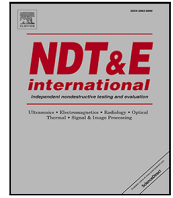




Contents lists available at ScienceDirect

NDT and E International

journal homepage: www.elsevier.com/locate/ndteint

Two-stream convolutional neural network for non-destructive subsurface defect detection via similarity comparison of lock-in thermography signals

Yanpeng Cao^{a,b}, Yafei Dong^{a,b}, Yanlong Cao^{a,b}, Jiangxin Yang^{a,b,*}, Michael Ying Yang^c

^a State Key Laboratory of Fluid Power and Mechatronic Systems, College of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China

^b Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, College of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China

^c Scene Understanding Group, University of Twente, Hengelosestraat 99, 7514 AE Enschede, The Netherlands

ARTICLE INFO

Keywords:

Non-destructive testing
Lock-in thermography
Convolutional neural network
Similarity comparison

ABSTRACT

Active infrared thermography is a safe, fast, and low-cost solution for subsurface defects inspection, providing quality control in many industrial production tasks. In this paper, we explore deep learning-based approaches to analyze lock-in thermography image sequences for non-destructive testing and evaluation (NDT&E) of subsurface defects. Different from most existing Convolutional Neural Network (CNN) models that directly classify individual regions/pixels as defective and non-defective ones, we present a novel two-stream CNN architecture to extract/compare features in a pair of 1D thermal signal sequences for accurate classification/differentiation of defective and non-defective regions. In this manner, we can significantly increase the size of the training data by pairing two individually captured 1D thermal signals, thereby greatly easing the requirement for collecting a large number of thermal sequences of specimens with defects to train deep CNN models. Moreover, we experimentally investigate a number of network alternatives, identifying the optimal fusion scheme/stage for differentiating the thermal behaviors of defective and non-defective regions. Experimental results demonstrate that our proposed method, directly learning how to construct feature representations from a large number of real-captured thermal signal pairs, outperforms the well-established lock-in thermography data processing techniques on specimens made of different materials and at various excitation frequencies.

1. Introduction

Hidden defects such as delaminations, disbonds, cracks, and subsurface corrosion will adversely decrease strength, plasticity, and fatigue resistance of materials (e.g., steel, aluminum alloy, or carbon fiber-reinforced polymer). Fast and accurate solutions for internal defect inspection are important in many industrial production tasks, performing quality control of final products and eliminating potential threats to human safety. In the recent years, many NDT&E techniques including X-rays [1], ultrasound [2], vibration analysis [3], guided-wave [4], holographic interferometry [5], image speckle processing [6] and infrared thermography [7] have been proposed for inspection of internal defects and assessment of structural integrity of materials. The ultrasonic C-scanning is capable of accurately depicting the characteristics of the defects (e.g., sizes, depths, categories), and is typically utilized as the standard NDT&E technique to assess the internal quality of materials. However, it is expensive, labor-intensive, and time-consuming to deploy C-scanning or X-rays to perform diverse routine maintenance activities on a daily basis.

Compared with the classic NDT&E techniques (ultrasonic or X-rays), active infrared thermography provides a safe, fast, and low-cost solution for subsurface defects inspection of large structures. Active thermography techniques typically apply external energy sources (e.g., flashlights or halogen lamps) to induce a heat flux to the surface of the specimen. Lock-in thermography (LT) and pulsed thermography (PT) are the most used active thermography techniques. LT typically sets up halogen lamps to generate modulated heat waves for periodic heating of specimens, while PT uses high power flashlights to provide a heat pulse to the specimen surface. A detailed comparison of LT and PT techniques is provided in references [8]. In this paper, we only focused our investigations on the LT technique. Subsurface defects will incur local barriers during the heat propagation and thus cause abnormal temperature patterns at the surface of the specimen, which can be detected using a thermal camera (e.g., 8–14 μm uncooled long-wave infrared camera).

After capturing a sequence of raw infrared images, it is critically important to develop signal processing methods that can effectively detect and/or characterize subsurface defects in the presence of severe

* Corresponding author.

E-mail address: yangjx@zju.edu.cn (J. Yang).

<https://doi.org/10.1016/j.ndteint.2020.102246>

Received 29 October 2019; Received in revised form 27 January 2020; Accepted 27 February 2020

Available online 6 March 2020

0963-8695/© 2020 Elsevier Ltd. All rights reserved.

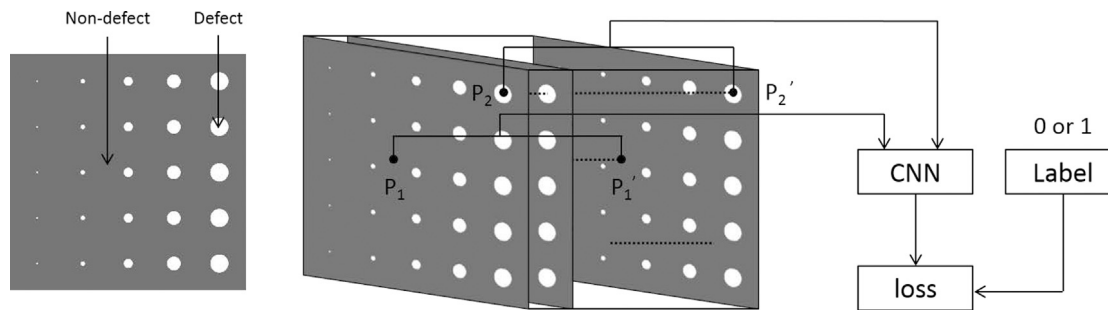


Fig. 1. An illustration of our proposed method. Different from most existing CNN models that directly classify individual regions/pixels as defective and non-defective ones, a novel two-stream CNN model is proposed for accurate classification/differentiation of defective and non-defective regions via similarity comparison of pairs of 1D thermal signal sequences.

Fixed Pattern Noise (FPN) and non-uniform heating effects. In the past decades, a wide variety of signal processing techniques have been proposed for active thermography NDT&E. It is reported in the literature that the phase information remains relatively independent of local heating conditions and less sensitive to noise disturbances [7,9,10]. Moreover, the analysis based on phase information allows deeper penetration into the material than the ones based on the amplitude signals [7]. Therefore, a standard technique is to extract the phase information in LT image sequences using the Four-point method [11,12] or discrete Fourier transformation method (DFT) [13,14]. Marinetti et al. and Rajic et al. successfully applied Principal component analysis (PCA), which is a well-known technique of multivariate linear data analysis, to thermal image sequences to enhance the thermal contrast between defects and non-defective areas [15,16]. However, the performance of the techniques mentioned above heavily depends on how well the extracted features (e.g., phase information or principal component) can characterize the abnormal thermal behaviors of subsurface defects. It is not a trivial task to construct/select the optimal feature representations for defect detection tasks in the presence of non-uniform heating, severe noise disturbance, and surface emissivity variations.

Recently, Convolutional Neural Network (CNN) models have significantly boosted the performance of various machine vision tasks including object detection [17,18], image segmentation [19] and target recognition [20]. Given a number of training samples, CNNs can automatically construct high-level representation by assembling the extracted low-level features. For instance, Simonyan et al. presented a very deep CNN model (VGG), which is commonly utilized as a backbone architecture for various computer vision tasks [18]. He et al. proposed a novel residual architecture to improve the training of very deep CNN models and achieve improved performance by increasing the depth of networks [17]. Moreover, some 3D CNN architectures have been proposed to extend the dimension of input data from 2D to 3D, processing video sequences for action recognition [21,22] or target detection [23].

Although CNN-based models have been successfully applied to solve many challenging image/signal processing tasks, very limited deep learning-based methods have been proposed to analyze thermal signal sequences for active thermography NDT&E. The major challenge is twofold. First, subsurface defects in materials typically cause abnormal thermal behaviors in individual video sequences of hundreds of image frames. Therefore, it is not applicable to fine-tune the pre-trained CNN models [17,18,24] working on single images. Moreover, it is extremely time-consuming to build a large-scale defect video dataset to train deep CNN models. Second, a complete LT experiment typically captures hundreds of frames (covering at least a periodical heating circle) to calculate the phase information robustly [7,11]. To handle such large video data, CNN models typically contain a large number of network parameters and are too slow for routine quality check tasks.

In this paper, we present a deep learning-based approach to process LT image sequences for NDT&E of subsurface defects in different

materials and at different excitation frequencies. Different from most existing CNN models that directly classify individual regions/pixels as defective and non-defective ones [25–27], a novel two-stream CNN architecture is proposed for accurate classification/differentiation of defective and non-defective regions via extracting/comparing features in a pair of 1D thermal signal sequences. In this manner, we can significantly increase the size of the training data by pairing two individually captured 1D thermal signals, thereby greatly ease the requirement for collecting a large number of thermal sequences of specimens with defects to train deep CNN models. Given thermal signal sequences captured at thousands of pixels, the proposed two-stream CNN model directly learns different thermal patterns/features presented by defective and non-defective regions instead of being hand-crafted before [11,14–16]. Moreover, we experimentally investigate a number of features/signals fusion alternatives (both fusion stages and functions) in an attempt to identify the optimal fusion scheme to perform the similarity comparison of two individual 1D thermal signals for nondestructive subsurface defect detection. Systematic evaluations are performed on specimens made of different materials (Q235 steel, aluminum alloy, and carbon fiber-reinforced polymer) at various excitation frequencies. Experimental results show that our proposed method outperforms the well-established LT data processing techniques, both qualitatively and quantitatively. This light-weight CNN model can process a 500-frame thermal image sequence within 60 s on a single NVIDIA Geforce Titan X GPU to facilitate routine quality inspection tasks. The contributions of this paper are summarized as follows:

(1) We formulate the subsurface defect NDT&E problem as a similarity comparison problem. To this end, a novel two-stream CNN model is presented to extract/compare features in a pair of 1D thermal signal sequences for accurate classification/differentiation of defective and non-defective regions. A noticeable advantage is that the size of the training data can be significantly increased by pairing two individually captured 1D thermal signals, thereby greatly easing the requirement for collecting a large-scale thermal video dataset to train deep CNN models.

(2) We experimentally evaluate the performance of architectures incorporating different fusion stages and functions, providing feasible design options to build up alternative CNN models for differentiating the thermal behaviors of defective and non-defective regions. Experimental results reveal that the CNN model based on middle-stage fusion generally performs better than alternatives based on early or late-stage fusion. Moreover, it is observed that the selection of fusion functions significantly affects the performance of subsurface defect detection, and is dependent on the choice of the fusion stage of two-stream signals/features.

(3) Our built-from-scratch model is trained using thousands of 1D thermal signal sequences captured at defective and non-defective areas/pixels without resorting to manually-designed features. Compared with the well-established LT data processing techniques [11,14–16], it requires no user-specific parameters or selection of the principal components and produces superior defect detection results in various materials at different frequencies both qualitatively and quantitatively.

The rest of this paper is organized as follows. Section 2 presents the details of LT experiments and the proposed CNN-based model for processing LT image sequences. Section 3 provides the performance analysis of different features/signals fusion alternatives and experimental comparison with other LT data processing solutions. Finally, Section 4 concludes the paper.

2. Methodology

After capturing raw infrared images of a specimen that contains subsurface defects, we aim to develop signal processing methods for effectively detecting and/or characterizing subsurface defects in the presence of non-uniform heating, FPN disturbances, and surface emissivity variations. Note it is time-consuming to capture a large-scale video dataset of specimens with defects in LT experiments. As a result, it is challenging to train deep CNN models that directly classify individual regions/pixels as defective and non-defective ones. In this paper, we propose a novel two-stream CNN architecture to extract/compare features in a pair of 1D thermal signal sequences for accurate classification/differentiation of defective and non-defective regions, as illustrated in Fig. 1. By pairing two individually captured 1D thermal signals, we can significantly increase the size of the training data to train deep CNN models. Our proposed method directly learns how to construct feature representations from a large number of real-captured thermal data, therefore it can implicitly take into account various types of disturbances during LT experiments (e.g., non-uniform heating, FPNs, and material/emissivity variations). Moreover, we experimentally investigate a number of network alternatives (performing signal/feature fusion at different stages and using different fusion functions) in an attempt to identify the feasible fusion schemes to process two-stream thermal signals for differentiating the thermal behaviors of defective and non-defective regions.

2.1. LT data acquisition system and specimens

The hardware configuration of the LT data acquisition system is illustrated in Fig. 2. Two 1000 W halogen lamps (the peak power is ~ 1500 W) and a signal modulator are utilized to generate sinusoidal waves of different frequencies (0.025, 0.05, 0.075, and 0.1 Hz) for heating specimens. The specimens are placed approximately 80 cm in front of the halogen lamps. A Xenics Gobi 640 long-wave infrared (LWIR) camera (working spectral band is 8–14 μm) is utilized to capture thermal sequences on the surfaces of specimens. The resolution of the infrared camera is 640×480 pixels, and its frame rate is 50 fps. The Noise Equivalent Temperature Difference (NETD) of this LWIR camera is 50mK at 30 $^{\circ}\text{C}$ with F/1 lens.

In total, we prepare six specimens using three different materials, including steel (Q235), aluminum alloy (Al), and Carbon Fiber Reinforced Polymer (CFRP). The size of a specimen is 300×300 mm, and its thickness is 4 mm. Note these specimens are manufactured and painted individually, and they present different thermal patterns during LT experiments. Three specimens (one Q235, Al, and CFRP sample each) are used to generate thermal signal sequences for training the CNN model while the others are used for testing. As illustrated in Fig. 3, the simulated defects are 25 bottom flat holes of different diameters (20, 15, 10, 5 and 2 mm) which are manufactured at various depths (1.5, 1.2, 1, 0.8 and 0.5 mm). Based on a single specimen, we can capture thermal images of defects with different sizes at the same depth (e.g., 5 mm and 10 mm circular defects both at 1 mm depth) and the ones with the same size at different depths (e.g., 5 mm circular defects at 1 and 1.5 mm depths). For each specimen, we set four different excitation frequencies at 0.1, 0.075, 0.05, and 0.025 Hz and capture LT image sequences. In our experiments, we recorded 24 individual thermal video sequences (on six specimens using four different excitation frequencies), each of which contains 2000 thermal images. Note a minimum 2000 frames are required to cover a complete periodic heating process (the frame rate of the LWIR camera is 50 fps) for the LT experiment using the lowest frequency (0.025 Hz).

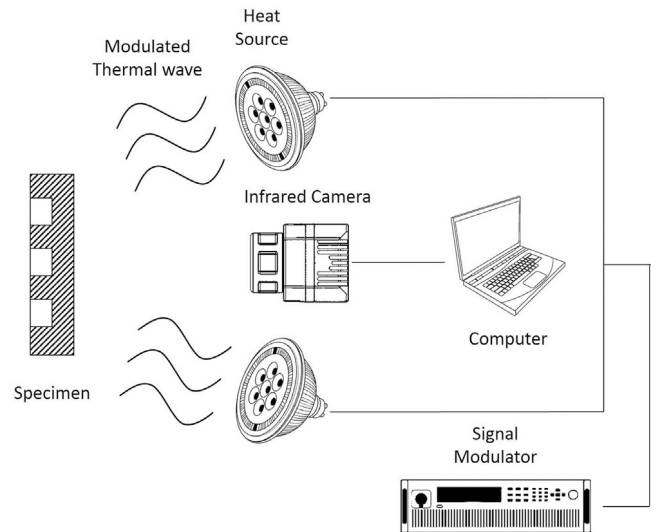


Fig. 2. The hardware configuration of the LT data acquisition system.

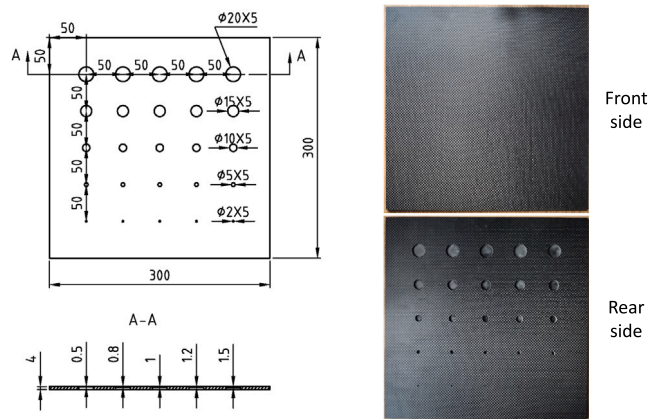


Fig. 3. The specimens used in LT experiments. Left: The schematic sketch of a specimen. Right: Front/rear sides of a CFRP specimen with simulated defects (25 bottom flat holes of different diameters and at various depths).

2.2. Thermal data pre-processing

In LT experiments, the periodical heating process of target specimens typically requires a few minutes for the surface temperature reaching the steady-state [28]. To speed up the data acquisition process, we start capturing infrared images after heating the specimens for 10 s. After capturing sequences of raw thermal signals, we firstly applied a second-order polynomial model to fit the time-varying surface temperatures at individual pixels. The pixel-wise fitted data is subtracted from the raw thermal images to compensate for the increasing trend in the surface temperature due to the dc component of the heating [29,30]. Then, we normalized all the data within a thermal image sequence to the range of 0 to 1. Finally, we decreased the frame number of each video sequence from 2000 to 500 through temporal downsampling to improve the computational efficiency of LT data analysis without sacrificing classification accuracy.

The square-size specimens are covered in an image region of 380×380 pixels. The total numbers of defective and non-defective pixels are $\sim 8\text{K}$ and $\sim 130\text{K}$, respectively. We uniformly selected a number of defective (6500) and non-defective (13,000) pixels in the captured images by referring to the location of pre-defined defects in specimens. Then we randomly paired thermal sequences captured at defective/non-defective pixels to generate training data with ground-truth labels.

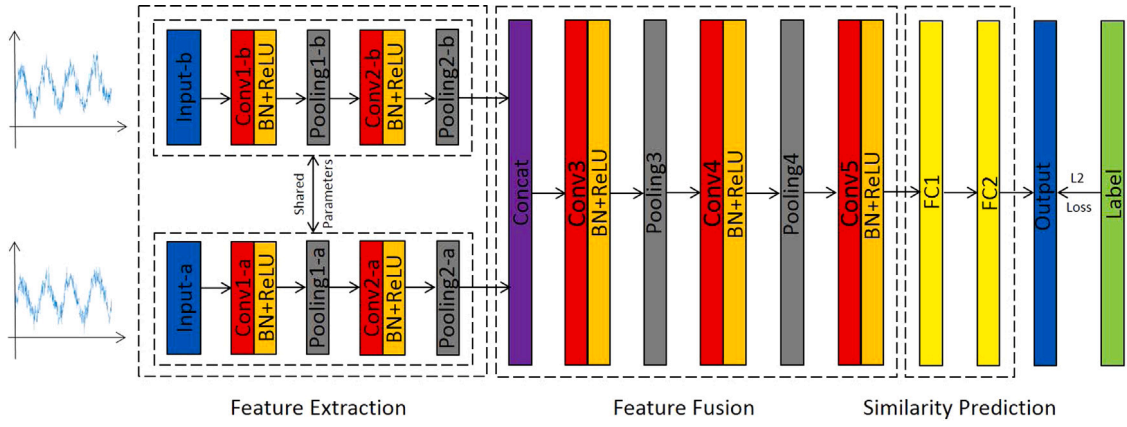


Fig. 4. Network architectures of our proposed TS-LTS model. Our TS-LTS contains three major parts: feature extraction, feature fusion, and similarity prediction. The detailed configurations of individual layers in the TS-LTS model are shown in Table 1.

Table 1

The detailed configurations of individual layers in the TS-LTS model. The filter parameters are indicated as $C \times H \times W$ where C is the channel number, H is the kernel height, and W is the kernel width.

Layers	Kernel size	Output size
Input		1×500
Conv1-a,b	$4 \times 1 \times 15$	$4 \times 1 \times 500$
Pooling1-a,b	1×5	$4 \times 1 \times 100$
Conv2-a,b	$4 \times 1 \times 15$	$4 \times 1 \times 100$
Pooling2-a,b	1×2	$4 \times 1 \times 50$
Conv3	$4 \times 1 \times 7$	$4 \times 1 \times 50$
Pooling3	1×2	$4 \times 1 \times 25$
Conv4	$8 \times 1 \times 7$	$8 \times 1 \times 25$
Pooling4	1×2	$8 \times 1 \times 13$
Conv5	$16 \times 1 \times 7$	$16 \times 1 \times 13$
FC1	32	32
FC2	1	1

More specifically, we assigned a positive label for a pair of thermal sequences if they are captured at two pixels of the same classes (e.g., both defective or non-defective pixels). In comparison, we assigned a negative label for two signals if they are captured at a defective pixel and a non-defective one. In this manner, we can significantly increase the size of the training data by pairing individually captured 1D thermal signals. In total, we generate 3,271,680 pairs of thermal signals with equal numbers of positive and negative labels for training deep CNN models.

2.3. Architecture

As illustrated in Fig. 4, we design a two-stream CNN model to process LT image sequences (TS-LTS) for NDT&E of subsurface defects in different materials. TS-LTS model consists of three major processing modules, including (1) feature extraction, (2) feature fusion, and (3) similarity prediction. The detailed configurations of individual layers are shown in Table 1. Given thermal signals captured at thousands of defective and non-defective pixels, the TS-LTS model learns how to generate the optimal feature representation to characterize/differentiate thermal behaviors of defective and non-defective regions.

Feature extraction: A two-stream feature extraction module is deployed to extract feature maps in a pair of 1D LT sequences and reduce the dimension of input signals. It consists of a number of stacked convolutional, normalization/activation, and pooling layers with shared parameters. Here we use convolutional layers with kernels of large sizes to handle the long 1D thermal signals (1×500). In our experiment, the kernel sizes of two convolutional layers are set to 15. The Batch Normalization (BN) operation is deployed before every activation function layer to normalize the means and variances

of each layer's inputs, and its transformation parameters are adaptively learned during the training stage. ReLU activation function is utilized to embed more nonlinear terms into the network. Finally, we use two pooling layers (the kernel sizes are set to 5 and 2 respectively) to reduce the size of the input signals, computing the compact yet distinctive representations for individual 1D thermal signals.

Feature fusion: A fusion layer is deployed after the two-stream feature extraction module to integrate the semantic feature maps computed on two individual channels. Here we consider three different fusion functions including f^{cat} (concatenation fusion), f^{max} (maximization fusion) and f^{sum} (summation fusion) [31–33]. Detailed comparisons of these three fusion functions are provided in Section 3.2. Given the channel-wise combined features, a number of convolutional and pooling layers are subsequently deployed to compute the optimal fused feature map for the following similarity comparison task.

Similarity prediction: The output of the feature fusion module is then fed to the similarity prediction module. It contains two fully connected linear decision layers and computes a single value output to predict the similarity level between two different thermal signals. The predicted score is compared with the ground-truth labels through the L2 loss, and the learning objective function is defined as

$$\mathcal{L} = \frac{\lambda}{2} \|\omega\|_2^2 + \frac{1}{2N} \sum_{i=1}^N \|y_i - o_i\|_2^2, \quad (1)$$

where ω are the parameters of the proposed TS-LTS model, and λ is the weight decay parameter to avoid data overfitting. For N pairs of training data, o_i is the i th predictions of the TS-LTS model, and y_i is its corresponding ground-truth label manually set 0 or 1. The TS-LTS model is trained in a strongly supervised manner by minimizing the loss function \mathcal{L} . During the testing phase, we randomly selected 20 reference pixels at different locations in an image. For each reference pixel, we associated it with the others and computed the sum of their similarity scores. Because of the low probability of defects occurrence, we use the reference pixel with the maximal summed similarity score as the reference non-defective pixel and compute its similarity levels with other pixels.

2.4. Fusion alternatives

In this section, we investigate a number of features/signals fusion alternatives (fusion functions and stages) in an attempt to identify feasible design options for comparing two individual 1D thermal signals.

Based on some commonly used CNN architectures for tracking/comparing 2D image patches (i.e., the Siamese and 2-channel networks [32,34]), we design three different architectures which perform feature/signal fusions at different stages as illustrated in Fig. 5. In

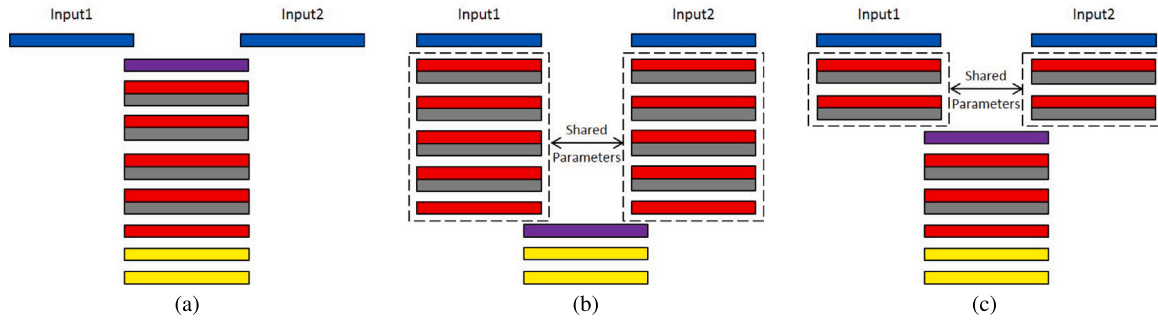


Fig. 5. The CNN architectures implementing different fusion schemes. (a) Early-stage fusion, (b) Late-stage fusion, and (c) Middle-stage fusion. Note that the blue blocks represent the input layers, the red blocks represent the convolutional and ReLU layers, the gray blocks represent the pooling layers, the purple blocks represent fusion layers, and the yellow blocks represent the fully connected layers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the first architecture (Fig. 5(a)), we directly combine two individual thermal signals without performing feature extraction (early-stage fusion) and then feed the 2-channel signal to the following convolutional layers to compute semantic feature maps for similarity prediction. In comparison, two branches of convolutional layers sharing the same parameters are deployed in the second architecture to extract semantic features in two individual channels (Fig. 5(b)). Then the individually extracted features are combined to generate the final prediction (late-stage fusion). Fig. 5(c) shows the architecture of our proposed TS-LTS. Different from the first and second architectures using the early and late fusion strategies, TS-LTS puts a fusion layer after a number of convolutional layers to combine the feature maps computed on two individual channels (middle-stage fusion). Moreover, a number of convolutional layers are utilized to fine-tune the channel-wise combined features before predicting similarity scores.

Moreover, we consider three different fusion functions (f^{cat} , f^{max} and f^{sum}) which have different working principles and thus lead to different detection performances. The concatenation fusion function f^{cat} stacks two-stream feature maps \mathbf{p} and \mathbf{q} at the same spatial locations but across the feature channels. In comparison, the maximization fusion function f^{max} outputs the maximum response of feature maps \mathbf{p} and \mathbf{q} at the same spatial locations and in the same channel. Similarly, the summation fusion function f^{sum} calculates the sum of feature maps \mathbf{p} and \mathbf{q} at the same spatial locations and in the same channel, combining the two-stream feature maps using equal weights.

For a fair comparison, we set the number of convolutional layers and kernel sizes to the same values for CNN architectures incorporating different fusion stages and functions. We will experimentally evaluate nine different design combinations (incorporating three different fusion stages and three different fusion function) and discuss the best scheme to perform features/signals fusion to achieve high-accuracy similarity prediction of two 1D thermal signals in Section 3.2.

3. Experiments and results

3.1. Implementation details

The publicly available Caffe platform is used for the proposed CNN model implementation [35]. We use 3,271,680 pairs for model training, and the batch size is set to 64. The maximum training iteration is set to 200,000. The learning rate LR is initially set to 0.01 and is reduced using the step strategy, updating LR by a 0.1 decreasing factor after 50,000 iterations. The training process is performed using the SGD training policy with momentum 0.9 and weight decay 0.0005. The momentum and weight decay are fixed during the entire training process. The proposed model is trained on a NVIDIA TITAN X GPU (12G memory) within 30 min.

3.2. Performance analysis

In this section, we set up experiments to evaluate the performance of CNN models based on different two-stream features/signals fusion alternatives (both fusion stages and functions). In total, we consider nine different design combinations which incorporate three different fusion stages and three different fusion function. Note the number of convolutional layers and kernel sizes are set to the same values for all three different architectures. The only difference among them is that the signal/feature fusion is performed in different stages and using different functions in CNN models. We train all models in a strongly supervised manner using the loss function defined in Eq. (1).

We adopt the global contrast index (C_G) between defective (Df) and non-defective ($N - Df$) regions to quantitatively evaluate the detection results of different models [10,36]. The C_G index computes the difference between all defective pixels and non-defective pixels in an image and then divides the value by the background noise (e.g., the standard deviation of the background pixels) as

$$C_G = \frac{|\text{mean}(S_{Df}) - \text{mean}(S_{N-Df})|}{\text{std}(S_{N-Df})}, \quad (2)$$

where S_{Df} and S_{N-Df} are the values of the defective and non-defective pixels, respectively. Note higher C_G index indicates a more distinctive contrast between defects and the background.

The performance (C_G index [10,36]) of CNN models incorporating different fusion schemes are quantitatively compared in Table 2. It is observed that the selection of signal/feature fusion stages and functions significantly affects the performance of subsurface defect detection. We made two important observations. Firstly, the proposed TS-LTS model (based on the middle-stage fusion) performs generally better than the CNN models incorporating early or late fusion schemes by achieving higher C_G values for different materials and at various excitation frequencies. The experimental results suggest that it is more suitable to firstly convert raw thermal signals to feature representations in individual channels and then perform feature-level fusion for comparing the similarity of two 1D thermal signals. Secondly, it is critically important to select the appropriate fusion functions for architectures that perform feature/signal fusions at different stages. For instance, the concatenation fusion function f^{cat} works well for architectures incorporating either early or middle stage fusion schemes but did not produce satisfactory results for the CNN model based on the late-stage fusion. Given the two-stream semantic features computed in late convolutional stages, it is better to identify the more distinct ones through the maximization fusion function (f^{max}) for the similarity prediction.

3.3. Comparisons with other LT data processing techniques

We compare the proposed TS-LTS model with well-established LT data analysis techniques, including the Four-point [11], DFT [13,14], and PCA [15,16] methods.

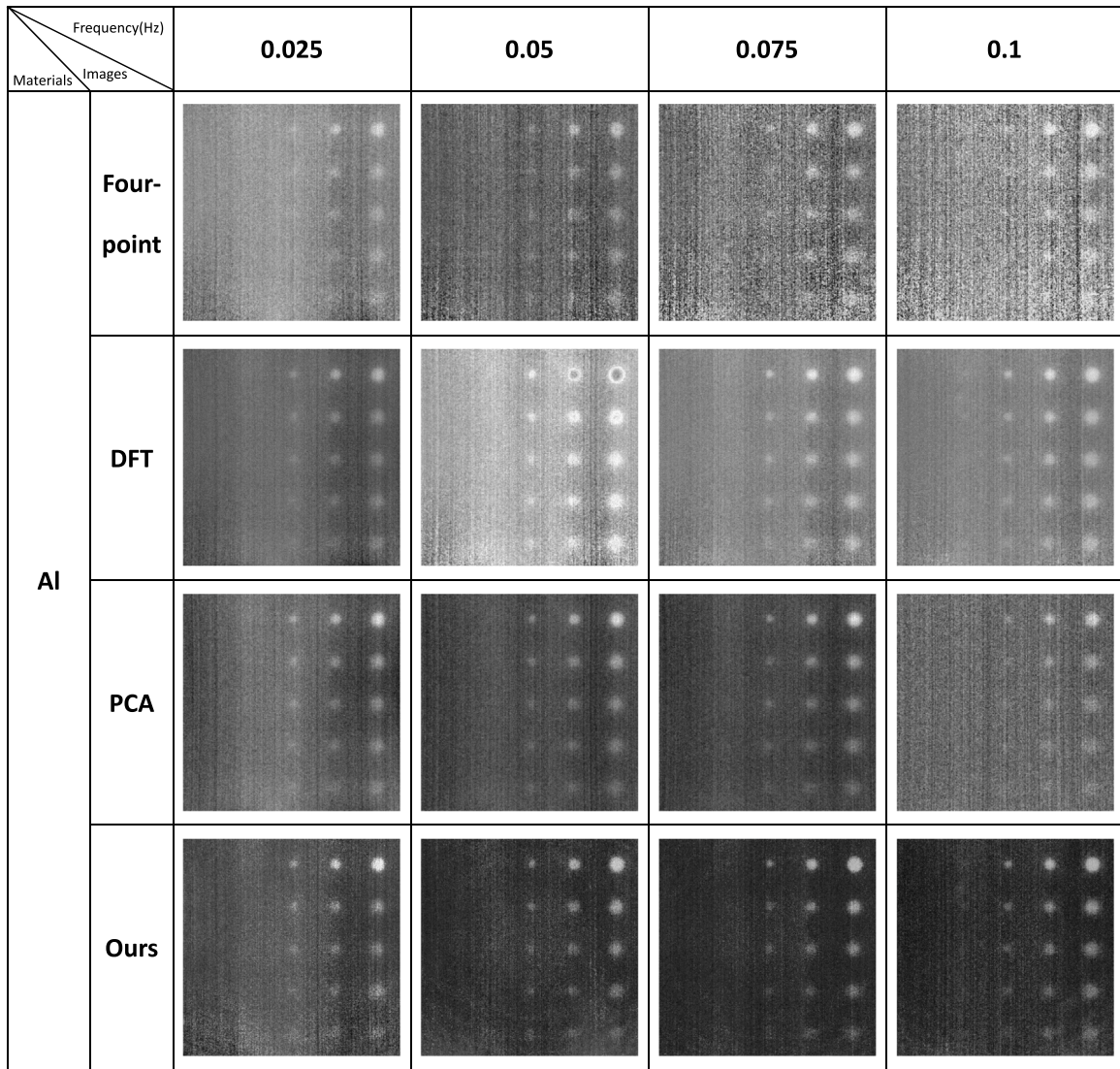


Fig. 6. Qualitative comparison of different LT data analysis techniques on a specimen made of aluminum alloy at four excitation frequencies (0.025, 0.05, 0.075, and 0.1 Hz). All images are normalized to the 0–1 range for visualization.

Table 2

Quantitative comparison (global contrast index C_G [10,36]) of CNN models which perform feature/signals fusion in different stages (early, middle, and late) and using different fusion functions (f^{sum} , f^{max} , and f^{cat}). The best results are highlighted in bold.

Materials	Frequency (Hz)	Early Fusion			Late Fusion			Middle Fusion		
		f^{sum}	f^{max}	f^{cat}	f^{sum}	f^{max}	f^{cat}	f^{sum}	f^{max}	f^{cat}
Al	0.025	0.2798	0.2943	1.2110	0.2790	1.2379	0.0215	1.4657	0.4514	1.5887
	0.05	0.1359	0.6043	2.3800	0.0925	2.0709	0.1730	2.0181	2.0111	2.2169
	0.075	0.1642	0.8451	1.3221	0.8303	2.0897	0.0484	2.1363	2.3969	2.9570
	0.1	0.1127	0.4107	1.3753	0.9250	1.7868	0.0334	1.2654	1.9880	2.5268
Q235	0.025	0.2102	0.4772	1.2822	1.0910	1.3228	0.5229	2.1806	1.9901	3.1886
	0.05	1.9533	2.7739	2.1958	0.4495	2.9495	0.6600	4.1505	4.9652	5.0440
	0.075	2.3480	3.5068	3.7097	1.6127	3.6364	0.9955	4.4956	3.3001	6.5652
	0.1	2.3135	3.8060	5.1393	0.4861	3.7772	0.4293	3.0931	4.0047	5.7612
CFRP	0.025	2.9486	2.6894	1.6298	1.7241	3.0321	0.8826	2.7007	2.6035	4.5238
	0.05	4.0327	4.5649	3.8356	0.5481	6.2490	0.1368	6.5712	4.0242	7.1299
	0.075	2.6450	3.1114	4.4915	1.5148	5.1927	0.1368	5.6409	5.5297	8.0582
	0.1	1.2772	1.6655	4.3467	0.4669	3.4182	0.9507	4.5032	3.3559	6.5646

Four-point method selects four equidistant data points S_1, S_2, S_3, S_4 in a thermal wave signal I and computes the magnitude (A) and phase (φ) information of this pixel as

$$A = \sqrt{(S_3 - S_1)^2 + (S_4 - S_2)^2}, \quad (3)$$

$$\varphi = \arctan\left(\frac{S_3 - S_1}{S_4 - S_2}\right). \quad (4)$$

DFT method transfers the 1D thermal wave signal I to a real part (Re) and an imaginary part (Im) and calculates the amplitude and

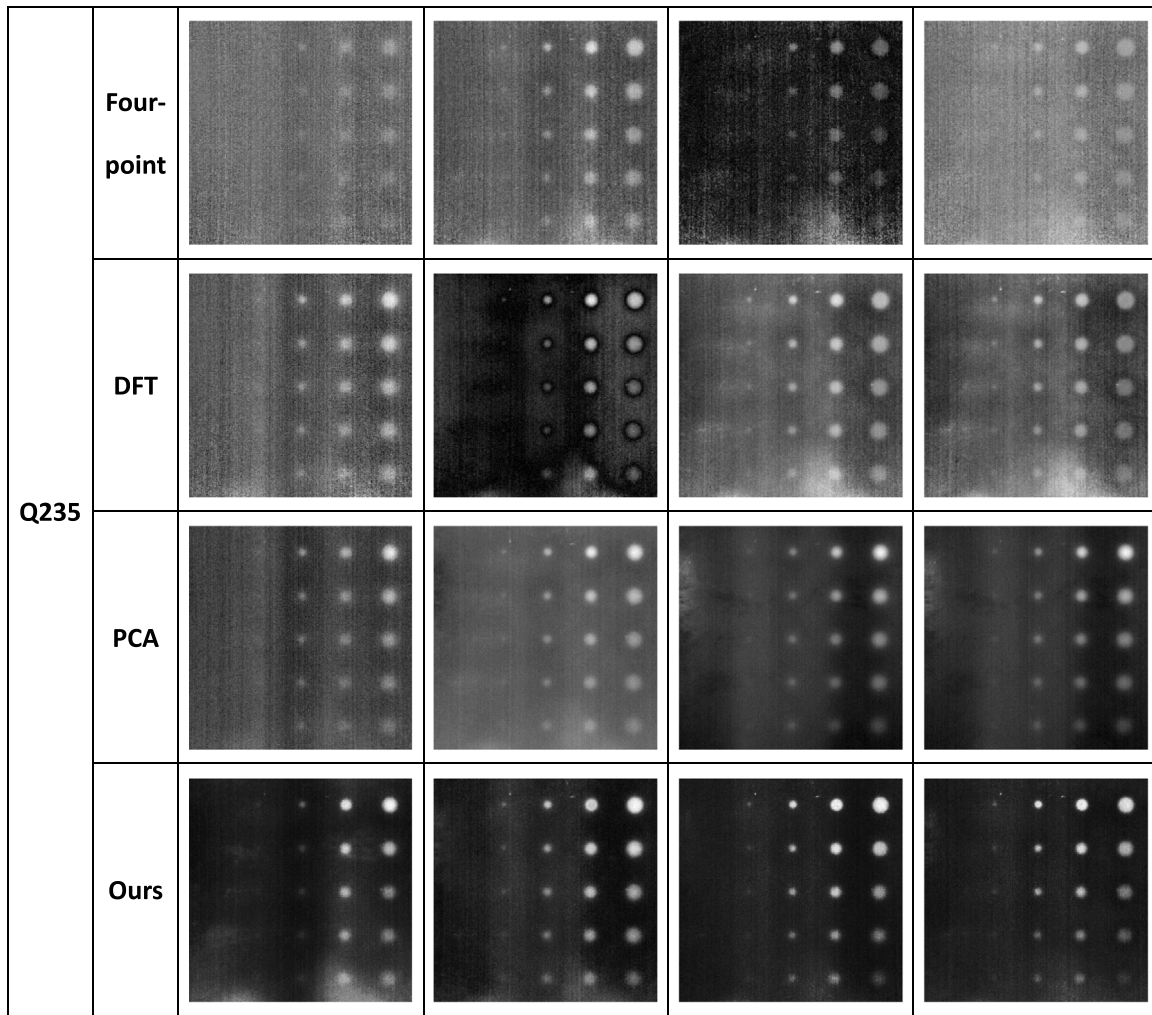


Fig. 7. Qualitative comparison of different LT data analysis techniques on a specimen made of Q235 steel at four excitation frequencies (0.025, 0.05, 0.075, and 0.1 Hz). All images are normalized to the 0–1 range for visualization.

phase information as

$$A = \sqrt{Re^2 + Im^2}, \quad (5)$$

$$\varphi = \arctan\left(\frac{Im}{Re}\right). \quad (6)$$

Note both Four-point and DFT methods make use of the computed phase data (φ), which remains relatively independent of local heating conditions and less sensitive to noise disturbances [7,9], for defect detection/characterization.

PCA method converts the 3D array (LT image sequences) to a 2D matrix A by reshaping pixels of each 2D image frame to a 1D column vector. And the scatter matrix S can be got by

$$S = (A - A_{mean})(A - A_{mean})^T \quad (7)$$

Then it performs singular value decomposition (SVD) of the transformed 2D matrix S as

$$S = UDU^T, \quad (8)$$

where U and D are the eigenvector and diagonal matrices of A . Then a principal eigenvector (U_p) is selected from the matrix U to compute the output as

$$A' = U_p^T A, \quad (9)$$

where A' is a 1D column vector and is further reshaped to a 2D image for visualizing defects. The selection of the optimal principal eigenvector is typically required to enhance the contrast between defects and

non-defective areas [37]. In our experiments, we found that the second principal component typically provides the most distinctive results to characterize subsurface defects, which is consistent with many previous research works [16,37,38]. It has been reported in previous literature works that the PCA technique can generate better subsurface defect detection results compared with the four-point or FFT methods [39–41], suppressing the influence of uneven heating and enhancing the contrasts of defects.

Table 3 summarizes the quantitative results on specimens made of three different materials (Al, Q235, and CFRP) at four excitation frequencies (0.025, 0.05, 0.075, and 0.1 Hz). Our proposed TS-LTS model outperforms other LT data analysis methods, achieving significantly higher C_G values. Such improvement is particularly evident when compared with the Four-point and DFT methods based on the computed phase information. The experimental results demonstrate the effectiveness of deep learning-based feature extraction, generating better representations of 1D thermal signals than hand-crafted ones. Another noticeable advantage of TS-LTS is that its feature extraction and comparison are automatically learned from thermal signal sequences captured at thousands of pixels in an end-to-end manner, requiring no further parameter fine-tuning or principal component selection which is heavily human expertise dependent. Moreover, this light-weight CNN model can process a 500-frame thermal image sequence within 60 s on a single NVIDIA Geforce Titan X GPU, which is fast enough for routine quality inspection tasks.

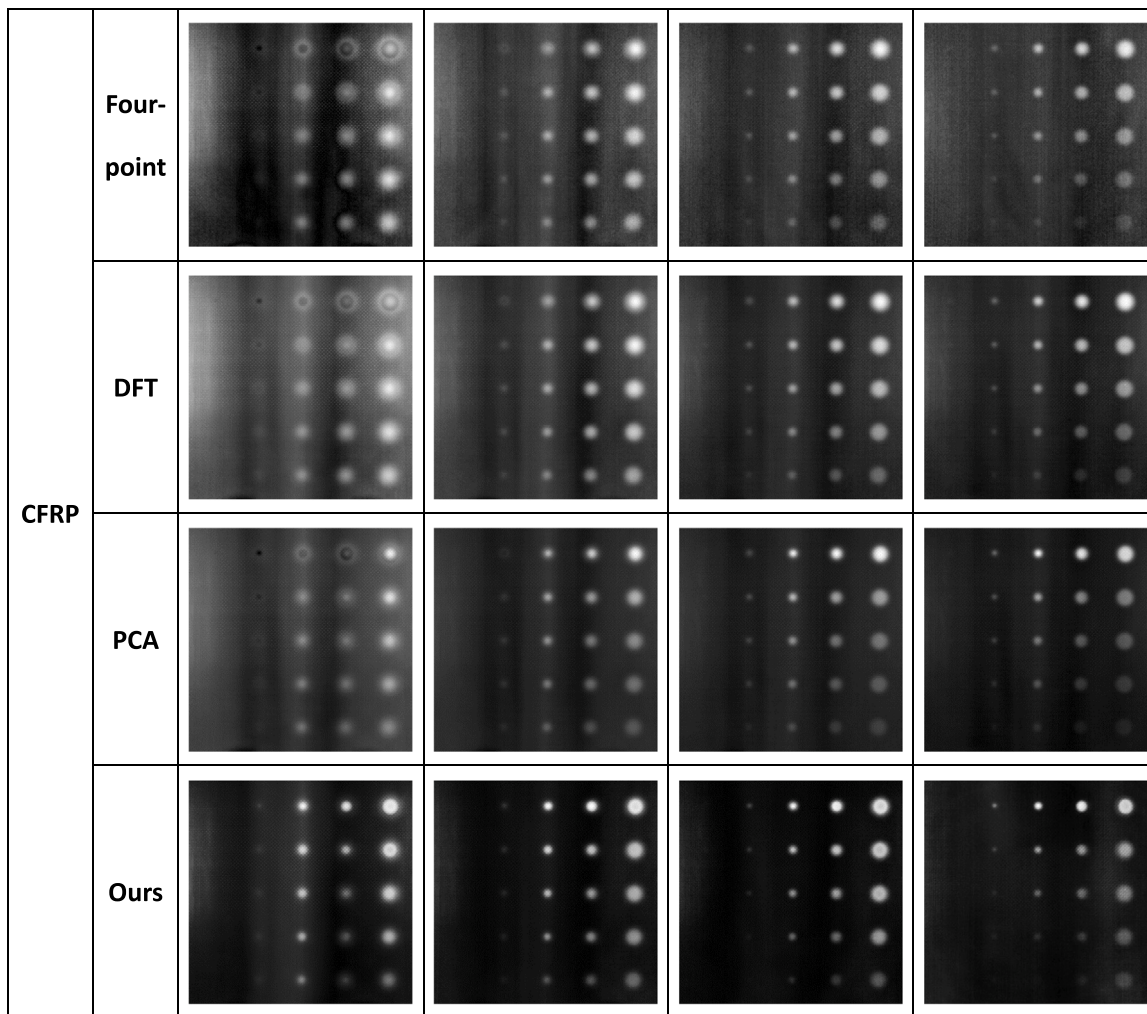


Fig. 8. Qualitative comparison of different LT data analysis techniques on a specimen made of CFRP at four excitation frequencies (0.025, 0.05, 0.075, and 0.1 Hz). All images are normalized to the 0–1 range for visualization.

Figs. 6, 7, and 8 visualize the qualitative results of defect detection on different materials at various excitation frequencies. Compared with other well-established LT data analysis techniques, our proposed TS-LTS model can achieve better defect detection results. It is able to generate more obvious contrast between defective and non-defective regions and remain more robust to thermal signal disturbances such as background noise and non-uniform heating.

4. Conclusion

In this paper, we describe a novel deep learning-based approach to achieve accurate subsurface defect detection by comparing the similarity of 1D thermal signals captured at defective and non-defective pixels. The size of the training data can be significantly increased by pairing two individually captured 1D thermal signals. Without resorting to manually-designed features, this built-from-scratch model is trained using thousands of 1D thermal signal sequences captured at defective and non-defective areas/pixels. Also, we set up experiments to evaluate the performance of network alternatives performing signal/feature fusion at different stages and using different functions. Compared with the well-established LT data processing techniques, our proposed method requires no user-specific parameters or selection of the principal components and generates better defect detection results (generating more obvious contrast between defective and non-defective

Table 3

Quantitative comparison (global contrast index C_G [10,36]) between our proposed TS-LTS model and other LT images analysis techniques on specimens made of three different materials (Al, Q235, and CFRP) at four excitation frequencies (0.025, 0.05, 0.075, and 0.1 Hz). The best results are highlighted in bold.

Materials	Frequency (Hz)	Four-point	DFT	PCA	Ours
Al	0.025	0.7566	1.2620	1.2166	1.5887
	0.05	1.0886	1.5426	2.1399	2.2169
	0.075	0.8065	2.3198	2.3272	2.9570
	0.1	0.8008	2.3898	1.1382	2.5268
Q235	0.025	1.0092	2.5077	3.0294	3.1886
	0.05	2.4321	2.3960	3.8187	5.0440
	0.075	1.3196	2.2781	2.7971	6.5652
	0.1	1.0555	1.5584	3.2271	5.7612
CFRP	0.025	2.0178	1.9855	2.4831	4.5238
	0.05	4.0850	4.4100	4.8676	7.1299
	0.075	4.4803	5.2142	5.7759	8.0582
	0.1	3.9216	4.5186	4.9764	6.5646

regions and remaining more robust to thermal signal disturbances such as background noise and non-uniform heating) in various materials at different frequencies. In the future, our research works will focus on developing new CNN models that are capable of extracting more distinctive features and identifying deeper and more complex subsurface defects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (No. 51605428 and No. 51575486).

References

- [1] Hassen Ahmed Arabi, Taheri Hossein, Vaidya Uday K. Non-destructive investigation of thermoplastic reinforced composites. *Composites B* 2016;97:244–54.
- [2] Hosur MV, Murthy CRL, Ramamurthy TS, Shet Anita. Estimation of impact-induced damage in CFRR laminates through ultrasonic imaging. *NDT & E Int* 1998;31(5):359–74.
- [3] Pérez Marco A, Gil Lluís, Oller Sergio. Impact damage identification in composite laminates using vibration testing. *Compos Struct* 2014;108:267–76.
- [4] Chillara Vamshi Krishna, Lissenden Cliff J. Review of nonlinear ultrasonic guided wave nondestructive evaluation: Theory, numerics, and experiments. *Opt Eng* 2015;55(1):011002.
- [5] Sfarra Stefano, Ibarra-Castaneda C, Lambiasi Francesco, Paoletti Domenica, Di Ilio Antoniomaria, Maldague X. From the experimental simulation to integrated non-destructive analysis by means of optical and infrared techniques: Results compared. *Meas Sci Technol* 2012;23(11):115601.
- [6] Sfarra S, Regi M, Santulli C, Sarasini F, Tirillò J, Perilli S. An innovative nondestructive perspective for the prediction of the effect of environmental aging on impacted composite materials. *Internat J Engrg Sci* 2016;102:55–76.
- [7] Maldague Xavier. Theory and practice of infrared technology for nondestructive testing. 2001.
- [8] Pickering Simon, Almond Darryl. Matched excitation energy comparison of the pulse and lock-in thermography NDE techniques. *NDT & E Int* 2008;41(7):501–9.
- [9] Palumbo Davide, Cavallo Pasquale, Galietti Umberto. An investigation of the stepped thermography technique for defects evaluation in GFRP materials. *NDT & E Int* 2019;102:254–63.
- [10] Choi Manyong, Kang Kisoo, Park Jeonghak, Kim Wontae, Kim Koungsuk. Quantitative determination of a subsurface defect of reference specimen by lock-in infrared thermography. *Ndt & E Int* 2008;41(2):119–24.
- [11] Ibarra-Castaneda Clemente, Piau Jean-Marc, Guilbert Stéphane, Avdelidis Nicolas P, Genest Marc, Bendada Abdelhakim, et al. Comparative study of active thermography techniques for the nondestructive evaluation of honeycomb structures. *Res Nondestruct Eval* 2009;20(1):1–31.
- [12] BuSse Gi, Wu D, Karpen W. Thermal wave imaging with phase sensitive modulated thermography. *J Appl Phys* 1992;71(8):3962–5.
- [13] Ibarra-Castaneda Clemente, Maldague Xavier. Pulsed phase thermography reviewed. *Quant Infrared Thermogr J* 2004;1(1):47–70.
- [14] Pitarresi G. Lock-in signal post-processing techniques in infra-red thermography for materials structural evaluation. *Exp Mech* 2015;55(4):667–80.
- [15] Marinetti Sergio, Grinzato Ermanno, Bison Paolo G, Bozzi Edoardo, Chimenti Massimo, Pieri Gabriele, et al. Statistical analysis of IR thermographic sequences by PCA. *Infrared Phys Technol* 2004;46(1–2):85–91.
- [16] Rajic Nikolas. Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures. *Compos Struct* 2002;58(4):521–8.
- [17] He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 770–778.
- [18] Simonyan Karen, Zisserman Andrew. Very deep convolutional networks for large-scale image recognition. *ICLR* 2015.
- [19] Long Jonathan, Shelhamer Evan, Darrell Trevor. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. pp. 3431–3440.
- [20] Li Haoxiang, Lin Zhe, Shen Xiaohui, Brandt Jonathan, Hua Gang. A convolutional neural network cascade for face detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. pp. 325–334.
- [21] Ji Shuiwang, Xu Wei, Yang Ming, Yu Kai. 3D convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intel* 2012;35(1):221–31.
- [22] Molchanov Pavlo, Gupta Shalini, Kim Kihwan, Kautz Jan. Hand gesture recognition with 3D convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2015. pp. 1–7.
- [23] Dou Qi, Chen Hao, Yu Lequan, Zhao Lei, Qin Jing, Wang Defeng, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE Trans Med Imaging* 2016;35(5):1182–95.
- [24] Iandola Forrest N, Han Song, Moskewicz Matthew W, Ashraf Khalid, Dally William J, Keutzer Kurt. SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size. 2016, [arXiv:1602.07360](https://arxiv.org/abs/1602.07360).
- [25] Yousefi Bardia, Kalhor Davood, Usamentiaga Fernández Rubén, Lei Lei, Castanedo Clemente Ibarra, Maldague Xavier PV. Application of deep learning in infrared non-destructive testing. In: QIRT 2018 proceedings. 2018.
- [26] Du Bolun, He Yigang, Duan Jiajun, Zhang Yaru. Intelligent classification of silicon photovoltaic cell defects based on eddy current thermography and convolution neural network. *IEEE Trans Ind Inf* 2019.
- [27] Li Xiaoxia, Yang Qiang, Lou Zhuo, Yan Wenjun. Deep learning based module defect analysis for large-scale photovoltaic farms. *IEEE Trans Energy Convers* 2018;34(1):520–9.
- [28] Liu Junyan, Yang Wang, Dai Jingmin. Research on thermal wave processing of lock-in thermography based on analyzing image sequences for NDT. *Infrared Phys Technol* 2010;53(5):348–57.
- [29] Chatterjee Krishnendu, Tuli Suneet. Image enhancement in transient lock-in thermography through time series reconstruction and spatial slope correction. *IEEE Trans Instrum Meas* 2011;61(4):1079–89.
- [30] Chatterjee Krishnendu, Tuli Suneet, Pickering Simon G, Almond Darryl P. A comparison of the pulsed, lock-in and frequency modulated thermography nondestructive evaluation techniques. *Ndt & E Int* 2011;44(7):655–67.
- [31] Hwang Soonmin, Park Jaesik, Kim Namil, Choi Yukyung, So Kweon In. Multi-spectral pedestrian detection: Benchmark dataset and baseline. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. pp. 1037–1045.
- [32] Zagoruyko Sergey, Komodakis Nikos. Learning to compare image patches via convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015. pp. 353–4361.
- [33] Feichtenhofer Christoph, Pinz Axel, Zisserman Andrew. Convolutional two-stream network fusion for video action recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 1933–1941.
- [34] Bertinetto Luca, Valmadre Jack, Henriques Joao F, Vedaldi Andrea, Torr Philip HS. Fully-convolutional siamese networks for object tracking. In: European conference on computer vision. Springer; 2016, p. 850–65.
- [35] Jia Yangqing, Shelhamer Evan, Donahue Jeff, Karayev Sergey, Long Jonathan, Girshick Ross, et al. Caffe: Convolutional architecture for fast feature embedding. 2014, [arXiv preprint arXiv:1408.5093](https://arxiv.org/abs/1408.5093).
- [36] Madruga Francisco J, Ibarra-Castaneda Clemente, Conde Olga M, López-Higuera José M, Maldague Xavier. Infrared thermography processing based on higher-order statistics. *NDT & E Int* 2010;43(8):661–6.
- [37] Alvarez-Restrepo CA, Benitez-Restrepo HD, Tobón LE. Characterization of defects of pulsed thermography inspections by orthogonal polynomial decomposition. *NDT & E Int* 2017;91:9–21.
- [38] Vavilov Vladimir Platonovich, Nesteruk Denis Alexeevich, Shiryayev Vladimir Vasilievich, Ivanov AI, Swiderski W. Thermal (infrared) tomography: Terminology, principal procedures, and application to nondestructive testing of composite materials. *Russ J Nondestruct Test* 2010;46(3):151–61.
- [39] Shrestha Ranjit, Choi Manyong, Kim Wontae. Thermographic inspection of water ingress in composite honeycomb sandwich structure: A quantitative comparison among lock-in thermography algorithms. *Quant Infrared Thermogr J* 2019;1–16.
- [40] Bo Chunqiang, Hu Hong, Lei Guobin, Liu Ze, Shao Junhao. Non-destructive testing of airfoil based on infrared lock-in thermography. In: 2018 IEEE international conference on information and automation (ICIA). IEEE; 2018, p. 1623–8.
- [41] Wang Qiang, Hu Qiuping, Qiu Jinxing, Pei Cuixiang, Li Xinyi, Zhou Hongbin, et al. Image enhancement method for laser infrared thermography defect detection in aviation composites. *Opt Eng* 2019;58(10):103104.