Entry(self.window, textvariable=self.img_path, width=50)
```

```

self.entry1.grid(row=1, column=1, padx=5, pady=6)

self.button_yolo_obj = tk.Button(self.window, text='path of image', bg='pink', relief=tk.RAISED,
width=14, height=1,
                                command=self.get_img_path1)
self.button_yolo_obj.grid(row=1, column=2, padx=10, pady=6)

self.multi_image = StringVar()
self.multi_image_label = tk.Label(self.window, text='choose directory: ')
self.multi_image_label.grid(row=2, column=0, pady=6)
self.entry2 = tk.Entry(self.window, textvariable=self.multi_image, width=50)
self.entry2.grid(row=2, column=1, padx=5, pady=6)

self.multi_image_button = tk.Button(self.window, text='path of image', bg='pink', relief=tk.RAISED,
width=14,
                                height=1, command=self.get_img_path2)
self.multi_image_button.grid(row=2, column=2, padx=10, pady=6)

self.button_start = tk.Button(self.window, text='start recognition', bg='orange', relief=tk.RAISED,
                                width=13, height=1, command=self.start)
self.button_start.grid(row=1, column=3, padx=10, pady=6)

self.button_get_result = tk.Button(self.window, text='start recognition', bg='orange',
relief=tk.RAISED,
                                width=13, height=1, command=self.test_all)
self.button_get_result.grid(row=2, column=3, padx=10, pady=6)

self.button_start = tk.Button(self.window, text='clear output', bg='red', relief=tk.RAISED, width=13,
height=1,
                                command=self.stop)
self.button_start.grid(row=3, column=2, padx=10, pady=10)

self.frm_ = tk.Frame()
self.frm_.grid(row=3, column=1, padx=5)

self.result_show = tk.Label(self.frm_)
self.result_show.grid(row=4, column=1, padx=5)

# self.output = tk.Label(self.window, text='output display', font=50)
# self.output.grid(row=3, column=2, padx=5, pady=5, sticky=tk.NW)

self.text = tk.Text(self.window, width=24, height=10)
self.text.grid(row=3, column=3, pady=5, sticky=tk.NW)

def get_img_path1(self):

```

Define the type of images:  
PIL. Image.

```

# self.path1 = askdirectory()
self.path1 = askopenfilename()
self.img_path.set(self.path1)
frame = cv2.imread(self.path1)
B, G, R = cv2.split(frame)
frame = cv2.merge([R, G, B])
img = Image.fromarray(frame)
img = self.resize(img)
imgtk = ImageTk.PhotoImage(img)
self.result_show.config(image=imgtk)
self.result_show.img = imgtk

```

1. Classification of objective;  
1. Define of Colors

```

def get_img_path2(self):
    self.path2 = askdirectory()
    self.multi_image.set(self.path2)

```

To obtain and connect the output later from YOLO

```

def test_all(self):
    images = os.listdir(self.path2)
    f = open(self.path2 + '/' + './location.txt', 'r+')
    for img in images:
        weightsPath = "material.weights"
        configPath = "material.cfg"
        labelsPath = "material.names"
        self.LABELS = open(labelsPath).read().strip().split("\n")
        COLORS = np.random.randint(0, 255, size=(len(self.LABELS), 3), dtype="uint8")
        boxes = []
        confidences = []
        self.classIDs = []
        net = cv2.dnn.readNetFromDarknet(configPath, weightsPath)
        if not os.path.exists(self.path2 + '/' + './result'):
            os.makedirs(self.path2 + '/' + './result')

        image = cv2.imread(self.path2 + '/' + img)
        # print(self.path1)
        (H, W) = image.shape[:2]

```

giving us the bounding box and associated probabilities with the loaded model

```

ln = net.getLayerNames()
ln = [ln[i] - 1 for i in net.getUnconnectedOutLayers()]

blob = cv2.dnn.blobFromImage(image, 1 / 255.0, (416, 416), swapRB=True, crop=False)
net.setInput(blob)
layerOutputs = net.forward(ln)

```



```

for output in layerOutputs:

    for detection in output:
        scores = detection[5:]
        classID = np.argmax(scores)
        confidence = scores[classID]

        if confidence > 0.5:

            bdx = detection[0:4] * np.array([W, H, W, H])
                (centerX, centerY, width, height) = box.astype("int")

                x = int(centerX - (width / 2))
                y = int(centerY - (height / 2))
                boxes.append([x, y, int(width), int(height)])

                confidences.append(float(confidence))
self.classIDs.append(classID)
idxs = cv2.dnn.NMSBoxes(boxes, confidences, 0.2, 0.3)

```

Non-Maximum  
Suppression

## B.1.2 Res-U-net for defect semantic segmentation

```

import time
import torch
import torch.optim as optim
import torch.backends.cudnn as cudnn
import torchvision.models as models
import numpy as np
from tqdm import tqdm
from torchvision import models
from torch.autograd import Variable
from PIL import Image
from torch import nn
from nets.unt import Unet
from nets.unet_training import CE_Loss, Dice_loss
from utils.metrics import f_score
from torch.utils.data import DataLoader
from dataloader import unetDataset, unet_dataset_collate

def get_lr(optimizer):
    for param_group in optimizer.param_groups:
        return param_group['lr']

```

```

def fit_one_epoch(net,epoch,epoch_size,epoch_size_val,gen,genval,Epoch,cuda,aux_branch):
    net = net.train()
    total_loss = 0
    total_f_score = 0

    val_toal_loss = 0
    val_total_f_score = 0
    start_time = time.time()
    with tqdm(total=epoch_size,desc=f'Epoch {epoch + 1}/{Epoch}',postfix=dict,mininterval=0.3) as pbar:
        for iteration, batch in enumerate(gen):
            if iteration >= epoch_size:
                break
            imgs, pngs, labels = batch

            with torch.no_grad():
                imgs = Variable(torch.from_numpy(imgs).type(torch.FloatTensor))
                pngs = Variable(torch.from_numpy(pngs).type(torch.FloatTensor)).long()
                labels = Variable(torch.from_numpy(labels).type(torch.FloatTensor))
                if cuda:
                    imgs = imgs.cuda()
                    pngs = pngs.cuda()
                    labels = labels.cuda()

            optimizer.zero_grad()
            if aux_branch:
                aux_outputs, outputs = net(imgs)
                aux_loss = CE_Loss(aux_outputs, pngs, num_classes = NUM_CLASSES)

                main_loss = CE_Loss(outputs, pngs, num_classes = NUM_CLASSES)
                loss = aux_loss + main_loss
                if dice_loss:
                    aux_dice = Dice_loss(aux_outputs, labels)
                    main_dice = Dice_loss(outputs, labels)
                    loss = loss + aux_dice + main_dice

            else:
                outputs = net(imgs)
                loss = CE_Loss(outputs, pngs, num_classes = NUM_CLASSES)
                if dice_loss:
                    main_dice = Dice_loss(outputs, labels)
                    loss = loss + main_dice

            with torch.no_grad():
                f_score = f_score(outputs, labels)

```

Determine if a secondary branch is used and pass it back

Calculation F-score

```

loss.backward()
optimizer.step()

total_loss += loss.item()
total_f_score += _f_score.item()

waste_time = time.time() - start_time
pbar.set_postfix(**{'total_loss': total_loss / (iteration + 1),
                    'f_score' : total_f_score / (iteration + 1),
                    's/step' : waste_time,
                    'lr'      : get_lr(optimizer)})
pbar.update(1)

start_time = time.time()

print('Start Validation')
with tqdm(total=epoch_size_val, desc=f'Epoch {epoch + 1}/{Epoch}', postfix=dict, mininterval=0.3) as pbar:
    for iteration, batch in enumerate(genval):
        if iteration >= epoch_size_val:
            break
        imgs, pngs, labels = batch
        with torch.no_grad():
            imgs = Variable(torch.from_numpy(imgs).type(torch.FloatTensor))
            pngs = Variable(torch.from_numpy(pngs).type(torch.FloatTensor)).long()
            labels = Variable(torch.from_numpy(labels).type(torch.FloatTensor))
            if cuda:
                imgs = imgs.cuda()
                pngs = pngs.cuda()
                labels = labels.cuda()

            if aux_branch:
                aux_outputs, outputs = net(imgs)

                aux_loss = CE_Loss(aux_outputs, pngs, num_classes = NUM_CLASSES)
                main_loss = CE_Loss(outputs, pngs, num_classes = NUM_CLASSES)
                val_loss = aux_loss + main_loss
                if dice_loss:
                    aux_dice = Dice_loss(aux_outputs, labels)
                    main_dice = Dice_loss(outputs, labels)
                    val_loss = val_loss + aux_dice + main_dice

            else:
                outputs = net(imgs)

```

```

    val_loss = CE_Loss(outputs, pngs, num_classes = NUM_CLASSES)
    if dice_loss:
        main_dice = Dice_loss(outputs, labels)
        val_loss = val_loss + main_dice

    _f_score = f_score(outputs, labels)

    val_toal_loss += val_loss.item()
    val_total_f_score += _f_score.item()

pbar.set_postfix(**{'total_loss': val_toal_loss / (iteration + 1),
                    'f_score' : val_total_f_score / (iteration + 1),
                    'lr'      : get_lr(optimizer)})
pbar.update(1)

print('Finish Validation')
print('Epoch:' + str(epoch+1) + '/' + str(Epoch))
print('Total Loss: %.4f || Val Loss: %.4f ' %
(total_loss/(epoch_size+1),val_toal_loss/(epoch_size_val+1)))
totalBig_loss = ('%.4f' % (total_loss/(epoch_size+1)))
val_loss1232= ('%.4f' % (val_toal_loss/(epoch_size_val+1)))
score = ('%.4f' % (val_total_f_score / (iteration + 1)))
file_handle2=open('train_loss.csv',mode='a+')

file_handle2.write(totalBig_loss+','+val_loss1232+'\n')
file_handle2.close()
file_handle3=open('score.csv',mode='a+')

file_handle3.write(score+'\n')
file_handle3.close()
print('Saving state, iter:', str(epoch+1))
torch.save(model.state_dict(), 'logs/Epoch%d-Total_Loss%.4f
Val_Loss%.4f.pth'%((epoch+1),total_loss/(epoch_size+1),val_toal_loss/(epoch_size_val+1)))
if __name__ == "__main__":
    inputs_size = [256,256,3]

    log_dir = "logs/"

```

Calculation of F-scores

## B.2 Mask-RCNN for defect instance segmentation in infrared thermography with synthetic data (Chapter 5)

```
import os
import random
import datetime
import re
import math
import logging
from collections import OrderedDict
import multiprocessing
import numpy as np
import tensorflow as tf
import keras
import keras.backend as K
import keras.layers as KL
import keras.engine as KE
import keras.models as KM
from keras import metrics

from mrcnn import utils

# Requires TensorFlow 1.3+ and Keras 2.0.8+.
from distutils.version import LooseVersion
assert LooseVersion(tf.__version__) >= LooseVersion("1.3")
assert LooseVersion(keras.__version__) >= LooseVersion('2.0.8')

#####
# Utility Functions
#####

def log(text, array=None):
    """Prints a text message. And, optionally, if a Numpy array is provided it
    prints it's shape, min, and max values.
    """
    if array is not None:
        text = text.ljust(25)
        text += ("shape: {:20} ".format(str(array.shape)))
        if array.size:
            text += ("min: {:10.5f} max: {:10.5f}".format(array.min(),array.max()))
        else:
            text += ("min: {:10} max: {:10}".format("", ""))
        text += " {}".format(array.dtype)
    print(text)
```

```

class BatchNorm(KL.BatchNormalization):
    def call(self, inputs, training=None):
        return super(self.__class__, self).call(inputs, training=training)

```

```

def compute_backbone_shapes(config, image_shape):
    """Computes the width and height of each stage of the backbone network.

```

Returns:  
 [N, (height, width)]. Where N is the number of stages  
 """

```

if callable(config.BACKBONE):
    return config.COMPUTE_BACKBONE_SHAPE(image_shape)

```

```

# Currently supports ResNet only
assert config.BACKBONE in ["resnet50", "resnet101"]
return np.array(
    [[int(math.ceil(image_shape[0] / stride)),
      int(math.ceil(image_shape[1] / stride))]
     for stride in config.BACKBONE_STRIDES])

```

```

#####
# Resnet Graph
#####

```

```

# Code adopted from:
# https://github.com/fchollet/deep-learning-models/blob/master/resnet50.py

```

```

def identity_block(input_tensor, kernel_size, filters, stage, block,
                  use_bias=True, train_bn=True):
    """The identity_block is the block that has no conv layer at shortcut
    # Arguments
        input_tensor: input tensor
        kernel_size: default 3, the kernel size of middle conv layer at main path

    conv_name_base = 'res' + str(stage) + block + '_branch'
    bn_name_base = 'bn' + str(stage) + block + '_branch'

    x = KL.Conv2D(nb_filter1, (1, 1), strides=strides,
                  name=conv_name_base + '2a', use_bias=use_bias)(input_tensor)
    x = BatchNorm(name=bn_name_base + '2a')(x, training=train_bn)
    x = KL.Activation('relu')(x)

```

Batch normalization has a negative effect on training if batches are small so this layer is often frozen (via setting in Config class) and functions as linear layer.

```

x = KL.Conv2D(nb_filter2, (kernel_size, kernel_size), padding='same',
              name=conv_name_base + '2b', use_bias=use_bias)(x)
x = BatchNorm(name=bn_name_base + '2b')(x, training=train_bn)
x = KL.Activation('relu')(x)

x = KL.Conv2D(nb_filter3, (1, 1), name=conv_name_base +
              '2c', use_bias=use_bias)(x)
x = BatchNorm(name=bn_name_base + '2c')(x, training=train_bn)

shortcut = KL.Conv2D(nb_filter3, (1, 1), strides=strides,
                    name=conv_name_base + '1', use_bias=use_bias)(input_tensor)
shortcut = BatchNorm(name=bn_name_base + '1')(shortcut, training=train_bn)

```

```

x = KL.Add()([x, shortcut])
x = KL.Activation('relu', name='res' + str(stage) + block + '_out')(x)
return x

def resnet_graph(input_image, architecture, stage5=False, train_bn=True):

    assert architecture in ["resnet50", "resnet101"]
    # Stage 1
    x = KL.ZeroPadding2D((3, 3))(input_image)
    x = KL.Conv2D(64, (7, 7), strides=(2, 2), name='conv1', use_bias=True)(x)
    x = BatchNorm(name='bn_conv1')(x, training=train_bn)
    x = KL.Activation('relu')(x)
    C1 = x = KL.MaxPooling2D((3, 3), strides=(2, 2), padding="same")(x)
    # Stage 2
    x = conv_block(x, 3, [64, 64, 256], stage=2, block='a', strides=(1, 1), train_bn=train_bn)
    x = identity_block(x, 3, [64, 64, 256], stage=2, block='b', train_bn=train_bn)

```

Build a ResNet graph architecture: Can be resnet50 or resnet101

stage: Boolean. If False, stage of the network is not created

## B.3. PLS-GANs algorithms for defect enhancement (chapter 6)

This section shows the Python and MATLAB codes correspond to applying GANs network (DCGAN; WGAN) and Partial least square thermography to build as the PLS-GANs network.

In order to apply these techniques, an OpenCV image analysis toolbox and MATLAB (R2019a) were used.

### B.3.1 Deep convolutional GANs

```
from __future__ import print_function
import argparse
import os
import random
import torch
import torch.nn as nn
import torch.nn.parallel
import torch.backends.cudnn as cudnn
import torch.optim as optim
import torch.utils.data
import torchvision.datasets as dset
import torchvision.transforms as transforms
import torchvision.utils as vutils
from skimage import io
```

DCGAN  
parameter  
configuration that  
can generate  
64×64 images

```
parser = argparse.ArgumentParser()
parser.add_argument('--dataset', default='mydata', help='cifar10 | lsun | imagenet | folder | lfw | fake')
parser.add_argument('--dataroot', default='data/', help='path to dataset')
parser.add_argument('--workers', type=int, help='number of data loading workers', default=2)
parser.add_argument('--batchSize', type=int, default=64, help='input batch size')
parser.add_argument('--imageSize', type=int, default=64, help='the height / width of the input image to network')
parser.add_argument('--nz', type=int, default=100, help='size of the latent z vector')
parser.add_argument('--ngf', type=int, default=64)
parser.add_argument('--ndf', type=int, default=64)
parser.add_argument('--niter', type=int, default=1000, help='number of epochs to train for')
parser.add_argument('--lr', type=float, default=0.002, help='learning rate, default=0.0002')
parser.add_argument('--beta1', type=float, default=0.5, help='beta1 for adam. default=0.5')
parser.add_argument('--cuda', default='True', action='store_true', help='enables cuda')
parser.add_argument('--ngpu', type=int, default=1, help='number of GPUs to use')
parser.add_argument('--netG', default='', help='path to netG (to continue training)')
```



```

parser.add_argument('--netD', default='', help="path to netD (to continue training)")
parser.add_argument('--outf', default='imgs/', help='folder to output images and model checkpoints')
parser.add_argument('--manualSeed', type=int, help='manual seed')

```

```

opt = parser.parse_args()
print(opt)

```

```

try:
    os.makedirs(opt.outf)
except OSError:
    pass

```

```

if opt.manualSeed is None:
    opt.manualSeed = random.randint(1, 10000)
print("Random Seed: ", opt.manualSeed)
random.seed(opt.manualSeed)
torch.manual_seed(opt.manualSeed)

```

```

cudnn.benchmark = True

```

```

if torch.cuda.is_available() and not opt.cuda:
    print("WARNING: You have a CUDA device, so you should probably run with --cuda")

```

```

if opt.dataset in ['imagenet', 'folder', 'lfw']:

```

```

    # folder dataset

```

```

    dataset = dset.ImageFolder(root=opt.dataroot,
                              transform=transforms.Compose([
                                  transforms.Resize(opt.imageSize),
                                  transforms.CenterCrop(opt.imageSize),
                                  transforms.ToTensor(),
                                  transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5))
                              ]))

```

Normalize grayscale images

```

elif opt.dataset == 'lsun':

```

```

    dataset = dset.LSUN(root=opt.dataroot, classes=['bedroom_train'],
                       transform=transforms.Compose([
                           transforms.Resize(opt.imageSize),
                           transforms.CenterCrop(opt.imageSize),
                           transforms.ToTensor(),
                           transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5)),
                       ]))

```

```

elif opt.dataset == 'cifar10':

```

```

    dataset = dset.CIFAR10(root=opt.dataroot, download=True,
                          transform=transforms.Compose([
                              transforms.Resize(opt.imageSize),
                              transforms.ToTensor(),

```

```

        transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5)),
    ])
elif opt.dataset == 'fake':
    dataset = dset.FakeData(image_size=(3, opt.imageSize, opt.imageSize),
                           transform=transforms.ToTensor())
elif opt.dataset == 'mydata':
    transforms = torchvision.transforms.Compose([
        torchvision.transforms.ToPILImage(),
        torchvision.transforms.Resize((64, 64)),
        transforms.ToTensor(),
        transforms.Normalize([0.5], [0.5])])

    dataset = torchvision.datasets.ImageFolder(opt.dataroot, transform=transforms, loader=io.imread)

assert dataset

dataloader = torch.utils.data.DataLoader(dataset, batch_size=opt.batchSize,
                                         shuffle=True, num_workers=int(opt.workers))

device = torch.device("cuda:0" if opt.cuda else "cpu")
ngpu = int(opt.ngpu)
nz = int(opt.nz)
ngf = int(opt.ngf)
ndf = int(opt.ndf)
nc = 1

# custom weights initialization called on netG and netD
def weights_init(m):
    classname = m.__class__.__name__
    if classname.find('Conv') != -1:
        m.weight.data.normal_(0.0, 0.02)
    elif classname.find('BatchNorm') != -1:
        m.weight.data.normal_(1.0, 0.02)
        m.bias.data.fill_(0)

class Generator(nn.Module):
    def __init__(self, ngpu):
        super(Generator, self).__init__()
        self.ngpu = ngpu
        self.main = nn.Sequential(
            # input is Z, going into a convolution
            nn.ConvTranspose2d(nz, ngf * 8, 4, 1, 0, bias=False),
            nn.BatchNorm2d(ngf * 8),
            nn.ReLU(True),
            # state size. (ngf*8) x 4 x 4

```

```

nn.ConvTranspose2d(ngf * 8, ngf * 4, 4, 2, 1, bias=False),
nn.BatchNorm2d(ngf * 4),
nn.ReLU(True),
# state size. (ngf*4) x 8 x 8
nn.ConvTranspose2d(ngf * 4, ngf * 2, 4, 2, 1, bias=False),
nn.BatchNorm2d(ngf * 2),
nn.ReLU(True),
# state size. (ngf*2) x 16 x 16
nn.ConvTranspose2d(ngf * 2, ngf, 4, 2, 1, bias=False),
nn.BatchNorm2d(ngf),
nn.ReLU(True),
# state size. (ngf) x 32 x 32
nn.ConvTranspose2d(ngf, nc, 4, 2, 1, bias=False),
nn.Tanh()
# state size. (nc) x 64 x 64
)

```

```

def forward(self, input):
    if input.is_cuda and self.ngpu > 1:
        output = nn.parallel.data_parallel(self.main, input, range(self.ngpu))
    else:
        output = self.main(input)
    return output

```

```

netG = Generator(ngpu).to(device)
netG.apply(weights_init)

```

```

if opt.netG != "":
    netG.load_state_dict(torch.load(opt.netG))
print(netG)

```

```

class Discriminator(nn.Module):
    def __init__(self, ngpu):
        super(Discriminator, self).__init__()
        self.ngpu = ngpu
        self.main = nn.Sequential(
            # input is (nc) x 64 x 64
            nn.Conv2d(nc, ndf, 4, 2, 1, bias=False),
            nn.LeakyReLU(0.2, inplace=True),
            # state size. (ndf) x 32 x 32
            nn.Conv2d(ndf, ndf * 2, 4, 2, 1, bias=False),
            nn.BatchNorm2d(ndf * 2),
            nn.LeakyReLU(0.2, inplace=True),
            # state size. (ndf*2) x 16 x 16
            nn.Conv2d(ndf * 2, ndf * 4, 4, 2, 1, bias=False),

```

```

nn.BatchNorm2d(ndf*4),
nn.LeakyReLU(0.2, inplace=True),
# state size. (ndf*4) x 8 x 8
nn.Conv2d(ndf * 4, ndf * 8, 4, 2, 1, bias=False),
nn.BatchNorm2d(ndf * 8),
nn.LeakyReLU(0.2, inplace=True),
# state size. (ndf*8) x 4 x 4
nn.Conv2d(ndf * 8, 1, 4, 1, 0, bias=False),
nn.Sigmoid()
)

```

Update D network:  
maximize  $\log(D(x)) + \log(1 - D(G(z)))$

```

def forward(self, input):
    if input.is_cuda and self.ngpu > 1:
        output = nn.parallel.data_parallel(self.main, input, range(self.ngpu))
    else:
        output = self.main(input)

    return output.view(-1, 1).squeeze(1)

```

```

netD = Discriminator(ngpu).to(device)
netD.apply(weights_init)
if opt.netD != "":
    netD.load_state_dict(torch.load(opt.netD))
print(netD)

```

```

criterion = nn.BCELoss()

```

```

fixed_noise = torch.randn(opt.batchSize, nz, 1, 1, device=device)
real_label = 1
fake_label = 0

```

```

# setup optimizer
optimizerD = optim.Adam(netD.parameters(), lr=opt.lr, betas=(opt.beta1, 0.999))
optimizerG = optim.Adam(netG.parameters(), lr=opt.lr, betas=(opt.beta1, 0.999))

```

```

for epoch in range(opt.niter):
    for i, data in enumerate(dataloader, 0):

        netD.zero_grad()
        real_cpu = data[0].to(device)
        batch_size = real_cpu.size(0)
        label = torch.full((batch_size,), real_label, device=device)

        output = netD(real_cpu)

```

```

errD_real = criterion(output, label)
errD_real.backward()
D_x = output.mean().item()

# train with fake
noise = torch.randn(batch_size, nz, 1, 1, device=device)
fake = netG(noise)
label.fill_(fake_label)
output = netD(fake.detach())
errD_fake = criterion(output, label)
errD_fake.backward()
D_G_z1 = output.mean().item()
errD = errD_real + errD_fake
optimizerD.step()
netG.zero_grad()
label.fill_(real_label) # fake labels are real for generator cost
output = netD(fake)
errG = criterion(output, label)
errG.backward()
D_G_z2 = output.mean().item()
optimizerG.step()

print('[%d/%d][%d/%d] Loss_D: %.4f Loss_G: %.4f D(x): %.4f D(G(z)): %.4f / %.4f'
      % (epoch, opt.niter, i, len(dataloader),
         errD.item(), errG.item(), D_x, D_G_z1, D_G_z2))
if i % 100 == 0:
    vutils.save_image(real_cpu,
                      '%s/real_samples.png' % opt.outf,
                      normalize=True)
    fake = netG(fixed_noise)
    vutils.save_image(fake.detach(),
                      '%s/fake_samples_epoch_%03d.png' % (opt.outf, epoch),
                      normalize=True)
torch.save(netG.state_dict(), '%s/netG_epoch_%d.pth' % (opt.outf, epoch))
torch.save(netD.state_dict(), '%s/netD_epoch_%d.pth' % (opt.outf, epoch))

```

Update G network: maximize  $\log(D(G(z)))$

do checkpointing

### B.3.2 Wasserstein GAN

```
import torch.nn as nn
```

```
class NetG(nn.Module):  
    def __init__(self, ngf, nz):  
        super(NetG, self).__init__()  
  
        self.layer1 = nn.Sequential(  
            nn.ConvTranspose2d(nz, ngf * 8, kernel_size=3, stride=2, padding=1, bias=False),  
            nn.BatchNorm2d(ngf * 8),  
            nn.ReLU(inplace=True)  
        )  
  
        self.layer2 = nn.Sequential(  
            nn.ConvTranspose2d(ngf * 8, ngf * 4, 3, 2, 0, bias=False),  
            nn.BatchNorm2d(ngf * 4),  
            nn.ReLU(inplace=True)  
        )  
  
        self.layer3 = nn.Sequential(  
            nn.ConvTranspose2d(ngf * 4, ngf * 2, 5, 2, 1, bias=False),  
            nn.BatchNorm2d(ngf * 2),  
            nn.ReLU(inplace=True)  
        )  
  
        self.layer4 = nn.Sequential(  
            nn.ConvTranspose2d(ngf * 2, ngf, 3, 3, 1, bias=False),  
            nn.BatchNorm2d(ngf),  
            nn.ReLU(inplace=True)  
        )
```

Define the Generator network G and the parameters on each layer of G; randomly noise input

```
self.layer5 = nn.Sequential(  
    nn.ConvTranspose2d(ngf, 3, 4, 2, 0, bias=False),  
    nn.Tanh()  
)
```

```
def forward(self, x):  
    out = self.layer1(x)  
    out = self.layer2(out)  
    out = self.layer3(out)  
    out = self.layer4(out)  
    out = self.layer5(out)  
    return out
```

Defining the forward propagation of Discriminator D

```

class NetD(nn.Module):
    def __init__(self, ndf):
        super(NetD, self).__init__()

        self.layer1 = nn.Sequential(
            nn.Conv2d(3, ndf, kernel_size=3, stride=3, padding=1, bias=False),
            nn.BatchNorm2d(ndf),
            nn.LeakyReLU(0.2, inplace=True)
        )

        self.layer2 = nn.Sequential(
            nn.Conv2d(ndf, ndf * 2, 3, 3, 0, bias=False),
            nn.BatchNorm2d(ndf * 2),
            nn.LeakyReLU(0.2, inplace=True)
        )

        self.layer3 = nn.Sequential(
            nn.Conv2d(ndf * 2, ndf * 4, 3, 3, 0, bias=False),
            nn.BatchNorm2d(ndf * 4),
            nn.LeakyReLU(0.2, inplace=True)
        )

        self.layer4 = nn.Sequential(
            nn.Conv2d(ndf * 4, ndf * 8, 5, 3, 0, bias=False),
            nn.BatchNorm2d(ndf * 8),
            nn.LeakyReLU(0.2, inplace=True)
        )

        self.layer5 = nn.Sequential(
            nn.Conv2d(ndf * 8, 1, 4, 1, 0, bias=False),
            nn.Sigmoid()
        )

```

Define the discriminator network D and the parameters on each layer of D

```

def forward(self, x):
    out = self.layer1(x)
    out = self.layer2(out)
    out = self.layer3(out)
    out = self.layer4(out)
    out = self.layer5(out)
    return out

```

Defining the forward propagation of Discriminator D

### B.3.3 Partial least square thermography

```
% Sampling frequency
```

```
fs = 88;
```

```
ts = (1/fs);
```

```
ncomp=10;
```

```
[a,b,t] = size(T);
```

```
X=reshape(T,(a*b),t);
```

```
X=X';
```

```
[n,m] = size(X);
```

```
time=ts:ts:(ts*t);
```

```
Y=time';
```

```
[XL,YL,XS,YS,BETA,PLSPctVar,PLSmsep,stats] = plsregress(X,Y,ncomp);
```

```
yfitPLS = [ones(n,1) X]*BETA;
```

```
%% Analysis
```

```
% 3D Loading Matrix
```

```
XLoading=reshape(XL,a,b,ncomp);
```

```
% 3D Weight Matrix
```

```
W=reshape(stats.W,a,b,ncomp);
```

```
% Decomposed Temperature Matrix
```

```
Xnew= XL*XS';
```

```
XfitPLS=reshape(Xnew,a,b,t);
```

Loading,  
transform and  
decomposed the  
temperature  
matrix;

```
%% ThermalMap Comp. 1
```

```
Xnew1=XL(:,1)*XS(:,1)';
```

```
% X_reg1=reshape(Xnew1,a,b,t);
```

```
%% ThermalMap Comp. 2
```

```
Xnew2=XL(:,2)*XS(:,2)';
```

```
X_reg2=reshape(Xnew2,a,b,t);
```

```
%% ThermalMap Comp. 3
```

```
Xnew3=XL(:,3)*XS(:,3)';
```

```
X_reg3=reshape(Xnew3,a,b,t);
```

```
%% ThermalMap Comp. 4
```

```
Xnew4=XL(:,4)*XS(:,4)';
```

```
X_reg4=reshape(Xnew4,a,b,t);
```

```
%% ThermalMap Comp. 5
```

```
ThermalMap Comp. 5
```

```
Xnew5=XL(:,5)*XS(:,5)';
```

```
X_reg5=reshape(Xnew5,a,b,t);
```



```

% % ThermalMap Comp. 6
ThermalMap Comp. 6
Xnew6=XL(:,6)*XS(:,6)';
X_reg6=reshape(Xnew6,a,b,t);

```

Cross-Validation

```

plot(1:ncomp,cumsum(100*PLSPctVar(2,:)),'-bo');
xlabel('Number of PLS components');
ylabel('Percent Variance Cumulative in X');

```

```
hold on;
```

```

plot(0:ncomp,PLSmsep(2:),'b-o');
xlabel('Number of components');
ylabel('Estimated Mean Squared Prediction Error');

```

```

for t=1:n;
    Xmean(t)=mean(X(t,:));
end

```

Selecting Group of Components  
2nd and 3rd two components  
Reconstructed Nom-centered Thermal Map

```

for t=1:n;
    T_reg(:,t)=Xnew(:,t)+Xmean(t);
end
T_reg_new=reshape(T_reg,a,b,t);

```

```

for t=1:n;
    C1(:,t)=Xnew(:,t)-Xnew1(:,t);
end

```

```
C1_new=reshape(C1,a,b,t);
```

```

for t=1:n;
    C1mean(t)=mean(C1(t,:));
end

```

```

for t=1:n;
    C1_reg(:,t)=C1(:,t)+C1mean(t);
end

```

Reconstructed  
Nom-centered  
Enhanced Images

```

plot(1:ncomp,cumsum(100*PLSPctVar(2,:)),'-bo');
xlabel('Number of PLS components');
ylabel('Percent Variance Cumulative in X');
hold on;
plot(0:ncomp,PLSmsep(2:),'b-o');
xlabel('Number of components');
ylabel('Estimated Mean Squared Prediction Error');

```

## B.4 Gated Recurrent Units (GRUs) for defect depth estimation (Chapter7)

### B.4.1 Model training

```
import numpy as np
import tensorflow as tf
from tensorflow import keras
import matplotlib.pyplot as plt

train_data = np.load('./train_data/train0.6_1.9.npy')

label1 = np.array([0.6]*20)
label2 = np.array([0.9]*20)
label3 = np.array([1.3]*19)
label4 = np.array([1.7]*20)
label5 = np.array([1.9]*20)

label = np.concatenate([label1, label2, label3, label4, label5])
label = np.reshape(label, (len(label),1))
```

```
x_train = train_data
```

```
x_train = tf.reshape(x_train, (-1, 1429, 1))
```

```
y_train = tf.reshape(label, (-1,1))
```

```
model = keras.models.Sequential([
    keras.layers.GRU(512, return_sequences=False, input_shape=[None, 1]),
    keras.layers.Dense(256, activation='relu'),
    keras.layers.Dense(64, activation='relu'),
    keras.layers.Dense(1)
])
```

```
model.compile(loss=keras.losses.mean_squared_error,
              optimizer=keras.optimizers.Adam(lr=0.001),
              metrics=['mae'])
```

```
history = model.fit(x_train, y_train,
                   epochs=2500,
                   batch_size=32,
                   verbose=2)
```

```
model.save('best0.6_1.9.h5')
```

Train the model

Retrieve data

Adjust the  
dimensionality

Building the  
GRU model

```
score_mse, score_mae = model.evaluate(x_train, y_train)
print(score_mse)
print(score_mae)
```

```
mae = history.history['mae']
loss = history.history['loss']
epochs = range(1, len(loss)+1)
plt.figure()
plt.plot(epochs, loss, 'b', label='Training loss')
plt.plot(epochs, mae, 'r', label='mae')
plt.legend()
plt.show()
```

## B.4.2 Validation and Reading the results

```
import numpy as np
import tensorflow as tf
from tensorflow import keras
import matplotlib.pyplot as plt

test_model = tf.keras.models.load_model('best0.6_1.9.h5')
#test_model.summary()

test_data = np.load('./test_data/test_data1.9.npy')

label = np.array([1.0]*20)

x_test = tf.reshape(test_data, (-1, 1429, 1))
y_test = tf.reshape(label, (-1, 1))

res = test_model.predict(x_test)
res = np.reshape(res, (-1, len(res[0])))
print("The test results are: " + str(res[:,-1]))
res1 = abs(res[:,-1]-1.9)
print("The test results of MAE: " + str(res1))
mean1 = res1.mean(axis=0)
print("The average of test results MAE: " + str(mean1))
```

Load the trained  
model to test the  
data