
Development of artificial neural network-based object detection algorithms for low-cost hardware devices



UNIVERSIDAD DE MÁLAGA

PhD THESIS

José Jesús de Benito Picazo


**Programa de Doctorado en Tecnologías Informáticas
Departamento de Lenguajes y Ciencias de La Computación
Escuela Técnica Superior de Ingeniería Informática
Universidad de Málaga**

June 2021



UNIVERSIDAD
DE MÁLAGA

AUTOR: Jose Jesús de Benito Picazo

 <https://orcid.org/0000-0001-9015-5804>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es



Development of artificial neural network-based object detection algorithms for low-cost hardware devices

*A thesis submitted in fulfillment for the degree of Doctor of Philosophy
presented by*

José Jesús de Benito Picazo

Directed by

**Enrique Domínguez Merino
Esteban José Palomo Ferrer**

**Programa de Doctorado en Tecnologías Informáticas
Departamento de Lenguajes y Ciencias de La Computación
Escuela Técnica Superior de Ingeniería Informática
Universidad de Málaga**

June 2021



UNIVERSIDAD
DE MÁLAGA



DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña JOSE JESÚS DE BENITO PICAZO

Estudiante del programa de doctorado TECNOLOGÍAS INFORMÁTICAS de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK-BASED OBJECT DETECTION ALGORITHMS FOR LOW-COST HARDWARE DEVICES

Realizada bajo la tutorización de EZEQUIEL LÓPEZ RUBIO y dirección de ENRIQUE DOMÍNGUEZ MERINO Y ESTEBAN JOSÉ PALOMO FERRER (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 30 de JUNIO de 2021

Fdo.: JOSE JESÚS DE BENITO PICAZO





UNIVERSIDAD
DE MÁLAGA



UNIVERSIDAD
DE MÁLAGA

**AUTORIZACIÓN PARA LA LECTURA E INFORME SOBRE LA TESIS DE D.
JOSE JESÚS DE BENITO PICAZO**

Ezequiel López Rubio, Catedrático de Universidad de Ciencia de la Computación e Inteligencia Artificial de la Universidad de Málaga, en calidad de tutor de la tesis doctoral de D. Jose Jesús de Benito Picazo, titulada **Development of artificial neural network-based object detection algorithms for low-cost hardware devices**; y Enrique Domínguez Merino y Esteban José Palomo Ferrer, profesores titulares del departamento de Lenguajes y Ciencias de la Computación de la Universidad de Málaga, en calidad de directores de dicha tesis, AUTORIZAN su lectura.

Asimismo, Ezequiel López Rubio, Enrique Domínguez Merino y Esteban José Palomo Ferrer, en calidad de tutor y directores de dicha tesis, INFORMAN que las publicaciones que avalan la tesis no han sido utilizadas en tesis anteriores.

Málaga, 30 de Junio de 2021.

Fdo.: Ezequiel López Rubio

Fdo.: Enrique Domínguez Merino

UNIVERSIDAD
DE MÁLAGA

Fdo.: Esteban José Palomo Ferrer





UNIVERSIDAD
DE MÁLAGA

*To my family,
for their support through
this wonderful journey.*



UNIVERSIDAD
DE MÁLAGA

Acknowledgements

*Let us be grateful to people who make us
happy; they are the charming gardeners
who make our souls blossom.*

Marcel Proust

I would like to start this acknowledgements section by saying thank you to Enrique and Esteban, my thesis' directors. This thesis would've never been a reality without their help and guidance. I would also like to thank Ezequiel for sharing his advice and wisdom with all of us. He was always ready to help in every situation, whether it was about a professional or personal issue. He always tended a friendly hand, and what he does every day as the head of our team is seriously appreciated.

Of course, I would like to thank my ICAI research team collaborators Rafa, Juanmi and especially my lab. mates: Karl (A.K.A. "Doctor"), Miguel Ángel, Jorge and Rosa. Their extreme professionalism and expertise make them a phenomenal team to work with, and their amazing personality makes me want to go to work every day with a smile.

I want to thank my father and mother for their absolute affection and support, for sharing my achievements and failures and for teaching me the most important lessons in life. I would also like to thank my younger brother, Fernando, for encouraging me to do things that I thought were beyond my capabilities. His talent and determination have set an excellent example for me to follow in my academic career and life.

Finally, I want to thank my dear Trini for her joy and warmth, for cheering me up when I needed it and for bringing adventure and fun to my life. But above all, I want to thank her for her unconditional love and understanding through all this time.

Thank you all for believing in me.



UNIVERSIDAD
DE MÁLAGA

Abstract

Brevity is the soul of wit.

William Shakespeare

The human brain is the most complex, powerful and versatile learning machine ever known. Consequently, many scientists of various disciplines are fascinated by its structures and information processing methods. Due to the quality and quantity of the information extracted from the sense of sight, image is one of the main information channels used by humans. However, the massive amount of video footage generated nowadays makes it difficult to process those data fast enough manually. Thus, computer vision systems represent a fundamental tool in the extraction of information from digital images, besides a major challenge for scientists and engineers.

This PhD Thesis' primary objective is automatic foreground object detection and classification through digital image analysis, using Artificial Neural Network-based techniques, specifically designed and optimised to be deployed in low-cost hardware devices. This objective will be complemented by developing individuals' movement estimation methods by using unsupervised learning and Artificial Neural Network-based models.

The cited objectives have been addressed through a research work illustrated in a series of four publications that have been selected to support this thesis. The first one was published in the *Integrated Computer-Aided Engineering* journal in 2018 and consists of a neural network-based movement detection system for Pan-Tilt-Zoom cameras deployed in a Raspberry Pi board. The second one was published in the *International Joint Conference on Neural Networks* in 2018 and consists of a deep learning-based automatic video surveillance system for PTZ cameras deployed in low-cost hardware. The third one was published in the *Integrated Computer-Aided Engineering* journal in 2020 and consists of an anomalous foreground object detection and classification system for panoramic cameras, based on deep learning and supported by low-cost hardware. Finally, the fourth work was published in the *International Joint Conference on Neural Networks* in 2020 and consisted of an individuals' position estimation algorithm based on a novel neural network model for environments with forbidden regions, named

Forbidden Regions Growing Neural Gas.

The results achieved by the author in these publications attest to the work of four years of research that gets summarised in this PhD Thesis memorandum document.

Contents

Acknowledgements	xi
Abstract	xiii
List of Abbreviations	xvii
List of Figures	xxi
Overview	1
1 Introduction	3
1.1 Context	3
1.2 Objectives	6
1.3 Methodology	8
1.4 Structure of this thesis	9
I Fundamentals and State of the Art	11
2 Fundamentals and State of the Art of Object Detection	13
2.1 Fundamentals on object detection and classification	14
2.1.1 Haar-like features	15
2.1.2 Histogram of Oriented Gradients (HOG)	16
2.1.3 Deformable Parts Models (DPM)	18
2.1.4 Object detection based on Artificial Neural Networks .	20
2.1.5 Deep Learning	29
2.2 Deep Learning-based video surveillance systems	35
2.3 PTZ camera based video surveillance systems	41
2.4 360° Panoramic camera-based video surveillance systems . . .	46
2.5 Microcontrollers, microcomputers and low-cost hardware-based surveillance systems.	50

II	Research work	57
3	Motion detection with low cost hardware for PTZ cameras	59
4	Deep learning-based anomalous object detection system powered by microcontroller for PTZ cameras	61
5	Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras	63
6	Image Clustering Using a Growing Neural Gas with Forbidden Regions	65
	Conclusion	67
7	Conclusion and future research lines	69
	7.1 Conclusion	69
	7.2 Future research lines	73
	7.3 Final reflection	78
	Appendices	81
A	Publications Summary	83
	A.1 Works supporting this PhD Thesis	83
	A.2 Additional Publications	86
B	Resumen en Español	89
	B.1 Introducción	90
	B.2 Estado del arte	92
	B.3 Trabajos de investigación que apoyan esta tesis	93
	B.4 Conclusiones y trabajo futuro	100
	B.4.1 Conclusiones	100
	B.4.2 Trabajo Futuro	103
	Bibliography	107

List of Abbreviations

- AI** Artificial Intelligence.
- ANN** Artificial Neural Network.
- AUC** Area Under the ROC Curve.
- BCSAE** Brain Connection based on Stacked Autoencoder.
- BING** Binarised Normed Gradients.
- CCTV** Closed-Circuit Television.
- CNN** Convolutional Neural Network.
- ConvNet** Convolutional Neural Network.
- CORE** Computing Research and Education Association of Australasia.
- CPU** Central Processing Unit.
- CSC** Convolutional Sparse Coding.
- CSCDRN** Convolutional Sparse Coding-based deep random vector functional link Network.
- CSS** Cloud Shadow Speed.
- DBI** Davies-Bouldin Index.
- DL** Deep Learning.
- DNN** Deep Neural Network.
- DPM** Deformable Parts Models.
- DQN** Deep-Q Network.
- DRBM** Deep Restricted Boltzmann Machine.
- DRN** Deep Random Network.

- EEG** Electroencephalography.
- ELM** Extreme Learning Machine.
- EPI** Epipolar Plane Image.
- ESSL** Energy-Saving Street Lighting.
- FBN** Functional Brain Network.
- FBP** Filtered Back Projection.
- FOV** Field Of View.
- FPGA** Field Programmable Gate Array.
- FRGNG** Forbidden Regions Growing Neural Gas.
- FRGNG3D** Forbidden Regions Growing Neural Gas 3D.
- FRSOFM** Forbidden Region Self Organising Feature Map.
- FRSOM** Forbidden Region Self Organising Feature Map.
- GCS** Growing Cells Structures.
- GG** Growing Grid.
- GGG** GII-GRIN-SCIE.
- GHNG** Growing Hierarchical Neural Gas.
- GNG** Growing Neural Gas.
- GPU** Graphics Processing Unit.
- GRMLP** Recurrent Multilayer Perceptron.
- HMM** Hidden Markov Model.
- HOG** Histogram of Oriented Gradients.
- IA** Inteligencia Artificial.
- ICAE** Integrated Computer-Aided Engineering.
- iEEG** Intracranial Electroencephalography.
- IGG** Incremental Growing Grid.
- IJCNN** International Joint Conference on Neural Networks.
- JCR** Journal Citation Report.

- KNN** K-Nearest Neighbours.
- LFW** Labeled Faces in the Wild.
- LOTS** Low-power Omni-directional Tracking System.
- LRN** Local Response Normalisation.
- MB-LBP** Multi-block Local Binary Patterns.
- ML** Machine Learning.
- MLP** Multilayer Perceptron.
- MSE** Mean Squared Error.
- MTP** Multimodal Temporal Panorama.
- NDC** Neural Dynamics Classification.
- NG** Neural Gas.
- NN** Neural Network.
- ODVS** Omni-directional Vision Sensor.
- PCB** Printed Board Circuit.
- PSA** Panoramic Scene Analysis.
- PSRI** position, scale, and rotation invariant.
- PTZ** Pan-Tilt-Zoom.
- PVI** Panoramic View Image.
- QCC** Quasi-Connected Components.
- RaspiCam** Camera specifically designed for Raspberry Pi.
- RBM** Restricted Boltzmann Machine.
- ReLU** Rectified Linear Unit.
- RES** Spanish Supercomputing Network.
- ResNet** Residual Network.
- RGB** Red Green Blue.
- RNN** Recurrent Neural Network.
- SMCNN** Stacked Multicolumn Convolutional Neural Network.

SNR Signal-To-Noise-Ratio.

SOC System-On-Chip.

SOM Self-Organising Map.

SVM Support Vector Machine.

TWH Thresholding With Histeresis.

UAV Unmanned Aerial Vehicle.

USAFIT United States Air Force Institute of Technology.

VOC Visual Object Classes.

VT Vision Tape.

WCCI IEEE World Congress on Computational Intelligence.

X-ray CT X-ray Computed Tomography.

List of Figures

1.1	Different threshold parameters that could affect the quality of possible segmentation processes (photo by Alexander Gardner, 1863).	4
1.2	Potential hardware architectures to be used in the deployment of artificial neural networks-based computer vision algorithms. Left, high performance Titan X NVIDIA card with a GPU. Right, Raspberry Pi 2 Model B microcomputer.	5
2.1	Objects detected in different image areas (Redmon et al. (2016))	14
2.2	Basic Haar feature kernels representation	15
2.3	HOG classification method. Left: Original image. Right: R-HOG descriptor weighted by respectively the positive and the negative SVM weights (Dalal and Triggs (2005)).	17
2.4	Cascade HOG classification method (Zhu et al. (2006)).	17
2.5	Results of two-step detection performed by the combined Haar and HOG feature-based detector in urban traffic (Wei et al. (2019)).	18
2.6	Results achieved by a DPM object detector with a person model (Felzenszwalb et al. (2008)).	19
2.7	Operation of the reconfigurable tree-shaped DPM detector presented in Lin et al. (2015).	19
2.8	Biologically inspired artificial neural networks (Meng et al. (2020)).	20
2.9	A taxonomy of neural network architectures (Gardner and Dorling (1998)).	21
2.10	Structure of a multilayer perceptron with two hidden layers. (Gardner and Dorling (1998)).	21
2.11	Structure of the system proposed by Arriola and Carrasco (1990): a) Main block diagram of the system; b) MLP basic structure; c) HMM example.	22
2.12	Right: Original gray-level fingerprint; left: Feature points extracted from thinned prints (Leung et al. (1991)).	23

2.13	Overall framework of H-ELM, presenting a multilayer forward encoding followed by the original ELM-based regression (Tang et al. (2016)).	24
2.14	Overview of a Self-Organising Map. Image source: http://www.cis.hut.fi	25
2.15	IGG typical operating (Blackmore and Miikkulainen (1993)).	26
2.16	Different dimensionality neural structures. k . (a) $k = 1$, (b) $k = 2$, (c) $k = 3$ (Fritzke (1994)).	27
2.17	Comparison between the adaptation of the NG (Martinetz and Schulten (1991)) to a probability distribution and the adaptation of the GNG (Fritzke (1995b)) to the same probability distribution.	28
2.18	GHNG Structure (Palomo and Lopez-Rubio (2017)).	29
2.19	Growing Grid network (Fritzke (1995a)). It starts with a 2×2 topology. Subsequently, full new rows and columns are being added while the network neurons number doesn't exceed 100. The last plot (down-right) represents the network with 102 units.	29
2.20	Examples of the final positions for the FRSOFM neuron prototypes, considering artificial data sets in Ramos et al. (2019)	30
2.21	Operating of a ConvNet for image classification (Lecun et al. (2015))	31
2.22	Left: Concept image of the electric activity of a brain in a normal operation and a brain in seizure; right: Detail of CNN structure used in Acharya et al. (2018)	32
2.23	Overview of the CNN-RF method proposed by Ansari et al. (2019)	32
2.24	Architecture of a Restricted Boltzmann Machine (Rafiei et al. (2017))	33
2.25	Architecture of the integrated DBM-SoftMax and DBM-BPNN model illustrated in Rafiei and Adeli (2018)	34
2.26	An example of X-ray CT reconstructions. First column on the left is the ground truth coming from an FBP reconstruction using 1,000 views. The rest of the columns are reconstructions from just 50 views using FBP, a regularized reconstruction, and from a CNN-based approach. The CNN-based reconstruction preserves more of the texture present in the ground truth and results in a significant increase in SNR (Jin et al. (2017)).	34
2.27	Graphical representation of combined denoising and classification architectures (Koziarski and Cyganek (2017)).	35
2.28	Illustration of rendered 3D virtual pavement surface (Zhang et al. (2019)).	36

2.29	Picture of the deep learning based inspection approach described in Liang (2019).	37
2.30	Depiction of the crack detection method using the encoder–decoder network proposed in Bang et al. (2019).	37
2.31	An overview of the CSCDRN developed in Maeda et al. (2019).	38
2.32	Operating of the system described Luo et al. (2019).	38
2.33	Structure of the convolutional neural networks with regional parallel structure described in Wang and Bai (2018).	39
2.34	The pipeline of Pairwise Filter Layer described in Liu et al. (2016). A pair of heterogeneous images are fed into the layer. After filtered by the learned pairwise filters, the two feature maps are summarised into the similarity map.	40
2.35	The structure of the SMCNN proposed in Shen et al. (2019).	41
2.36	Density map generation example in Shen et al. (2019). Left, original image. Centre, density map generated by the system in Shen et al. (2019). Right, density map superimposed on the original image.	41
2.37	Image of a PTZ camera.	42
2.38	Saliency detection performed by the model described in Chen et al. (2016).	42
2.39	Example of the foreground subtraction performed in Ferone and Maddalena (2014). (a) Original frame, (b) ground truth, (c) moving object detection masks, (d) representation of the neural background model.	43
2.40	Foreground-Background segmentation obtained by different models of the state of the art and the method proposed in Allebosch et al. (2019).	44
2.41	Parallel lane markings and their vanishing point used by the PTZ calibration system presented in Song and Tai (2006).	44
2.42	Various architectures for networks exploiting PTZ cameras. Micheloni et al. (2010).	45
2.43	Dynamic camera control images. Blue regions mark the Field Of View (FOV); darker regions identify camera overlap in the system developed in Ding et al. (2012)	45
2.44	Left, Structure of the Deep-Q Network utilised in the work by Kim et al. (2019); right, target of the PTZ camera control.	46
2.45	Omni directional object tracking system presented in Boulton et al. (2004).	47
2.46	System described in Gandhi and Trivedi (2004). (a) Omni-directional Vision Sensor (ODVS). (b) A typical image from an ODVS. (c) Transformation to a perspective plan view.	48
2.47	Pedestrian tracking by the system presented in Scotti et al. (2005).	48

2.48	MTP representation used in the work by Wang and Zhu (2012).	49
2.49	Left, image of a Ladybug5+ 360° panoramic camera; right, 360° frame captured by the Ladybug5+.	50
2.50	(a) Underwater drone. (b) Convolutional Neural Network (c) Fish classes observed and identified by the system developed in Meng et al. (2018).	50
2.51	Moving object detection embedded in the video encoding phase performed in Tong et al. (2014).	51
2.52	Left, Schematic representation of the system proposed in Angelov et al. (2017); right, real life experiment object detection results with the Angelov et al. (2017) proposal.	52
2.53	Left, smart camera composed by a Raspberry Pi and a RaspiCam used in Dziri et al. (2016); right, tracking pipeline powered by low-cost hardware developed in Dziri et al. (2016) proposal.	52
2.54	Left, image of the Vision Tape in curved configuration; right, examples of integration of the VT onto different substrates (Dobrzynski et al. (2012)).	53
2.55	(a) Cloud Shadow Speed (CSS) presenting a weather-proof enclosure. (b) Simplified schematic of the luminance sensor. 3.1 Luminance (c) Sensor arrangement presenting the luminance sensors distribution through the enclosure. Fung et al. (2014).	54
2.56	Motion detection examples for a pedestrian video with the SOM-based video surveillance system developed in Ortega-Zamorano et al. (2016).	54

Overview



UNIVERSIDAD
DE MÁLAGA

Chapter 1

Introduction

*The only way of discovering the limits of
the possible is to venture a little way past
them into the impossible.*

Arthur C. Clarke

ABSTRACT: This chapter presents a brief introduction to this PhD Thesis, where the reader can find the various circumstances that motivated its development, the objectives pursued by the author and the methodologies used for the research process. At the end of this chapter, the structure of this thesis is also summarised.

1.1 Context

The human brain is the most powerful, complete and versatile learning device known so far. As a consequence, scientists of various disciplines are fascinated by its structures and processing mechanics. Computer science is one of these disciplines, as it emulates the human brain to design some of the most powerful machine learning methods and algorithms.

Due to how the human brain processes the information supplied by the sense of sight and the massive amount of information it can extract from it, image is one of the leading information channels used by human beings. This fact has led modern society to develop methods for storing and interpreting cited images through increasingly advanced techniques. Nowadays, the improvement and lowering costs of those devices have promoted that any person at any moment can obtain and store images in a fast and inexpensive way. This fact has contributed to significant advances in many fields, such as industrial processes control, identity verification systems, security, leisure

and entertainment. However, the vast amount of visual information generated makes it extremely difficult for a human operator to extract features from those images at such speed that this process is effective. These reasons postulate image processing through computer vision algorithms as an essential tool for the correct operation of those systems, and at the same time, an extremely challenging task for scientists and engineers.

In order for an automatic system to extract information from an image representing a scene of any environment, first, this image must be captured using any image capture device and conveniently stored in a digital format, previously codified in an RGB values matrix. Second, the system must have the capability to identify the objects present in the scene and, if these are in motion, register their trajectories, so it helps the system estimate the activities that are taking place. Hence, it will be critical to differentiate the image background from the objects that are in the foreground. For such a task, a process called *segmentation* will be performed. This process consists of applying a series of algorithms that will eventually yield two different sets of pixels: One set corresponding to each of the foreground objects in the image and another set corresponding to the image background.

However, those algorithms' capabilities for detecting foreground objects are not 100% effective because of their intrinsic limitations. These limitations are caused by several factors such as the models' parameter calibration, the models' adaptation to different kinds of images and the variability in the supplied images' quality. Indeed, these images might have been captured in unfavourable environments or using low-quality cameras, which may result in some inconvenience, such as high noise levels or insufficient resolutions.

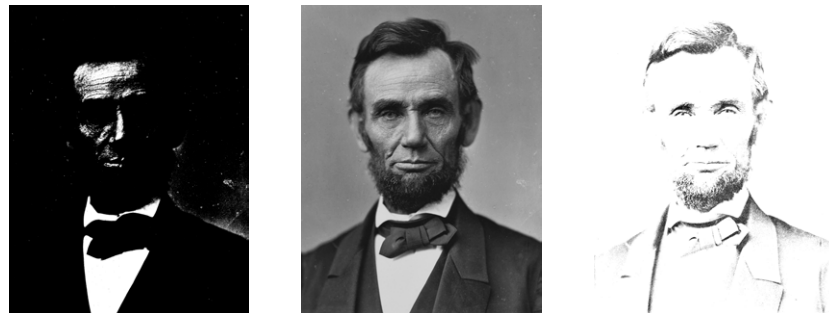


Figure 1.1: Different threshold parameters that could affect the quality of possible segmentation processes (photo by Alexander Gardner, 1863).

Because of these reasons, in the last years has arisen a new generation of algorithms capable of performing object segmentation in an image, identifying and localising them in a fast way with acceptable accuracy. These artificial neural network-based algorithms have reached new levels of speed and accuracy in object detection tasks, thanks to new ways of training these

artificial neural networks and especially to the performance improvements achieved by the adaptation of these Artificial Neural Network-based mathematical models to high-performance hardware devices such as the new NVIDIA GPU-based graphic cards which allow the training of networks with huge amounts of data in a reasonable time.

Despite the large improvement introduced in computer vision processes through these techniques, these tasks also present one issue: the enormous computing power needed to work efficiently. This fact forces them to be deployed on expensive hardware devices with high electric power consumption, hindering their deployment in the cheap and low power consumption hardware that can be often found in portable devices. Thus, it appears the need for designing computer vision-based algorithms involving artificial neural networks, but in such a way that they are optimised to get the best possible performance in low computing power and low-cost hardware systems.



Figure 1.2: Potential hardware architectures to be used in the deployment of artificial neural networks-based computer vision algorithms. Left, high performance Titan X NVIDIA card with a GPU. Right, Raspberry Pi 2 Model B microcomputer.

This way, the power of artificial neural network-based computer vision algorithms would be combined with the accessibility and operating economy from cited low-cost hardware devices, allowing the creation of a new generation of cheap, efficient portable systems with built-in Artificial Intelligence (AI).

Our research's main target is designing and developing object detection algorithms based on artificial neural networks, which can solve the difficulties commented above, heading towards their deployment in low-cost hardware devices.

The first stage of this work is the development of object detection algorithms based on artificial neural networks so that they work properly whatever the origin of the processed images would be, either isolated images or being part of digital video streams shot with a camera that may be either stationary or in motion. In order to accomplish that objective, the various types of neural networks that can be used will be assessed, so we can state which one works better in each case. Moreover, cited algorithms

will be optimised to be deployed in low computational capabilities and low power consuming hardware devices as much as those are the systems they are intended to be deployed in.

In the second stage, due to the notable performance demonstrated by the novel deep learning-based techniques, the possibility of using them will also be studied. Deep Learning (DL) techniques imply the use of certain types of artificial neural networks featuring a significantly high amount of layers, also known as Deep Neural Networks (DNNs). More precisely, in the field of digital image processing, these techniques use to rely upon Convolutional Neural Networks (CNNs). These networks have the ability to detect in each one of the layers of their three-dimensional structure various object features in a hierarchically growing way, attending to the complexity of cited features. Along the same line, part of this work aims to achieve the deployment of these types of networks in low-cost hardware devices with limited computing power as well as low electrical power requirements. Indeed, this last objective will be presented as the most important challenge of this thesis due to the huge amount of computational resources demanded by deep learning-based recognition and identification systems.

Concurrently with the two phases described above, and as it is illustrated in Chapter 2, it is absolutely critical the fulfilment of a constantly evolving survey of the different low-cost hardware platforms, aiming to find the most suitable in terms of a balance between computing power and electric power consumption, for the newly developed algorithms to be deployed. This way, we aspire to develop algorithms adaptable to the highest amount of hardware platforms possible to constitute the foundations for agile engineering systems development, such as AI arrangements for robotics, personal identification, security, video surveillance, and autonomous driving.

1.2 Objectives

The main target pursued in this research is to perform the detection and identification of foreground static or in motion objects in scenes through the analysis of digital images provided by several various types of cameras, using artificial neural network-based techniques specially designed and optimised to be deployed in low-cost hardware devices.

Artificial neural network-based techniques for detecting and identifying foreground objects comprehend a wide field of study, so briefly, this work is going to be oriented towards three specific objectives:

- As pointed out above, one of the main areas of research in this thesis is the construction of low-cost and power-efficient video surveillance systems based on Machine Learning (ML) techniques. A good starting point to design and implement these kinds of systems is the use

of shallow artificial neural networks aimed to implement movement detectors to process the output of video surveillance cameras. Accordingly, this research's first objective will be the construction of video surveillance-oriented movement detectors to be deployed in low-cost hardware devices, with the capability of detecting movement in scenes supplied by a Pan-Tilt-Zoom (PTZ) camera while performing panning, tilting and zoom movements.

- Because of its intrinsic nature, unsupervised learning constitutes the pure essence of the learning task for both living organisms and machines. As a result, unsupervised machine learning is one of the most promising areas in the machine learning field of study, with a remarkable amount of possible applications in science and engineering. One of the most important artificial neural networks-based unsupervised learning systems nowadays is the Self-Organising Map (SOM). Thus, our second objective will be to study and develop new SOM-based neural networks as a powerful tool to be used in machine learning tasks such as object location prediction or data compression. These models will be put to practice in a research work where we will tackle the surveillance of the migratory movements of biological species, so behaviour predictions in these species' migratory habits can be made.
- Deep Learning is the use of deep artificial neural networks applied to machine learning. Because of the development of new programming paradigms applied to extremely powerful hardware devices, deep learning arises as an unstoppable force that has brought a revolution among all the machine learning techniques existent nowadays. However, the large amount of computing power required for handling these neural networks often makes it mandatory to use expensive and high power-consuming hardware devices. This fact makes it difficult to obtain cheap and low power consuming deep learning-based systems. However, there are many situations and environments where, for any reason, it is difficult to deploy high power-consuming expensive hardware systems. In terms of automatic video surveillance, there are many environments where, because of the morphology of the general power network installation or because of the distribution of the different spaces within, installing a high power-consuming device is not suitable. Consequently, it would be very convenient for those environments to have the possibility of mounting some low cost and low power consuming surveillance system, portable, autonomous and with a significant fraction of a deep learning system's accuracy and performance. Therefore, this thesis' last and more important objective is to optimise deep learning techniques so they can be deployed in low cost and low power consuming hardware devices but presenting accuracy and speed

as near as possible to the performance reached by the expensive high power-consuming device-based systems.

1.3 Methodology

The methodology followed to develop the research illustrated in this thesis is dictated by the research's intrinsic nature. Considering the demands of the applications for which foreground object detection algorithms are often intended, cited algorithms must offer the most accurate results possible. Simultaneously, they will be required to have a fast response that allows making decisions on a time scale that frequently will be very near to real-time. Therefore, it is mandatory to set a balance between speed and accuracy in the design of cited algorithms. Moreover, it is important to have mechanisms that allow quantifying the goodness of the obtained results and whether the algorithms end properly in an acceptable time.

These considerations led us to make use of several methodological principles as the foundations of this research, namely the following:

- **Scientific method.** This method states that every project must be subject to the principles of reproducibility and refutability, *id est*, any experiment performed in the framework of this research must be in conditions of being reproduced in a way that equal data produces equal results and considering that any scientific hypothesis may be refuted.
- **Incremental development methodology.** This methodology consists of the development of a functional algorithm that is capable of solving a simple version of the problem that is being dealt with, the complexity of which will be increasing in the subsequent iterations by incorporating new functionalities until it reaches a version that can answer the whole problem.
- **Implementation methodology.** Aiming to ease their development, maintenance and code reusing, algorithms must be implemented following a correct, modular, maintainable and well-documented programming style.
- **Evaluation criteria.** The algorithm goodness measurement must be clearly stated. In order to achieve this objective, error and similarity measurements will be used between the ground truth and the results generated by the algorithms.
- **Comparison of the results with other models found in the state of the art.** This way, we have the certainty that our research presents enough novelty, and its contribution is significant enough.

1.4 Structure of this thesis

This thesis is structured in three different parts: The first one details the background and state of the art that the author has used as a starting point for the research process. The second part gathers the most important works that have been published as a result of the research process. The third and last part of this thesis presents the conclusions obtained from the research process and the possible future works emerging from those in order to continue the development of the different models and systems presented.

Comprising mainly the second chapter of this document, the first part of this thesis presents the theoretical background and state of the art this research is based on. In this chapter, we detail the main concepts and ideas supporting the object detection and classification models, the different types of existing video surveillance systems, with a special mention to those based on neural networks, as well as the known issues for deploying these types of systems in low-cost hardware devices such as microcontrollers and System-On-Chip-based microcomputers (SOCs).

The second part consists of four chapters, one for each work supporting this doctoral thesis. Thus, in Chapter 3 we can find the first contribution presented in this thesis, which consists of a neural network-based movement detection system deployed in a Raspberry Pi for Pan-Tilt-Zoom (PTZ) cameras, that was published in the *Integrated Computer-Aided Engineering (ICAE)* Q1 JCR journal. This article can be framed in the automatic video surveillance systems development and describes the design of an algorithm that is able to detect foreground object movement in a scene taken by a PTZ camera that is performing pan, tilt, or zoom movements at a constant speed by using a new mathematical model that compensates the drift of a PTZ camera in any of its movements in combination with a Multilayer Perceptron (MLP). This algorithm is optimised to be deployed in a Raspberry Pi 3 System-On-Chip microcomputer.

Chapter 4 gathers the second work supporting this thesis. Firstly presented in the 2018 *International Joint Conference on Neural Networks (IJCNN)* in Rio de Janeiro, Brasil, it consists of an automatic video surveillance system powered by low-cost hardware and equipped with a novel detection and classification system for anomalous foreground objects in video streams taken by a PTZ camera. The cited detection and classification system presents an optimised potential detection generator in charge of formulating possible anomalous object locations all over the frame and passing each of those potential detections to a convolutional neural network, which will judge whether an actual anomalous object has been found in the image framed by the cited potential detection. As convolutional neural networks are usually high resource-consuming models, the ones used in this work have been optimised by using some software libraries that allow them to exploit the whole com-

puting power of the limited resources featured by low-cost hardware devices, such as the Raspberry Pi 3 Model B: the hardware device which the system will be deployed in.

Chapter 5 presents the evolution of the above-referenced detection and classification system in a new article published in the Q1 JCR journal *Integrated Computer-Aided Engineering*. This new research illustrates the design and implementation of an anomalous foreground object detection and classification system based on deep learning and powered by low-cost hardware, but in this case, the system will be able to detect and categorise anomalous moving objects in a video stream shot by a panoramic 360° camera. This work consists of an algorithm presenting an architecture that integrates a potential detection generator based on three different statistical models, each one featuring a multivariate homoscedastic distribution and a CNN-based classification module, conveniently optimised to achieve acceptable results when deployed in low-cost and low power demanding microcontroller-based hardware devices.

One of the most important parts of this research is the study of the suitability that different artificial neural network models present to be employed in the construction of automatic surveillance systems. Thus, arises the issue of using artificial neural network models, other than the convolutional ones, to create systems with the ability to predict the possible behaviour of individuals. Regarding this matter, the fourth work in this thesis, illustrated in Chapter 6, explores the possibility to predict the movement of individuals in a certain area with some movement restrictions, through clustering processes. Because of their special nature, some categories of artificial neural networks have demonstrated to be especially efficient when it comes to clustering tasks. We are referring to unsupervised learning networks such as Self-Organising Maps (SOMs), neural gasses and their derivatives. Hence, this work, presented in the *International Joint Conference on Neural Networks*, in the year 2020, describes the building of a Growing Neural Gas-inspired model that keeps its neuron prototypes out from a previously specified set of regions through its training process. This model is named Forbidden Regions Growing Neural Gas (FRGNG) and is intended to perform clustering tasks in data distributions with some regions where no data can be found.

Finally, the last part of this memorandum is dedicated to looking back at the different research works developed in this thesis in order to extract a set of conclusions that summarise the main ideas and results presented in them. Moreover, this chapter will also describe the main future research lines inspired by those works as they will constitute the starting point of new research activities.

Part I

Fundamentals and State of the Art



UNIVERSIDAD
DE MÁLAGA

Chapter 2

Fundamentals and State of the Art of Object Detection

*I was obliged to be industrious. Whoever
is equally industrious will succeed . . .
equally well.*

Johann Sebastian Bach

ABSTRACT: This chapter explains the basic concepts, theory and related work behind this PhD thesis. First, foreground object detection and classification system fundamentals are enumerated, indicating what foreground object detection and classification is. Diverse problems emerging are also addressed, and a complete state of the art, presenting different solutions to these problems, is provided. As this thesis is mainly oriented to artificial neural networks, this section also presents an introduction to artificial neural networks explaining the different types used in this thesis. Spanning from the multilayer perceptron to the new and popular Deep Neural Networks, through the Self-Organising Map-based networks, this section illustrates the different features of each of these models with a significant mention to the machine learning techniques based on them, supported by notable works of some of the most acclaimed experts in the field.

The second part is dedicated to video surveillance fundamentals; more precisely, it is explained the segmentation phase and the existing problems in the phase of background modelling. Besides, a set of algorithms that solve these problems with different success levels is provided. This section also features one classification of different automatic video surveillance systems attending to the main devices used for image and video stream provisioning. Those devices are the PTZ cameras and the 360° panoramic cameras, and this section also mentions their main features as well as the different advantages and disadvantages they present.

The last part of this chapter consists of an introduction to the different kinds of microcontrollers and low-cost hardware-based surveil-

lance systems. This introduction is followed by an explanation of the different issues found in the process of optimisation and deployment of the object detection and classification algorithms in these devices, aiming to get cheap and power-efficient automatic video surveillance systems.

2.1 Fundamentals on object detection and classification

Object detection and classification has been one of the toughest challenges faced by computer vision researchers since the need for artificial entities to understand what they were seeing emerged. This process consists of detecting semantic object instances from a certain class in digital images by acknowledging the features that define the objects of that class (Dasiopoulou et al. (2005)). The first approaches to this task refer back to the '70s of the past century when computer scientists, impelled by biomedical, civil engineering and military industries, started to formulate different methods to automatically detect, localise and classify objects in digital images, considering those methods as critical tasks for automation and inspection procedures (Kazmierczak (1978)). The first approaches to achieve object localisation and classification consisted of using classical computer vision to look into some regions of the image supplied by the camera and trying to find some of the features that allow identifying the presence of a certain object in the area just as it can be seen in Figure 2.1.

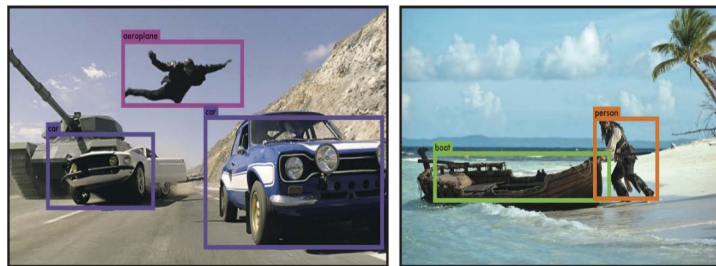


Figure 2.1: Objects detected in different image areas (Redmon et al. (2016))

This is the case of the object detection and localisation performed by the sliding window-based algorithms such as the model presented in Glumov et al. (1995), where the authors describe a procedure for detecting objects employing a sliding window algorithm assisted by classical computer vision techniques and a linear classifier. From this point on, sliding window algorithms became an essential resource when implementing object detection algorithms that were not only capable of addressing whether an object was

in the scene but also were capable of pointing out the position of that object. Consequently, they were improved to be faster and more reliable. We can find an excellent example in work described in Li and Lee (2005), where the authors propose an improved sliding window method employing multi adaptive thresholds to adapt the sliding windows localisation in order to detect cuts in video streams. This method uses possibility values produced from different thresholds to measure the possibility that a cut has been performed in the video file. This system has some interesting applications, such as surveillance cameras forensic inspection.

Sliding-window algorithms continued their evolution and improvement as a technique to perform object detection and localisation by smartly leveraging certain object image features. This evolution resulted in faster and more accurate models, such as the system presented in Cheng et al. (2014), where it is observed that generic objects with well-defined closed boundaries can be detected by looking at the norm of gradients resizing their corresponding image windows to a small fixed size. Based on this fact, this work proposes the resizing of the objects' related windows to an 8×8 size and use the gradients norm as a simple $64D$ feature to describe it for explicitly training an *objectness* measure. This study's authors further demonstrate how the binarised version of this feature named Binarised Normed Gradients (BING), can be used for accurate objectness estimation in a particular image area when improved with a novel fast segmentation method.

2.1.1 Haar-like features

Once the detection and localisation process is performed, it is time to move towards the classification task by considering the features that will enable the system to decide the class of the currently localised object. In order to build accurate and versatile enough systems, it is crucial to develop methods to identify a generic set of features that the designed models can use to figure out the class of the object that has been detected and localised. We can find a good example of these in the Haar-like features. Briefly explained, Haar features are convolutional filters used to detect certain types of lines and areas that are considered promising for performing object detection.

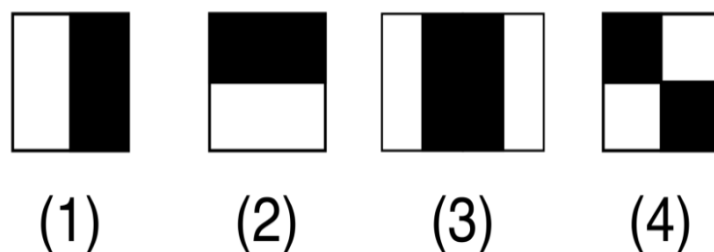


Figure 2.2: Basic Haar feature kernels representation

The four Haar basic feature kernel representations can be seen in Figure 2.2. The first two features are considered as *edge* features, majorly used to detect edges. The third is named *line feature*, and the fourth one is used mainly to detect slanted lines in the image. Of course, there are more sophisticated Haar-like features that allow building more accurate classifiers. We can find an example of this in work developed by the authors in Viola and Jones (2001), where Haar-like features have been used to build high-speed classifiers with the highest accuracy rates at the moment this research was published. More precisely, in this research, three objectives were pursued: The first one was designing a new image representation that allows more efficient processing of the features used in the detector. The second is a learning algorithm based on AdaBoost (Freund and Schapire (1995)) that selects a reduced number of features to implement extremely efficient classifiers. The last contribution of this work corresponds to a method for combining classifiers in cascade, specifically designed to discard background regions of the image so the classifier can focus only on the regions having a higher probability of framing an actual object.

The work by Freund and Schapire (1995) placed the Haar-like features as a powerful method for object detection and classification tasks in images, inspiring researchers who wanted to develop fast classifiers with affordable computing costs. As a consequence, Haar-like features turned into a research field in continuous development by soon bringing new improvements and extensions such as the research performed by Lienhart and Maydt at the Intel Labs that is presented in Lienhart and Maydt (2002). In this work, the authors describe new rotated Haar-like features that improve the number of false positives. The authors also describe optimisation procedures for a given boosted cascade that improves the average false positives. Even though other alternatives to the Haar-like features appeared, such as the Multi-block Local Binary Patterns (MB-LBP) addressed at Zhang et al. (2007), these continued to be one of the most popular and used techniques in object detection and classification, especially in tasks such as face recognition Mita et al. (2005) and vehicle classification Wen et al. (2015) until the hardware technology advances allowed the arrival of Deep Learning.

2.1.2 Histogram of Oriented Gradients (HOG)

Apart from the Haar-like features, some other object detection techniques and classification arose in the last years of the twentieth century. One of the most important is a feature descriptor named Histogram of Oriented Gradients (HOG) (Dalal and Triggs (2005)). Published in 2005, this work describes a method based on evaluating well-normalised local histograms of image gradient orientations in a dense grid. Basically, the idea relies on the fact that local object appearance and shape can be characterised acceptably by distributing local intensity gradients or edge directions, even without

precise knowledge of the corresponding gradient or edge positions. This technique is implemented by dividing the image into small spatial regions or *cells*. Each cell accumulates a local 1-D histogram of gradient directions over the pixels of the cell. Once the image has been tiled with a dense grid of HOG descriptors, the combined feature vector is fed to a classifier based on a Support Vector Machine (SVM) (Figure 2.3) .

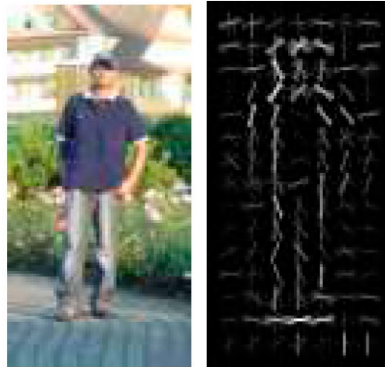


Figure 2.3: HOG classification method. Left: Original image. Right: R-HOG descriptor weighted by respectively the positive and the negative SVM weights (Dalal and Triggs (2005)).

As it would be expected, this type of classifiers also started to evolve fast as more researchers started to utilise them by implementing new cascade approaches such as in Zhu et al. (2006). In this work, Mitsubishi corporation researchers integrate the cascade-of-rejectors approach with histograms of oriented gradients features to build human detection systems with a high level of accuracy and fast delivery. In this case, the considered features are HOGs of variable size blocks of the image, finding an appropriate set of blocks from a more extensive set of image blocks using AdaBoost as feature selector (Figure 2.4).

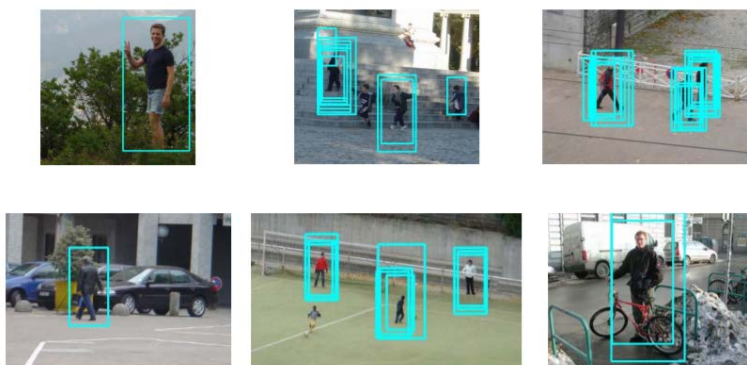


Figure 2.4: Cascade HOG classification method (Zhu et al. (2006)).

Along the same lines, Déniz et al. (2011) presents a HOG-based face recognition algorithm with improved capabilities on compensating errors caused by occlusions, pose and illumination changes. It features the fusion of HOG descriptors at different scales, allowing to capture important structures for face recognition and identifying the necessity for dimensionality reduction in order to make the process less prone to overfitting.

Moreover, HOG features-based classification algorithms have been evolving through the years until nowadays, reaching faster and more powerful yet complex models by combining HOG features and other generic features such as the above referenced Haar-like features. An interesting example of this technique can be found in the recent work by the team of Dr Yun Wei presented in Wei et al. (2019). In this research, the authors propose a two-step detection algorithm based on combining the Haar features and HOG to develop a tracking and detection system to be used in multi-vehicle targets swarming in complex urban environments.



Figure 2.5: Results of two-step detection performed by the combined Haar and HOG feature-based detector in urban traffic (Wei et al. (2019)).

2.1.3 Deformable Parts Models (DPM)

As an improvement to pure HOG-based object detectors, around 2008 emerged a new type of algorithm capable of performing object detection tasks with high levels of accuracy and speed, drawing the attention of the whole computer vision scientific community, hence becoming an integral component of many classification, segmentation, person layout and action recognition tasks. Those are the so-called Deformable Parts Models (DPM), which, briefly explained, work by learning the relationships between objects' HOG features using a latent SVM (Figure 2.6).

One of the first articles where this model can be found is Felzenszwalb et al. (2008). In this work, the authors describe the design of a multi-scaled, discriminatively trained, deformable parts model for object detection that outperforms the results achieved by the best models in the Pascal VOC 2007 challenge in ten out of twenty categories. The system described in this paper combines a margin-sensitive approach for data mining hard negative examples with latent SVMs, leading to non-convex training problems. This model was developed and improved by the same authors over two years, res-

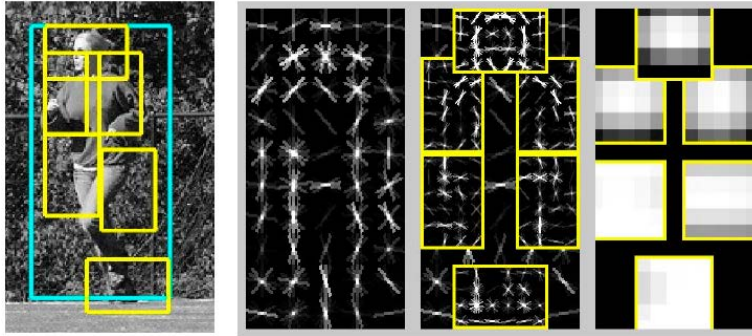


Figure 2.6: Results achieved by a DPM object detector with a person model (Felzenszwalb et al. (2008)).

ulting in a more mature model portrayed in a new article that was published in the Pattern Analysis and Machine Intelligence (PAMI) journal in 2010 (Felzenszwalb et al. (2010)).

DPM-based object detectors continued their evolution, improving their discriminative training algorithms by introducing new adjustments, such as adding different mixtures of different multi-scale deformable part-based models, as it was done in Cheng et al. (2013), or endowing them new reconfigurable graph-shaped structures (Figure 2.7), such as in Lin et al. (2015). These improvements turned DPM-based algorithms into the most sophisticated and powerful object detectors until the arrival of artificial neural network-based object detection and classification systems that ultimately led towards Deep Learning (DL): the revolutionary neural network-based technique that is the most popular nowadays and brought a qualitative jump in machine learning in general, and particularly in object detection processes.

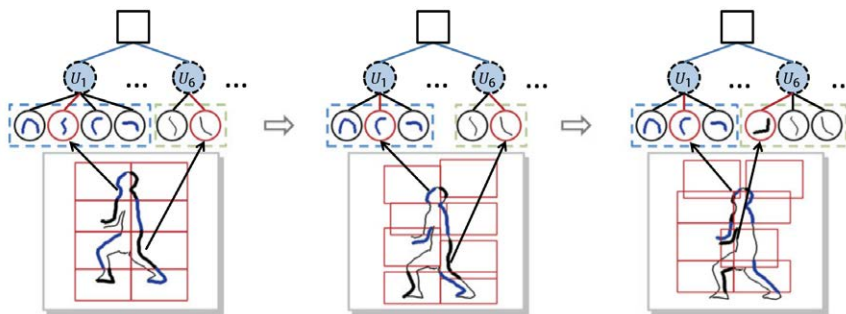


Figure 2.7: Operation of the reconfigurable tree-shaped DPM detector presented in Lin et al. (2015).

2.1.4 Object detection based on Artificial Neural Networks

2.1.4.1 A note on Artificial Neural Networks

Briefly explained, Artificial Neural Network (ANN) are brain-inspired mathematical models intended to replicate the way that humans learn (Figure 2.8).

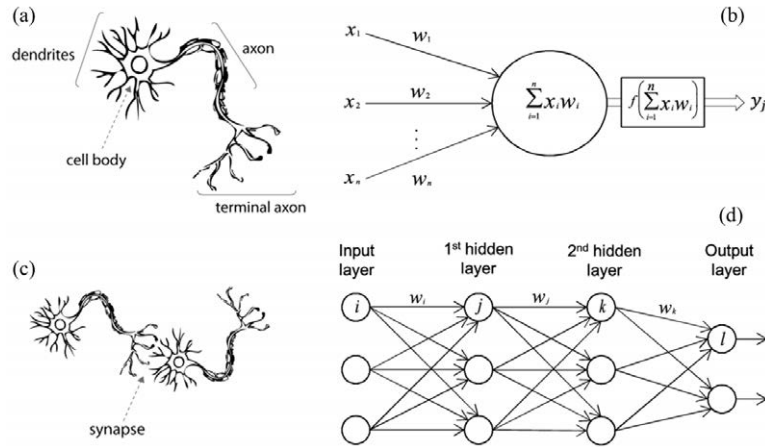


Figure 2.8: Biologically inspired artificial neural networks (Meng et al. (2020)).

As it appears in Schmidhuber (2015), a standard Neural Network (NN) consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Some neurons, defined as input neurons, can be activated through the action of any sensor capable of perceiving the world conditions. Some others can be activated by means of the activation signals supplied through the weighted connections with neurons belonging to a previous layer. The learning process consists of finding weights that make the NN exhibit desired behaviour through a training process that would vary these weights according to some specific criterion.

Once again, depending on their specific structure, topology and operating modes, there are different kinds of artificial neural networks. Because of their critical relation with the research presented in this thesis, we will focus on three different ones: The Multilayer Perceptron (MLP), the Self-Organising Map (SOM) and the Convolutional Neural Network (CNN).

2.1.4.2 The Multilayer perceptron

Artificial neural networks are one of the categories artificial intelligence is divided on. Considering the taxonomy of artificial neural networks appearing in Figure 2.9, multilayer perceptron constitutes one of the main types of artificial neural networks and, as illustrated in Figure 2.10, it consists of a

system of simple neurons, or nodes, connected by weights and output signals which are a function of the sum of the inputs of each node modified by a nonlinear transfer, or activation, function (Gardner and Dorling (1998)).

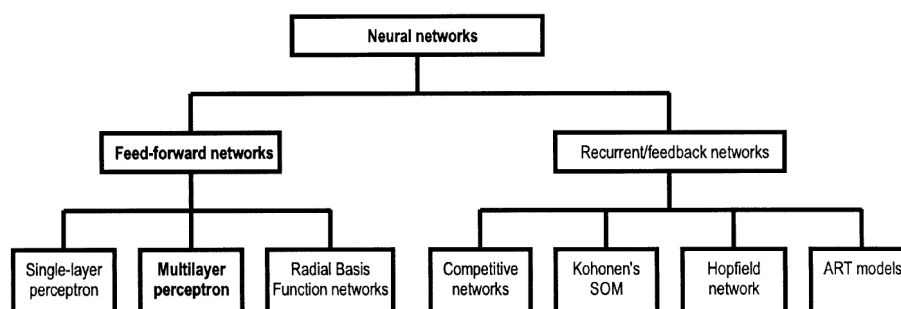


Figure 2.9: A taxonomy of neural network architectures (Gardner and Dorling (1998)).

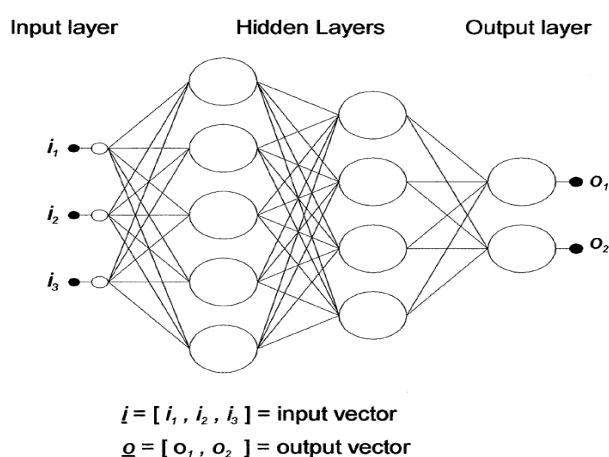


Figure 2.10: Structure of a multilayer perceptron with two hidden layers. (Gardner and Dorling (1998)).

The architecture of a multilayer perceptron is variable, but in general, it will consist of an input layer, an output layer and several intermediate layers, namely *hidden layers*. This type of neural network is described as fully connected, which means that each neuron of one layer is connected to all the neurons of the following layer. Multilayer perceptrons can learn through training in a supervised way, and by selecting a suitable set of connecting weights and transfer functions, they are capable of approximating any smooth and measurable function between the input layer vectors and the output layer vectors (Hornik et al. (1989)).

The multilayer perceptron is one of the most successful and used artificial neural network models over the years because of its versatility, ease of use

and the vast documentation existent referred to it.

After the formalisation of the multilayer perceptron model in the 1960s, was in the 1980s decade when this kind of model started to be considered in engineering as a suitable candidate to be applied in some tasks considered as critical in artificial intelligence and machine learning, such as the speech synthesis, speech recognition and object classification in images.

In the earliest related works like the paper from McCulloch et al. (1988), the authors describe the MLP and discuss its application in essential areas of speech recognition technology such as speech synthesis and vowel recognition. In work by Arriola and Carrasco (1990), authors present the implementation of a new speech recognition system based on the integration of an acoustic processor, a multilayer perceptron that maps the acoustic feature sequences to phonemes and a Hidden Markov Model (HMM), which produces a final identification of the entire utterance as a consequence of the computations of the probabilistic phonetic observations output by the MLP. This structure can be observed in Figure 2.11.

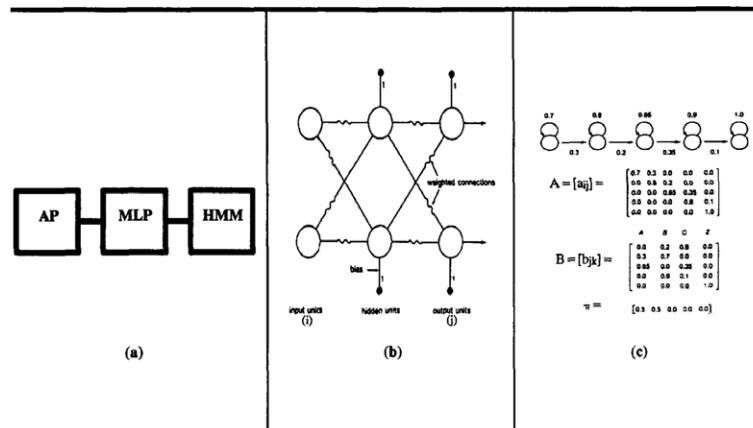


Figure 2.11: Structure of the system proposed by Arriola and Carrasco (1990): a) Main block diagram of the system; b) MLP basic structure; c) HMM example.

Image classification was another field where multilayer perceptrons were considered a promising tool in both the civil and military industry back in the 1980s and at the beginning of the 1990s decades of the past century. We can find an excellent approach in work developed by Troxel et al. (1988), researchers of the United States Air Force Institute of Technology (USAFIT), where a multilayer perceptron is used to perform classification of multifunction laser radar data of tanks and lorries that could be considered as possible tactical targets in combat. The classification is performed by comparing the feature space vector of the processed image to the feature vector of several stored templates representing the different classes. Cited feature vectors are

obtained from transforming the template, and the processed Doppler images into the position, scale, and rotation invariant (PSRI) feature space. According to the authors, the performed experiments' accuracy is near 100%, attesting to the effectiveness of early multilayer perceptron-based image classification systems.

PSRI feature space was widely used also in the 1990s decade to perform image classification just as it is documented in Leung et al. (1991), where a multilayer perceptron-based fingerprint classifier is implemented by determining the existence and position of the fingerprints' minutiae¹ (Figure 2.12)



Figure 2.12: Right: Original gray-level fingerprint; left: Feature points extracted from thinned prints (Leung et al. (1991)).

Simulation results illustrate good detection ratios and low failure rate, hence constituting a reliable method for a system with a small set of fingerprint data.

Various works presented above illustrate how multilayer perceptrons were successfully used in the late 80s and early 90s of the past century to perform simplified versions of some tasks we are today tackling using deep learning-based proposals and large amounts of computing power barely imaginable at that time. However, those days the popularity of a new mathematical model, the Support Vector Machine (SVM), was about to increase to the detriment of the multilayer perceptron in such tasks, exhibiting better performances at lower prices in terms of computing power, leading to the scientific community to abandon the neural networks-based image and speech classifiers and generators until the arrival of deep learning.

Nevertheless, there are some tasks multilayer perceptrons have always

¹The most prevalent system at that time for automated fingerprint classification was the "Minutiae-Coordinate" model.

been very good at. One of them is their use as universal function approximators where they act as compelling classifier algorithms as it is presented in Atkinson and Tatnall (1997).

As it can be easily expected, even though they were not the most popular models for image classification tasks, multilayer perceptrons have continued their evolution over the years by integrating more complex systems that, propelled by the scientists' inexhaustible creativity, can be the basis for the application of new learning algorithms. Moreover, despite the upswing of deep learning-based machine learning algorithms, multilayer perceptrons are still subject to new advances in machine learning, with new MLP-based models appearing every day, continually growing in both complexity and power. This is the case of the work developed in Tang et al. (2016) where a new ELM-based hierarchical learning framework is proposed for multilayer perceptron. The Extreme Learning Machine (ELM) is an emerging learning algorithm for the generalised single hidden layer feedforward neural networks where the hidden node's parameters are randomly generated, and the output weights are computed analytically. In this work, the authors propose a two component-based architecture, namely H-ELM, consisting of self-taught feature extraction, followed by supervised feature classification bridged by randomly initialised weights as it appears in Figure 2.13.

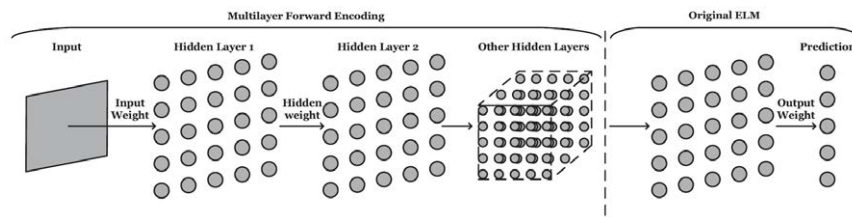


Figure 2.13: Overall framework of H-ELM, presenting a multilayer forward encoding followed by the original ELM-based regression (Tang et al. (2016)).

According to this work, once the previous layer is established, the current layer's weights are fixed without fine-tuning. Therefore, this algorithm reaches higher learning efficiency than deep learning-based models. Experimental results show that the approach presented in this work achieves better and faster convergence than the hierarchical learning methods appearing in the state of the art.

2.1.4.3 Self-Organising Map-based networks

Among the mathematical models applied to machine learning and artificial intelligence, we can find certain networks whose learning algorithms search for previously undetected patterns in a dataset where no labels are previously supplied and where human intervention is reduced to the minimum. We name these *unsupervised learning models*, and Kohonen's Self-Organising

Map (SOM) is one of the most popular ones. Self-Organising Maps are particularly useful for representing data distributions by describing the numerical and topological relations over the different clusters. Originally proposed by Teuvo Kohonen in Kohonen (1990), the Self-Organising Map consists of a sheet-like artificial neural network, the cells of which become specifically tuned to various input signal patterns or classes of patterns through an unsupervised learning process. In the basic version, only one cell or local group of cells at a time gives the active response to the current input. Each cell or local cell group acts as a separate decoder for that input (Figure 2.14). Thus, it is the presence or absence of an active response at that location, and not so much the exact input-output signal transformation or magnitude of the response, that provides an interpretation of the input information. This fact turns SOM into a feasible system to represent input data distributions by defining the relations both topological and statistical among the different data clusters extracted from the cited data distribution, making SOMs useful to generate low dimensional visualisations of high dimensional data.

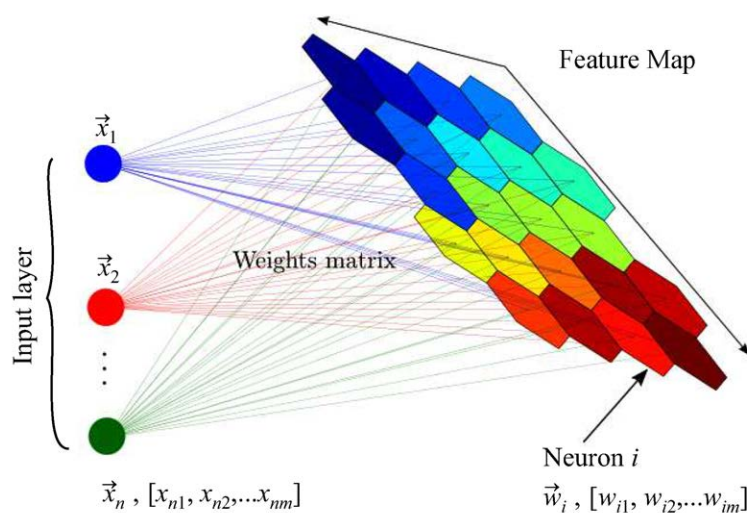


Figure 2.14: Overview of a Self-Organising Map. Image source: <http://www.cis.hut.fi>

As always, after the presentation of a new model, such as the Self-Organising Map, several alternatives and variants emerged, intending to improve the mapping or representation quality. Some of them are the SOM-based neural network models, having the capability of growing as the training process requires it in any way. In general, the training algorithms for these networks start with a relatively small number of neurons that will grow up with the addition of new units. These neurons will be added in specific iterations of the training algorithm until a stop condition is fulfilled. In some models, there are links between the different neurons which are added or

eliminated over the training process according to the value of the parameters of the different neurons belonging to the network, hence conditioning the relation with their neighbours by helping to create a better separation between the generated clusters.

For example, the Incremental Growing Grid (IGG) (Blackmore and Miikkulainen (1993)) is formed initially by four neurons connected in a rectangular grid. During the training process, both the structure and the connectivity of the network is adapted dynamically by adding new neurons in the border of the neuron adjacent to the neuron whose quantification error is maximum, so more space can be provided in the map for the representation of the input data, as illustrated in Figure 2.15.

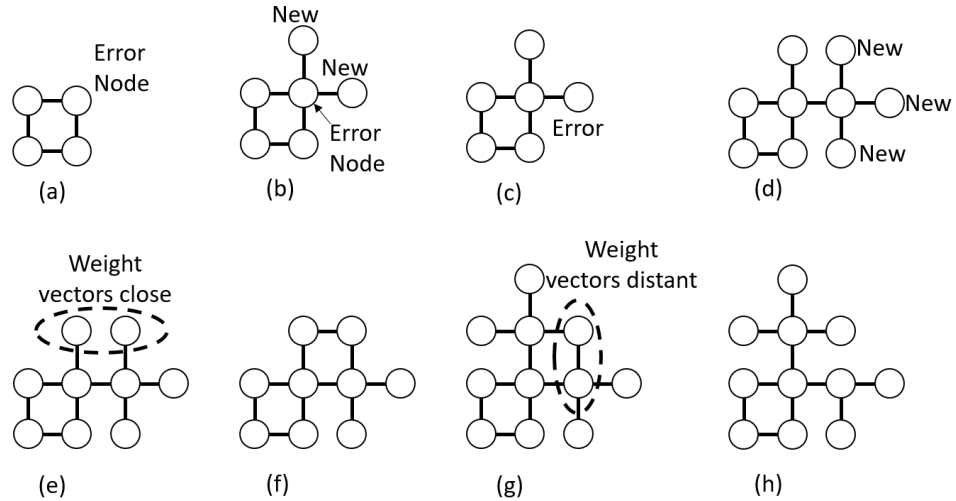


Figure 2.15: IGG typical operating (Blackmore and Miikkulainen (1993)).

Similarly, other types of algorithms increase their neuron number as the training process is completed. This is the case of the Growing Cells Structures (GCS), (Fritzke (1994)), where the number of neurons is increased, and the states of the connections between these neurons keep varying more flexibly in terms of the spatial distribution of the map, as it can be seen in Figure 2.16.

Bernd Fritzke's Growing Neural Gas (GNG) (Fritzke (1995b)) presents a similar algorithm, but its learning rule is different. Moreover, we are talking about an incremental network capable of learning the topological relations in a set of input vectors through a learning rule inspired by Hebb's learning rule. In contrast with other learning algorithms such as the Neural Gas (NG) described in Martinetz and Schulten (1991), this model does not have any parameters that change with time and is capable of continuing its learning process, adding new neurons and connections until a stop criterion is fulfilled. The evolution of the GNG and NG learning algorithms can be

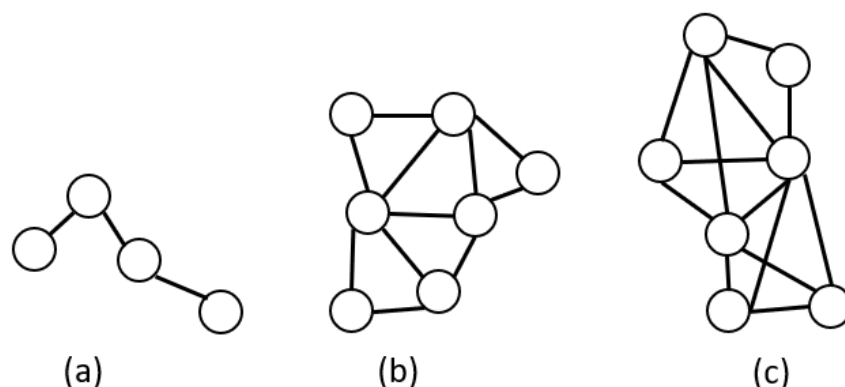


Figure 2.16: Different dimensionality neural structures. k . (a) $k = 1$, (b) $k = 2$, (c) $k = 3$ (Fritzke (1994)).

seen in Figure 2.17. Even though both algorithms seem to get similar results, the main advantage of the GNG over the NG is that certain parameters, such as the final network size, are not meant to be fixed at the beginning of the learning process, resulting in a more powerful and adaptable model whose performance is not constrained by a fixed number of neurons.

The GNG resulted in being a very successful model for clustering tasks, and so did its hierarchical variants, such as the one proposed in Palomo and Lopez-Rubio (2017). In this work, a hierarchical version of the GNG designed to learn a tree of graphs, the Growing Hierarchical Neural Gas (GHNG) (Figure 2.18), is presented, where the original GNG is improved by making a distinction between a growth phase where more neurons are added until no significant improvement in the quantisation error is achieved, and a convergence phase where no unit creation is allowed.

Some growing models based on the SOM add entire rows and columns of neurons to the network instead of adding units one by one. This is the case of the Growing Grid (GG) model, also by Bernd Fritzke (Fritzke (1995a)), where complete rows and columns integrated by neurons are added to the network, keeping a rectangular grid until the training process ends. New neurons are inserted between the neuron having the best score and the neuron presenting the weights vector with the highest difference with the weights vector of the first one, among all its neighbours, whilst the connections between the neurons remain unaltered, as it can be checked in Figure 2.19.

Again, there are also several variants of the method described above. Thus, the work in Bauer and Villmann (1997) illustrates a variant to the above-referenced method, presenting a hypercubic adaptive output space.

However, there are many datasets where we can find some regions in

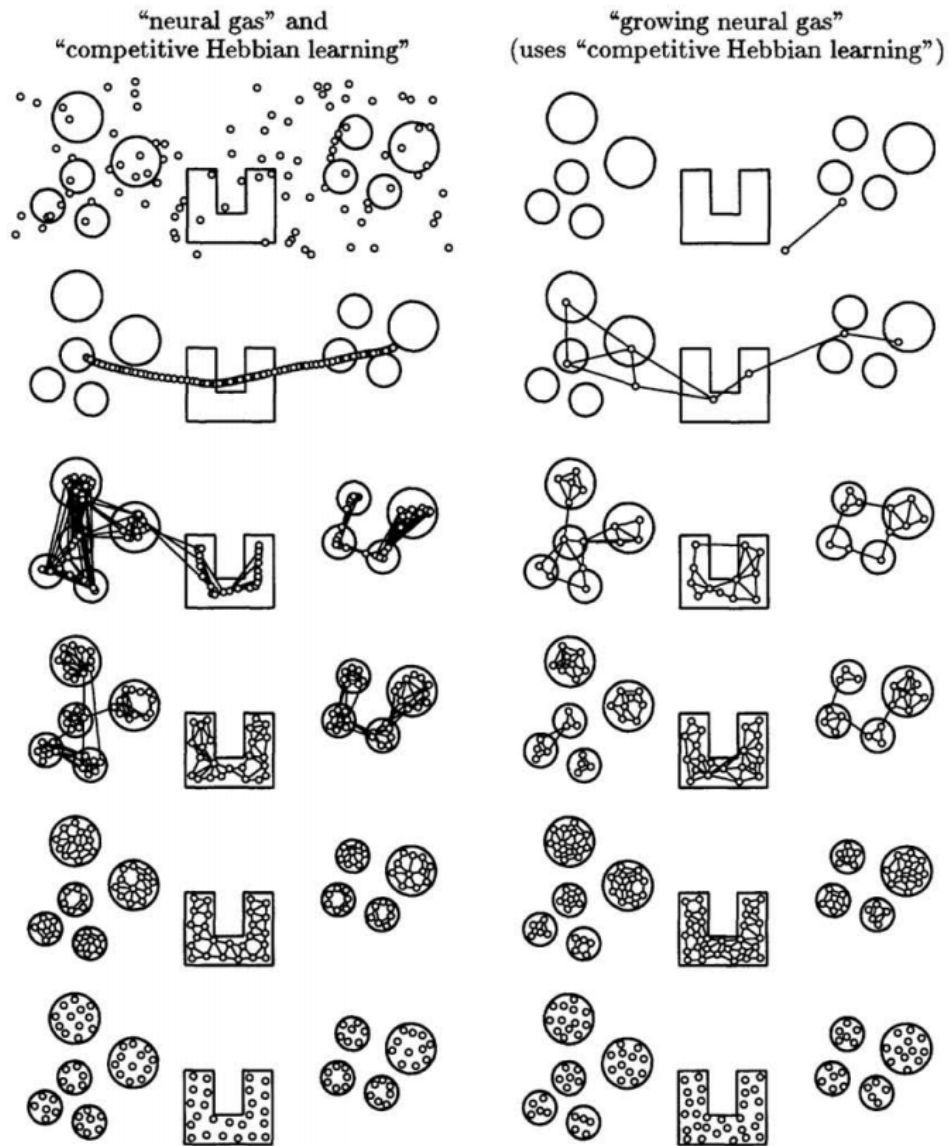


Figure 2.17: Comparison between the adaptation of the NG (Martinetz and Schulten (1991)) to a probability distribution and the adaptation of the GNG (Fritzke (1995b)) to the same probability distribution.

the input vector space where no sample will be observed. Consequently, for the training algorithms, those regions will be excluded to place any of the network neurons. Those will be referred to as *forbidden regions*, and they could be modelling physical barriers that usually will be defined by convex polyhedral sets. Hence, in the most recent state of the art, we can find several algorithms for achieving clustering tasks with forbidden regions. As an

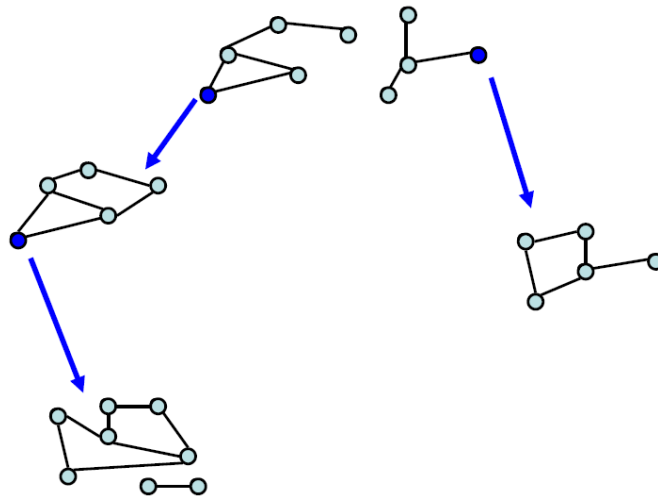


Figure 2.18: GHNG Structure (Palomo and Lopez-Rubio (2017)).

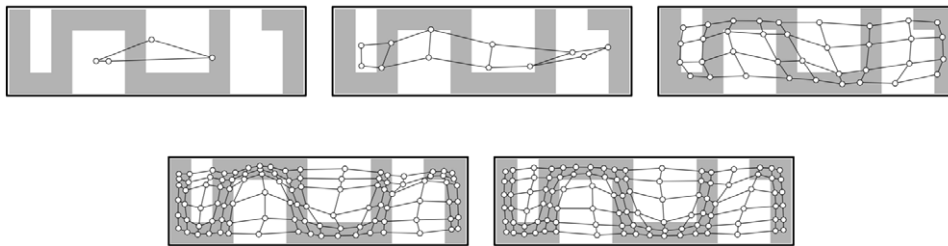


Figure 2.19: Growing Grid network (Fritzke (1995a)). It starts with a 2×2 topology. Subsequently, full new rows and columns are being added while the network neurons number doesn't exceed 100. The last plot (down-right) represents the network with 102 units.

example, we highlight a modification of the SOM model known as the Forbidden Region Self Organising Feature Map (FRSOFM) whose prototypes are able to avoid a set of polyhedral forbidden regions in the plane. This model, presented in Ramos et al. (2019), proposes a new SOM-based model that guarantees that all its neuron prototypes will avoid crossing through any of the mentioned forbidden polyhedral convex regions in their drift through all the considered area as the training algorithm completes its job (Figure 2.20).

2.1.5 Deep Learning

Deep Learning (DL) is one of the most powerful, popular and versatile supervised machine learning and artificial intelligence-oriented techniques. Reaching new levels of accuracy and performance, these methods led to a *renais-*

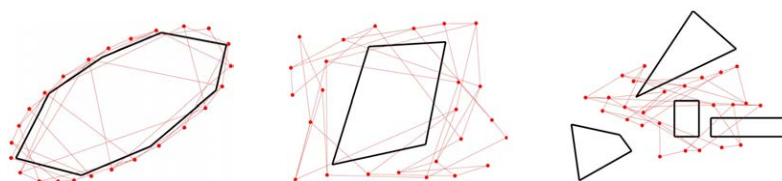


Figure 2.20: Examples of the final positions for the FR-SOFM neuron prototypes, considering artificial data sets in Ramos et al. (2019)

sance of the artificial neural network models in artificial intelligence, almost forsaken around the 1990s decade, when the existing hardware could not satisfy their computational demands.

Before the arrival of DL, many practical applications of machine learning relied on linear classifiers that used hand-engineered feature extractors. Since the 1960s decade of the 20th century, scientists know about the limitations of such linear classifiers as they can only carve the input space into very simple regions such as half-spaces separated by a hyperplane (Duda and Hart (1973)). However, problems such as image recognition or speech processing require the input-output function to be insensitive to variations in the input, such as changes in the position, illumination, orientation of an object, or variations in the intonation, pitch and accent of the speaker; while being very sensitive to other variations that can be very small but very relevant, such as the position of the spots in the skin of a certain animal, or the difference of pronunciation of two different vowels.

A Deep Learning architecture is a multilayer stack of simple modules, all (or most) of which are subject to learning and many of which compute nonlinear input-output mappings (Lecun et al. (2015)). Each module in the stack transforms its input to increase both the selectivity and the representation's invariance. As LeCun says in the cited work, “with multiple nonlinear layers [...], a system can implement extremely intricate functions of its inputs that are simultaneously sensitive to minute details and insensitive to irrelevant variations that can be important, such as the background, pose, lighting and surrounding objects”.

One of the most used neural network models in deep learning and the main one used in the development of the work detailed in this document is the Convolutional Neural Network. Sometimes referred to as *ConvNets* or CNNs, these feed-forward models are designed to process input data supplied in multiple arrays, for example, a colour image composed of three 2D arrays containing pixel intensities in the three colour channels. As it is detailed in Lecun et al. (2015), the architecture of a typical ConvNet is structured as a series of stages (Figure 2.21).

The first few stages are composed of convolutional layers and pooling layers. Roughly speaking, the result of the locally weighted sum of these

layers is then passed through a non-linearity such as a Rectified Linear Unit (ReLU). Hence, a ConvNet network general structure will rely on two or three convolution stages, non-linearity and pooling, stacked, followed by more convolutional and fully-connected layers. Backpropagating gradients through a ConvNet is as simple as through a regular deep network, allowing all the weights in all the filter banks to be trained (Lecun et al. (2015)).

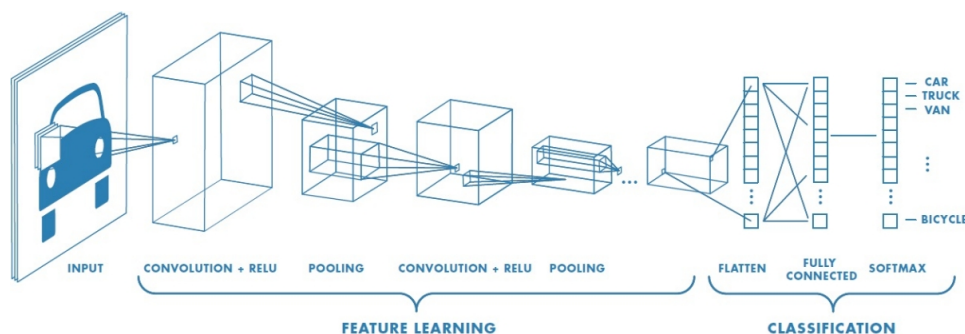


Figure 2.21: Operating of a ConvNet for image classification (Lecun et al. (2015))

Thanks to new hardware devices, whose computing power is in constant improvement, deep learning has been applied successfully to multiple research areas (Liu et al. (2017)). One of the most important research fields that have benefited from deep learning has been medicine. For example, in Antoniadis et al. (2018), an ensemble deep learning architecture for the non-linearly mapping scalp to Intracranial Electroencephalography (iEEG) data is proposed with the objective of exploiting the information from a limited number of joint scalp intracranial recordings to design a novel methodology for detecting the epileptic discharges from the scalp Electroencephalography (EEG) of a general population of subjects, circumventing the unavailability of intracranial EEG and the limitations of scalp EEG. The work presented in Hua et al. (2018) and Hua et al. (2019) illustrates the implementation of a novel semi-data-driven method of computing functional Brain Connection based on Stacked Autoencoder (BCSAE), aiming to detect the proficiency of operators during their mineral grinding process control, based on Functional Brain Network (FBN) models.

Some of the healthcare-oriented applications of deep learning are used to diagnose specific medical conditions. We can find an example of this in the field of neurology, more specifically in the field of epilepsy-caused seizure. Taking into account that an EEG is a commonly used ancillary test to aid in the diagnosis of epilepsy and that the EEG signal contains information about the electrical activity of the brain, we can find several examples of seizure detection based on deep learning algorithms. The first one we would like to present is illustrated in work by Acharya et al. (2018), where a 13-

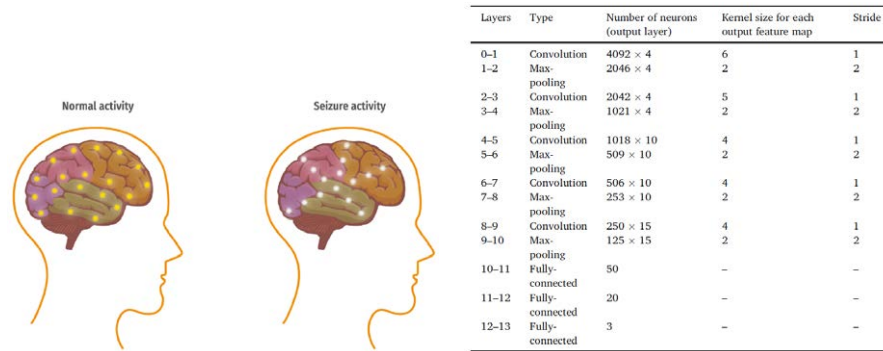


Figure 2.22: Left: Concept image of the electric activity of a brain in a normal operation and a brain in seizure; right: Detail of CNN structure used in Acharya et al. (2018)

layer deep convolutional network is used as the foundation of an algorithm intended to be used in normal, preictal and epilepsy seizure detection from EEGs (Figure 2.22).

Detecting seizure from EEGs has turned into a prevalent use of deep learning in its different modalities. Hence, along the same lines as the work in Acharya et al. (2018), but in a more specific way, we can find papers such as the one describing the work by Ansari et al. (2019), that illustrates the implementation of a seizure detector for neonatal children using raw multichannel EEGs as inputs of a classifier, made from a deep CNN and a random forest (Figure 2.23), that will have “seizure” and “non-seizure” as output classes.

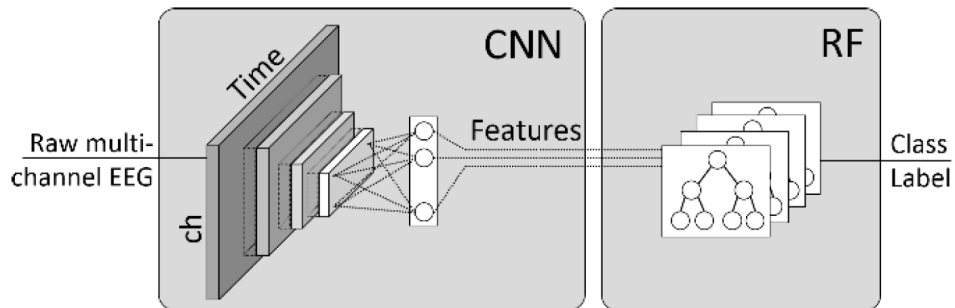


Figure 2.23: Overview of the CNN–RF method proposed by Ansari et al. (2019)

Engineering is another research field where deep learning is commonly used because of the ability this technology presents to create accurate classifiers. There are some neural network structures, namely Restricted Boltzmann Machines (RBMs), which, combined with other RBMs or other types of

structures, conform deep learning-based models that are enjoying high popularity levels in the field of civil engineering. According to Rafiei et al. (2017), a Restricted Boltzmann Machine is an unsupervised learning model with a two-layer interconnected stochastic neural network consisting of a visible input layer and a hidden layer with binary units (Figure 2.24). Indeed, this is the case of the structures employed in works like Rafiei et al. (2017), where an advanced computing model for estimating concrete properties employing a Deep Restricted Boltzmann Machine (DRBM) can be found. Structurally speaking, a Deep Restricted Boltzmann Machine is an auto-encoder consisting of two parts: encoder and decoder. The encoder is a feature extractor made of several layers of RBM intended to simulate the brain's perception of input data. The decoder is symmetric to the encoder and intends to simulate the brain's ability to remember perceptions, reconstruct the inputs, and reduce the dimensionality of data mathematically (Adeli and Wu (1998)). It shows to be effective for the solution of classification and pattern recognition problems such as classification of work descriptions in construction projects, structural impairment detection, structural system identification, and structural defect detection.

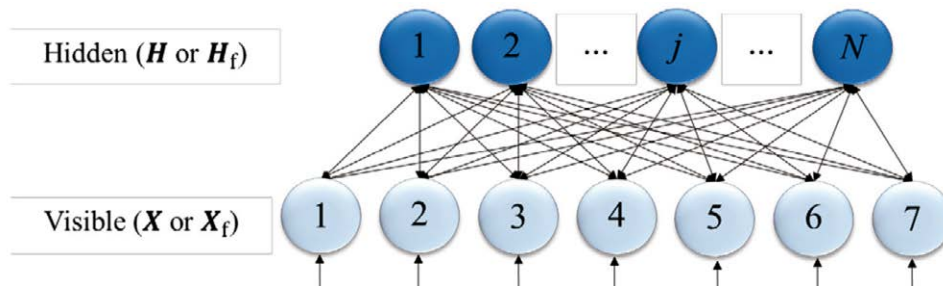


Figure 2.24: Architecture of a Restricted Boltzmann Machine (Rafiei et al. (2017))

We can find similar research by reading the work presented in Rafiei and Adeli (2017), where a new model for detecting damage in high-rise building structures is described, based on a RBM and a Neural Dynamics Classification (NDC) algorithm.

Finally, the work presented in Rafiei and Adeli (2018) presents a cost estimation model for new buildings construction, including advanced deep learning contraptions such as the integration of a deep Boltzmann machine approach with a softmax layer and some regression models (Figure 2.25).

Deep neural network-based algorithms can also be behind several new techniques applied to image quality improvement tasks like image denoising, deconvolution, superresolution and medical reconstruction. The work developed in McCann et al. (2017) illustrates this application in a review on the uses of deep learning for solving such inverse imaging problems (Figure

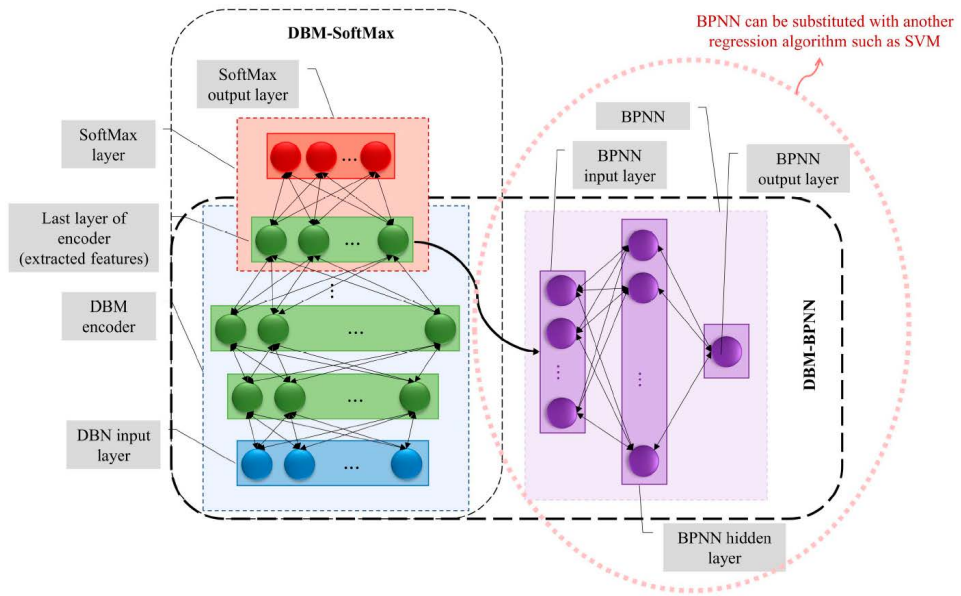


Figure 2.25: Architecture of the integrated DBM-SoftMax and DBM-BPNN model illustrated in Rafiei and Adeli (2018)

2.26).

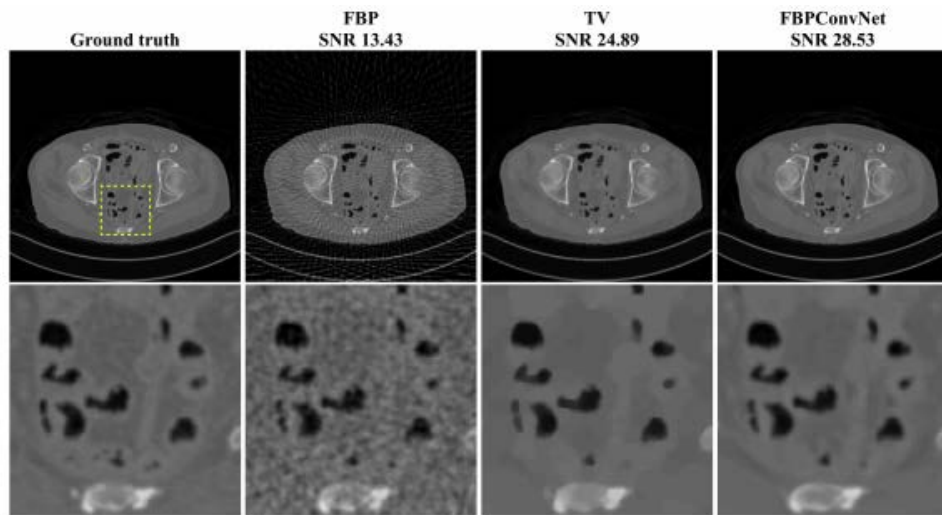


Figure 2.26: An example of X-ray CT reconstructions. First column on the left is the ground truth coming from an FBP reconstruction using 1,000 views. The rest of the columns are reconstructions from just 50 views using FBP, a regularized reconstruction, and from a CNN-based approach. The CNN-based reconstruction preserves more of the texture present in the ground truth and results in a significant increase in SNR (Jin et al. (2017)).

Along the same lines, the work by Koziarski and Cyganek (2017) illustrates the experimental examination of the influence of different types of noise, the proposition and the construction of a deep neural network-based denoising filter, as well as the outlining of a practical method for deep neural network training with noisy patterns for image recognition improvement against noisy test patterns (Figure 2.27).

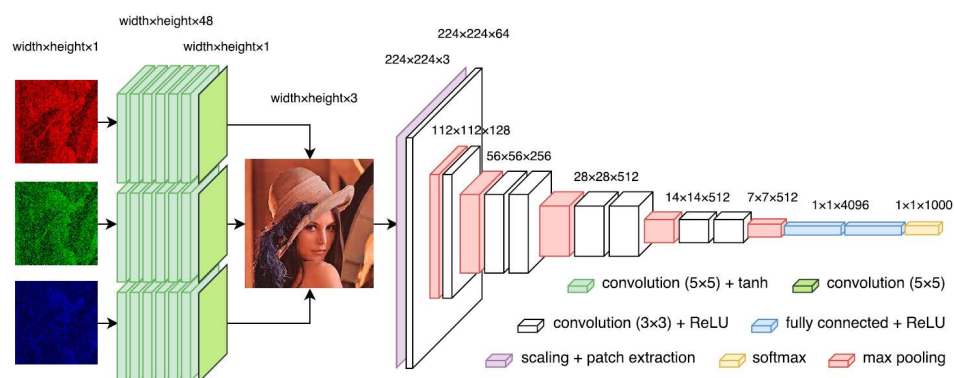


Figure 2.27: Graphical representation of combined denoising and classification architectures (Koziarski and Cyganek (2017)).

Object recognition and image classification stand as one of the most common uses of deep neural networks, especially in video surveillance, where the latest models, empowered by massively trained deep neural networks supported by extremely efficient GPU-based hardware devices, have presented qualitative jump with respect to the preceding object detection and classification techniques. These reasons led the author of this thesis to emphasise the importance, popularity and performance of these systems by dedicating the next section of this document to present some of the most relevant works in state of the art referred to deep learning-based video surveillance systems.

2.2 Deep Learning-based video surveillance systems

In order to offer a neat and rigorous view of deep learning-based automatic video surveillance systems, we have divided their most common uses into three separate parts: general verification of civil engineering structure condition, object detection in video feeds supplied by surveillance cameras and human behaviour monitoring, all of them presenting deep learning-based operating, relying on different kinds of deep neural networks.

Civil structures' condition verification is a relevant application of automatic video surveillance systems since fast and exhaustive damage detection in civil engineering structures is critical for human beings safety. Deep learning-based detection models are also a useful technique in the construction of these types of systems. Therefore, in this field we can find works

like the research presented in Zhang et al. (2019), where a fully automated system for crack detection on three-dimensional asphalt pavement surfaces is described (Figure 2.28). The system foundation consists of a Recurrent Neural Network (RNN), namely *Crack-Net-R* that relies upon a new unit, namely Recurrent Multilayer Perceptron (GRMLP), which is in charge of recursive updating the internal memory of the Crack-Net-R. According to the results of the experiments performed by the authors, Crack-Net-R has approximately a 5% better performance, in terms of Precision and Recall, than its immediate predecessors of state of the art. Besides, it is significantly faster.

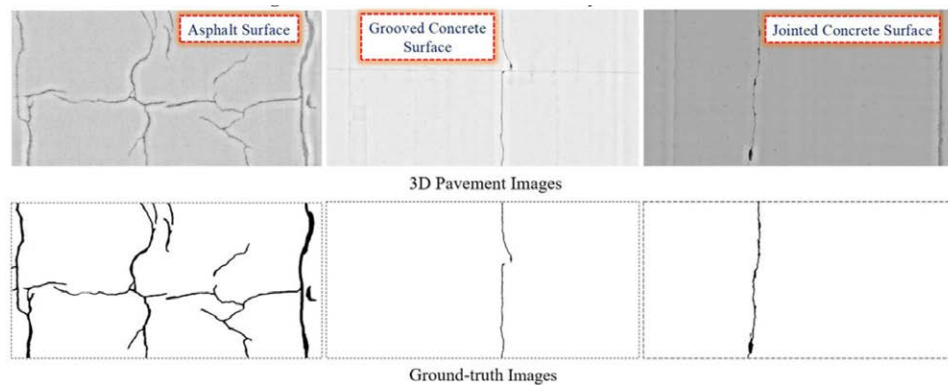


Figure 2.28: Illustration of rendered 3D virtual pavement surface (Zhang et al. (2019)).

Along the same lines of civil engineering condition inspection through computer vision, we can find the research developed by Liang (2019), where the authors describe a three-level image-based approach for post-disaster inspection of the reinforced concrete bridges, using deep learning with novel training strategies. The system uses a convolutional neural network for conducting system-level failure classification, component-level bridge column detection, and local damage-level localisation, plus Bayesian optimisation (Figure 2.29).

Many of the most important types of structures in civil engineering are roads. Indeed, road inspection is a critical task insofar as the security of the passengers of the different vehicles depends on the good state of conservation of cited structures. The work by Bang et al. (2019) exemplifies this by proposing an optimal pixel-level detection method for identifying road cracks in black-box camera images, using a deep convolutional encoder-decoder network in order to perform timely monitoring of the location and severity of the cracks compromising the adequate road condition. The encoder-decoder structure presented by the network described in this system (Figure 2.30) presents an encoder part consisting of convolutional layers of the Residual Network (ResNet) for extracting crack features, and the decoder consists of

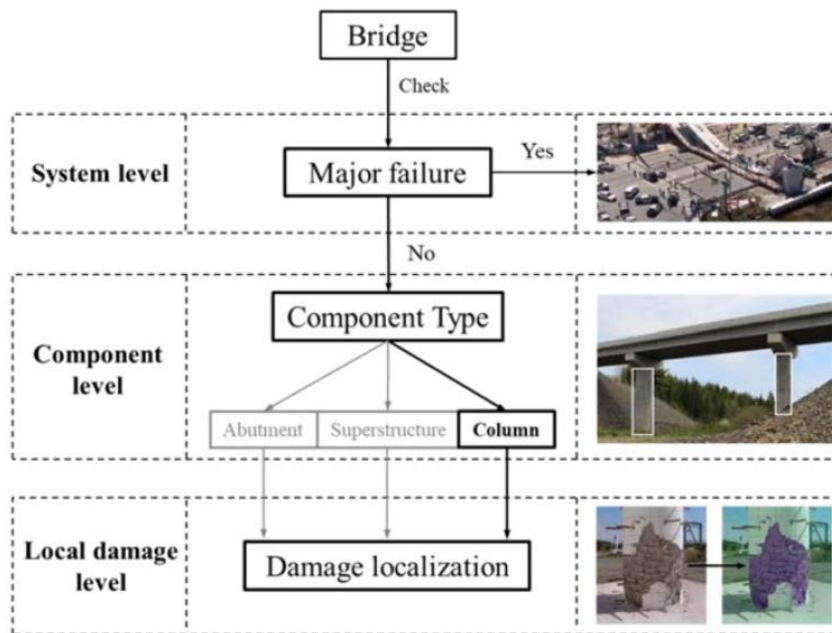


Figure 2.29: Picture of the deep learning based inspection approach described in Liang (2019).

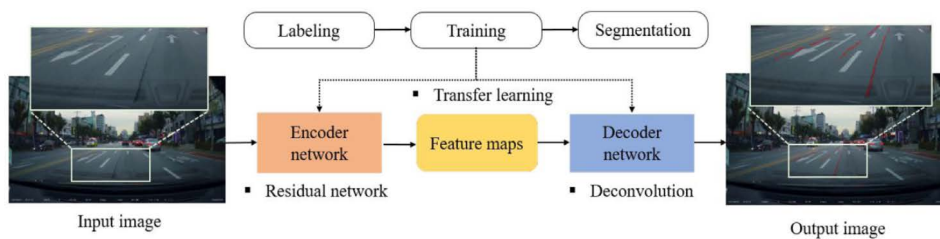


Figure 2.30: Depiction of the crack detection method using the encoder–decoder network proposed in Bang et al. (2019).

deconvolutional layers in charge of localising the pavement cracks in an input image.

Roads' inspection is also the issue dealt with in work by Maeda et al. (2019). This research approaches the problem of distress classification of asphalt through the design of a novel neural network, namely Convolutional Sparse Coding-based deep random vector functional link Network (CSCDRN). It presents a Convolutional Sparse Coding (CSC)-based deep random vector functional link structure that consists of the combination of a CSC-based feature extractor, a pooling layer, a Local Response Normalisation (LRN) layer and a Deep Random Network (DRN) classifier (Figure 2.31). CSC's unique design makes it a model capable of extracting visual features representing the characteristics of target images by training from a

small number of distress images having diverse visual characteristics.

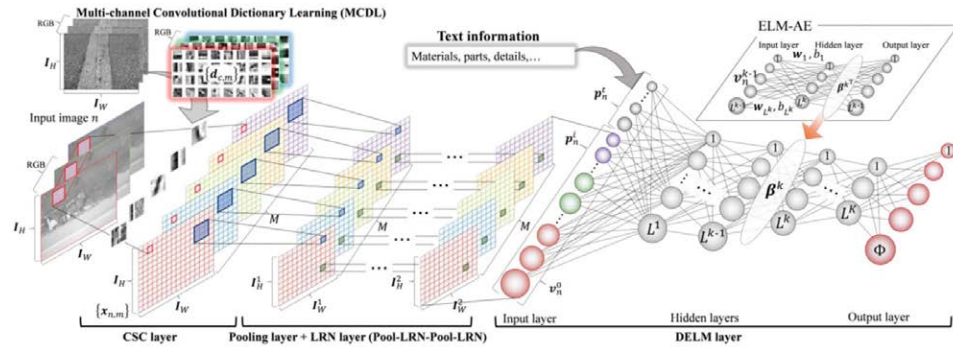


Figure 2.31: An overview of the CSCDRN developed in Maeda et al. (2019).

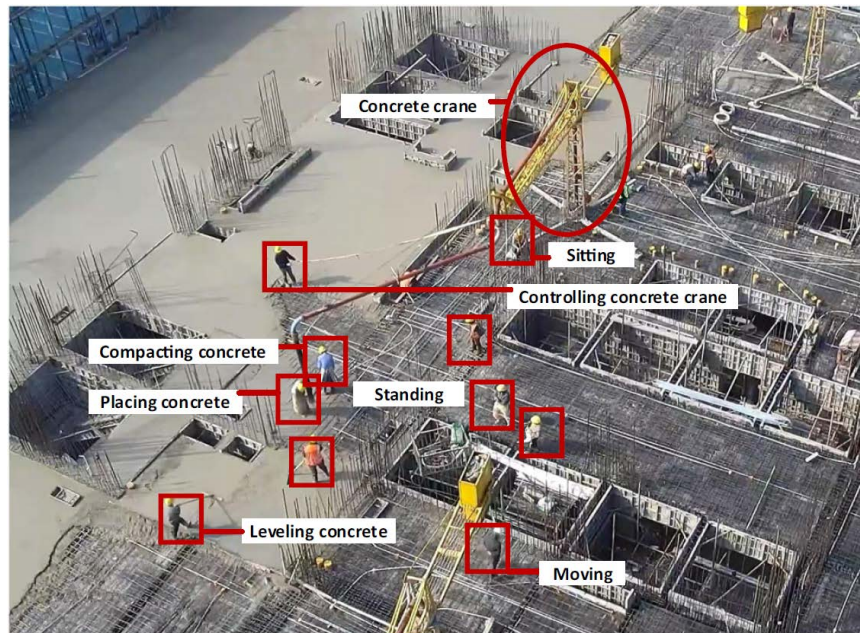


Figure 2.32: Operating of the system described Luo et al. (2019).

When it comes to deep neural network-based human behaviour monitoring, several relevant works have been published. We can find an example in work presented in Luo et al. (2019), where a hierarchical statistical method for recognising the activities performed by workers in far-field videos shot by surveillance cameras is outlined (Figure 2.32). Here, deep learning-based action recognition is used to recognise workers' actions and a Bayesian non-parametric hidden semi-Markov model is employed to infer the workers' activities by processing action sequences.

Face identification is often a critical issue in human behaviour monitoring systems as many of them are intended to trace the actions and the different locations a person can be found at. In this field, several neural network models have been designed. The work described in Schroff et al. (2015) is a good example of these models as it presents a system, called FaceNet, that directly learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity. The work presented in Liu et al. (2016) describes DeepIris: a deep learning framework based on convolutional neural networks, for heterogeneous iris verification, which learns relational features to measure the similarity between pairs of iris images. The operating of this system can be seen in Figure 2.34. Other example can be found at Sun et al. (2016), where the authors present the design of a hybrid convolutional network (ConvNet)-Restricted Boltzmann Machine (RBM) model for face verification that achieves competitive face verification performance on the Labeled Faces in the Wild (LFW) dataset. Along the same lines, Wang and Bai (2018) describes a thermal infrared face identification system relying on a convolutional neural network, where a regional parallel structured CNN algorithm (*RPS net*) is proposed to obtain multi-scale features based on edge information. Finally, the work presented in Perdana and Prahara (2019) describes a light-CNN based on a modified VGG16 model for face recognition with a limited dataset producing good performances with 94.4% accuracy.

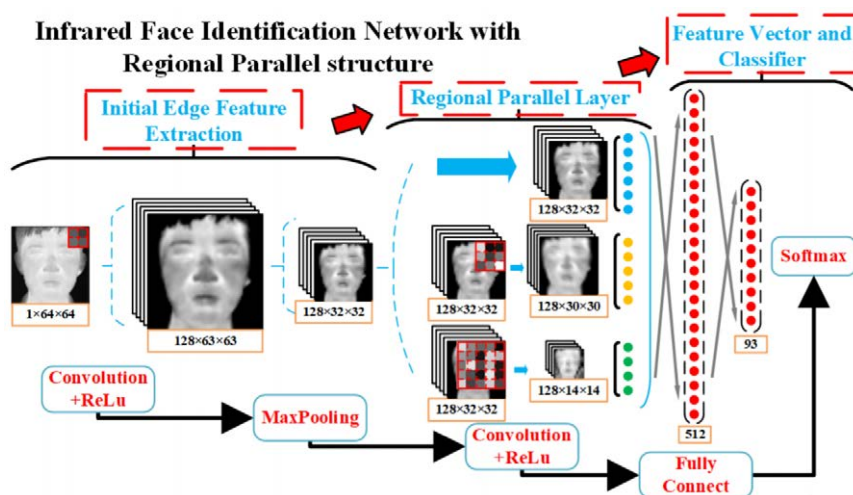


Figure 2.33: Structure of the convolutional neural networks with regional parallel structure described in Wang and Bai (2018).

Deep learning is a powerful tool for face identification systems even when the person identity might be hidden intentionally. This is the case of the system described in Kohli et al. (2018), where the authors utilise deep learning based transfer learning approach for face verification with disguise variations

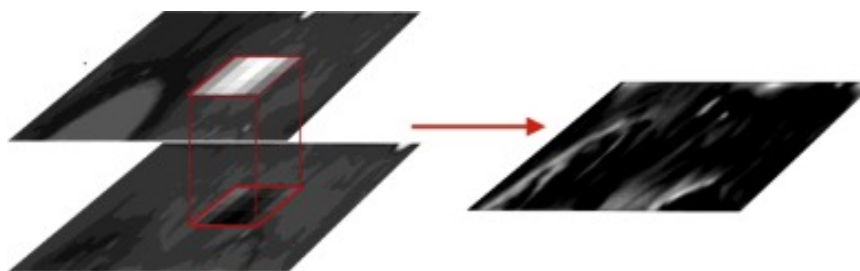


Figure 2.34: The pipeline of Pairwise Filter Layer described in Liu et al. (2016). A pair of heterogeneous images are fed into the layer. After filtered by the learned pairwise filters, the two feature maps are summarised into the similarity map.

by employing Residual Inception network framework for learning inherent face representations.

Pedestrian counting is another important issue when trying to monitor the actions of people in crowded scenarios. This task has several applications such as resource management, transportation engineering, urban design or advertising. That was the motivation for Shen et al. (2019) to develop a CNN-based system for pedestrian counting in crowded scenes. It consists of the use of a novel network model which makes use of a Stacked Multicolumn Convolutional Neural Network (SMCNN), (Figure 2.35) to generate density maps revealing the number of people being present in a particular scene, as it is illustrated in Figure 2.36. In this image we can observe, from left to right, the original picture, the density map generated by the system developed in Shen et al. (2019) and a composition of this density map superimposed on the original picture.

Stimulated by the abundance of social conflicts, public security has emerged as a very important field of study for computer scientists and engineers, apart from a very lucrative business. One of the main facets of human monitoring tasks relies upon video surveillance systems. A critical part of those will depend on the monitoring of video devices capable of encompassing the physical environment. Because of their versatility and ease of installation, PTZ cameras are handy devices when it comes to human behaviour video monitoring tasks. Therefore, the next section is fully dedicated to PTZ camera-based video surveillance systems.

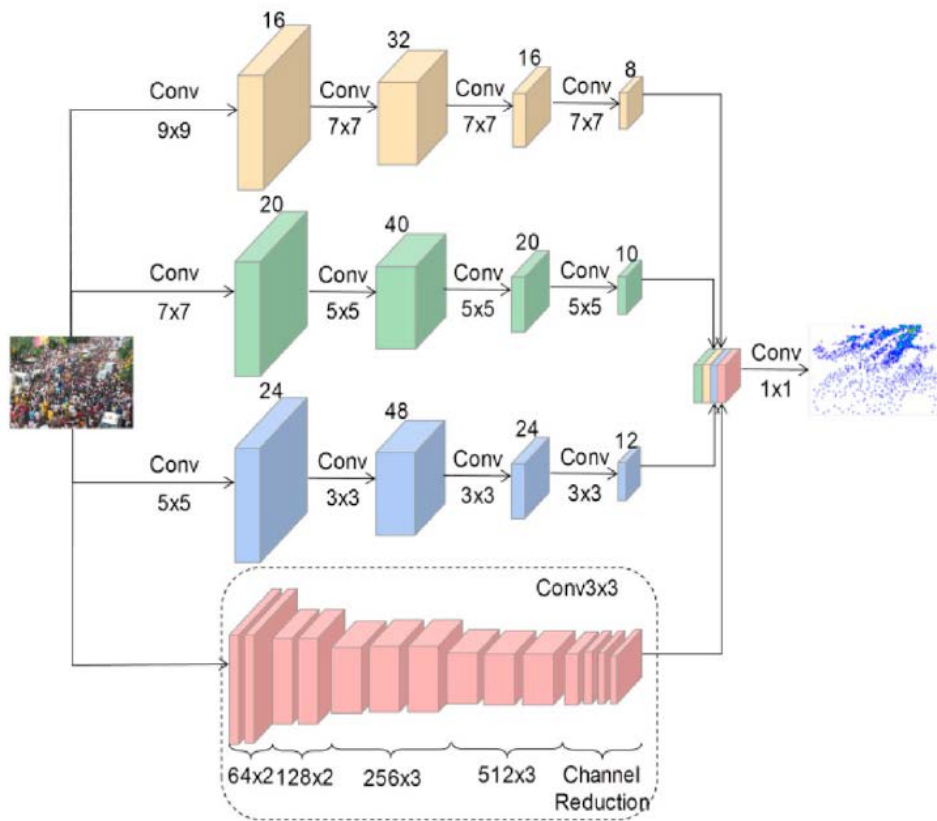


Figure 2.35: The structure of the SMCNN proposed in Shen et al. (2019).

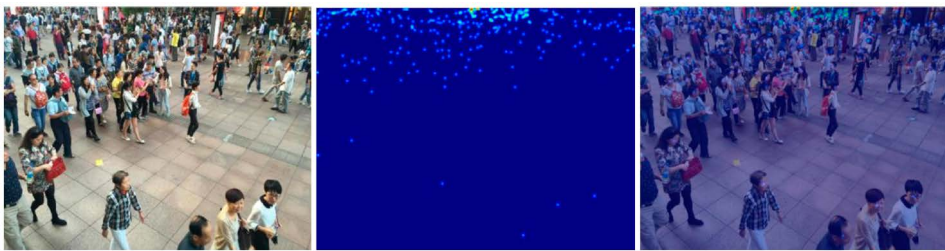


Figure 2.36: Density map generation example in Shen et al. (2019). Left, original image. Centre, density map generated by the system in Shen et al. (2019). Right, density map superimposed on the original image.

2.3 PTZ camera based video surveillance systems

Pan-Tilt-Zoom (PTZ) camera-based foreground object detection, identification and tracking has been a wide field of study in computer vision for many years. A PTZ camera could be defined as a robotised camera which is cap-



Figure 2.37: Image of a PTZ camera.

able of performing movements in the horizontal and vertical planes and is also capable of executing a zoom movement over any scene to obtain a more detailed image of the object to be observed (Figure 2.37).

PTZ camera-based automatic video surveillance systems are the foundation of many sophisticated and accurate human behaviour monitoring tasks. A good example can be found in work by Chen et al. (2016). In it, a novel salient motion detection method for non-stationary videos recorded by hand-held cameras and Pan-Tilt-Zoom cameras is presented (Figure 2.38). This method's main idea is to incorporate the respective advantages of matching and modelling-based methods into a unified low-rank analysis driven tracking-by-detection framework. By the time the paper was released, and according to its authors, this method achieved better results in saliency detection tasks than its competitors in state of the art.



Figure 2.38: Saliency detection performed by the model described in Chen et al. (2016).

One of the main tasks to be accomplished when designing any automatic video surveillance system is an efficient object detector. However, this task requires a good background subtraction algorithm. Concerning this, we find works like the research presented in Komagal and Yogameena (2018), where a survey on the PTZ camera-based foreground segmentation methods is presented. This survey also provides an overview of various techniques from state of the art that address the challenges and solutions of the PTZ

camera-based foreground segmentation methods, besides the existing datasets for experimentation and future possibilities and directions for computer vision researchers in the field of PTZ camera-based background-foreground modelling. We can find an example in work by Ferone and Maddalena (2014). Indeed, this paper presents a neural-based background subtraction approach to moving object detection for image sequences shot employing PTZ cameras (Figure 2.39). The system relies on constructing the sequence background model using a self-organising map in charge of learning image sequence variations, seen as pixel trajectories over time.

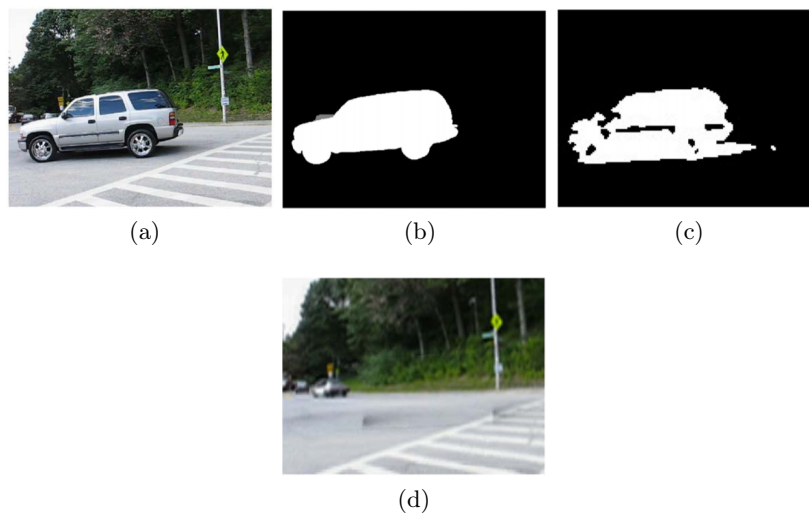


Figure 2.39: Example of the foreground subtraction performed in Ferone and Maddalena (2014). (a) Original frame, (b) ground truth, (c) moving object detection masks, (d) representation of the neural background model.

The research presented in Sajid et al. (2016) aims to a similar target presenting the design of a background subtraction algorithm for PTZ cameras that overcomes this task without the need for explicit image registration, fixing the false detection and long-term results drift affecting traditional background subtraction algorithms based on simple per-pixel models of scene appearance and static cameras. Finally, the work by Allebosch et al. (2019) presents a robust method for compensating the panning and tilting motion of PTZ cameras, applied to foreground-background segmentation. This method works by determining the camera internal parameters through feature-point extraction, in addition to tracking and establishing two motion models for points in the image plane, which will be used at runtime to compensate for the motion of the camera in the background model. The results yielded by this model can be seen in Figure 2.40.

However, not all of the PTZ related works consist of just implementing

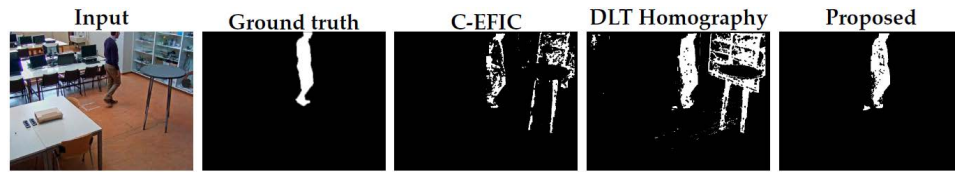


Figure 2.40: Foreground-Background segmentation obtained by different models of the state of the art and the method proposed in Allebosch et al. (2019).

video surveillance systems. It is true that thanks to their flexibility and capabilities for varying the angle of vision, PTZ cameras constitute the basis of many automatic video surveillance systems. However, such systems' performance often depends on the quality of the calibration of the cited PTZ camera. Traditionally, the calibration of these devices has been made manually. Nevertheless, this is an inefficient process that must be done each time the environment circumstances change, being desirable to be done automatically. This is the topic covered by the research described in Song and Tai (2006). This work presents a novel calibration method for PTZ cameras, designated to overlook traffic scenes, that requires no manual operation to select the positions of the special features by automatically using a set of parallel lane markings and the lane width to compute the camera parameters, namely, focal length, tilt angle and pan angle.

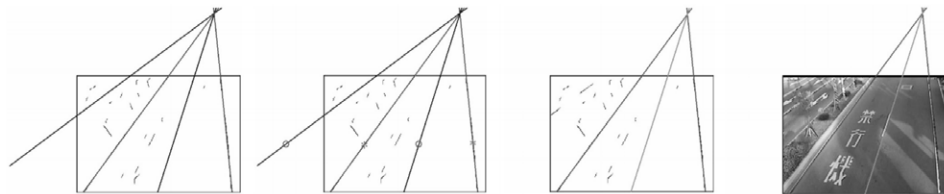


Figure 2.41: Parallel lane markings and their vanishing point used by the PTZ calibration system presented in Song and Tai (2006).

A reader who reached this point can infer that there are many different PTZ camera-based video surveillance systems. For example, some systems just are focused on video surveillance models for PTZ camera networks, such as the one described in Micheloni et al. (2010). In this paper, the authors present a comprehensive introduction to PTZ camera networks to be used in active surveillance, covering low-level techniques for autonomous object detection and more high-level methods for managing the cooperation of different moving cameras. The purpose of such a system is the construction of coordinated PTZ camera networks capable of accurately track moving objects through different environments.

Researchers have approached this topic from different angles. For ex-

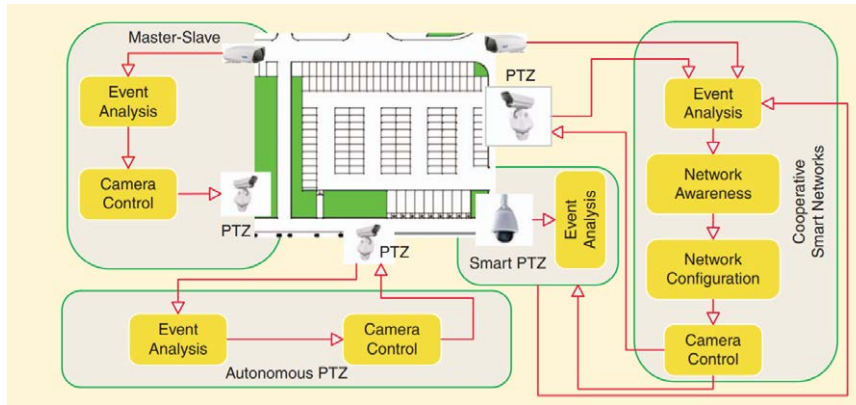


Figure 2.42: Various architectures for networks exploiting PTZ cameras. Micheloni et al. (2010).

ample, in Ding et al. (2012) the construction of coordinated PTZ camera networks is addressed by proposing an integrated analysis and control framework for a PTZ camera network to maximise various scene understanding criteria such as tracking accuracy, best shot and image resolution. This system utilises dynamic camera-to-target assignment and efficient feature acquisition, considering situations where the image processing tasks are distributed over the network in an arrangement where each camera must collaborate with the others to decide how they can be dynamically assigned to targets.



Figure 2.43: Dynamic camera control images. Blue regions mark the FOV; darker regions identify camera overlap in the system developed in Ding et al. (2012)

The above-referenced work was later improved again by the same research team in Ding et al. (2017). In this paper, the authors tackle the problem of automatically controlling the fields of the cameras belonging to a PTZ camera network, leading to improve situation awareness in a region of interest. This system, which brings together computer vision and network control techniques, attempts to observe the entire region of interest at some minimum resolution while opportunistically acquiring high-resolution images of critical events in real-time and deciding to focus on individuals or groups of people by understanding the actions displayed in the images. At the same

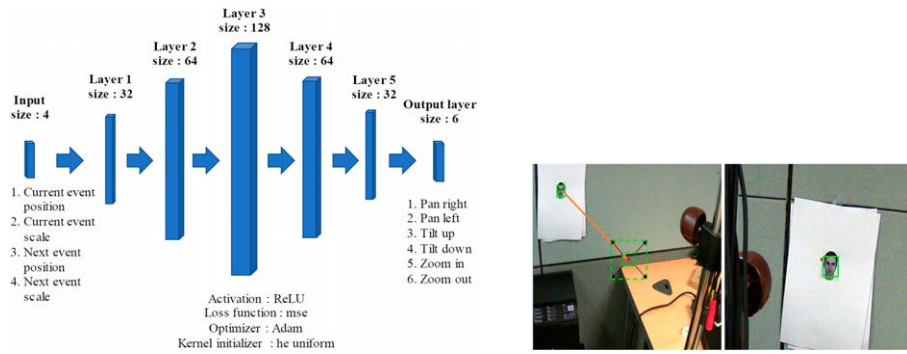


Figure 2.44: Left, Structure of the Deep-Q Network utilised in the work by Kim et al. (2019); right, target of the PTZ camera control.

time, this work proposes camera control strategies to improve a metric that quantifies the quality of the source imagery.

Nowadays, deep learning-based models constitute the basis of the most advanced and accurate automatic video surveillance systems powered by PTZ cameras. We can find a good example of these systems in Dimou et al. (2016), where a combination of deep convolutional networks, heterogeneous training data and data augmentation is explored to improve the detection rate of object detection and tracking tasks in challenging CCTV scenes supplied by PTZ cameras. In other systems, deep learning techniques are used for properly controlling the movements of a PTZ camera to improve object tracking in video surveillance systems. This is the case of the work developed in Kim et al. (2019), where indeed, the authors describe an automatic camera control method based on a Deep-Q Network (DQN) for improving the recognition accuracy of anomalous actions in the video surveillance system (Figure 2.44).

Finally, in the most recent times, we are witnessing an increment of PTZ camera-based video surveillance systems, powered by deep learning techniques, oriented to natural catastrophe detection. This is precisely the topic considered in works like the research presented in Son et al. (2019), where we can find a deep learning-based system with an architecture specially designed to detect floods in certain susceptible areas overlooked by PTZ cameras with two different view angles, obtaining a 95% accuracy in flood classification accuracy.

2.4 360° Panoramic camera-based video surveillance systems

Because of their dynamic capabilities, PTZ cameras are very popular in the implementation of video surveillance systems as they allow the incorpora-

tion of physical movement-based tracking systems into their gear. However, types of devices present some limitations caused by their operating's mechanical nature, such as the fact that they can cover just one section of the environment at a time. This fact motivated us to start searching for some device capable of solving PTZ camera limitations by offering some kind of panoramic 360°, Omni-directional image to be mounted in automatic video surveillance systems. Of course, several researchers have thought the same, and consequently, different models of 360° video cameras have been used in computer vision systems since the last decade of the past century (Yagi (1999)). It is quite common that targets of interest often try to hide from the watcher's eye by blending in with the environment, so they tend to be visible only when they are in motion, reducing the effectiveness of any standard PTZ-based approach. In order to overcome the cited limitations, several systems have been developed. This is the case of the work presented in Boulton et al. (2004), where a Low-power Omni-directional Tracking System system is proposed. This system, also known by its acronym, LOTS, presents a set of Quasi-Connected Components (QCC) that combines gap filling, Thresholding With Hysteresis (TWH) and a novel region merging/cleaning method. The system also incorporates a multi-background modelling system and dynamic thresholding features that are very appropriated for adapting itself to challenging situations such as outdoor tracking in highly populated environments. Omni-directional imaging also reduces issues such as image warping and backprojection.

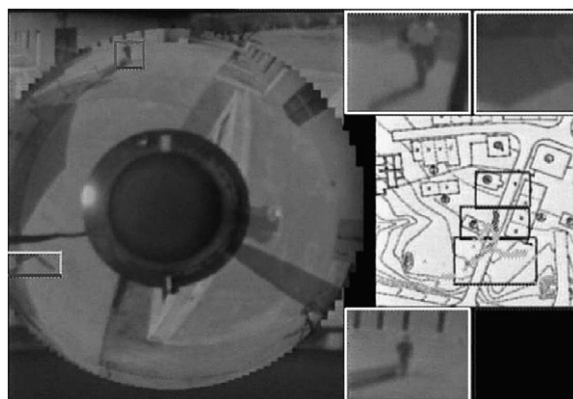


Figure 2.45: Omni directional object tracking system presented in Boulton et al. (2004).

There are other Omni-directional camera-involving proposals, such as the one described in Gandhi and Trivedi (2004), that dive directly into the theoretical issues of Omni-directional camera-based video surveillance systems by describing a formulation and application of parametric egomotion compensation to avoid 360° image distortion for a mobile platform equipped

with an Omni-directional Vision Sensor (ODVS) (Figure 2.46).

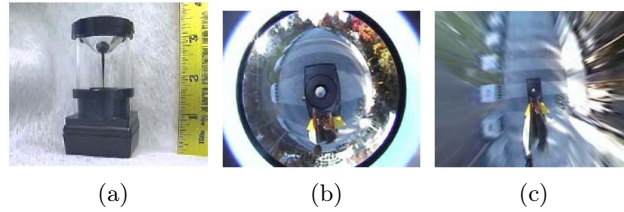


Figure 2.46: System described in Gandhi and Trivedi (2004). (a) Omni-directional Vision Sensor (ODVS). (b) A typical image from an ODVS. (c) Transformation to a perspective plan view.

Given the power, versatility, and competitive price of the PTZ cameras, they could be expected to be utilised as a part of the new Omni-directional imaging video surveillance systems, as it is illustrated in work presented in Scotti et al. (2005), where Panoramic Scene Analysis (PSA) is used to develop an integrated multi-camera system intended to be used in video surveillance applications. Two kinds of image sensors were used in the proposed system: A catadioptric sensor and a PTZ camera, both working symbiotically to integrate a kind of sensor that is able to automatically track, at a higher zoom level, any moving object within the guarded area. Hence, the catadioptric sensor is in charge of detecting any kind of movement in the scene, and once this happens, the system adjusts the PTZ camera parameters to the object location, causing the object that triggered the movement alert to be appropriately targeted and tracked as it moves through the scene.



Figure 2.47: Pedestrian tracking by the system presented in Scotti et al. (2005).

Integration between Omni-directional cameras and PTZ cameras is quite common, mainly because of the capability of a PTZ camera to perform a zoom movement towards a moving object targeted by the Omni-directional camera, as it can be found in works like Sato et al. (2008), where a surveillance system based on a combination of an Omni-directional camera and a networked PTZ camera is proposed.

Other proposals include highly sophisticated methods to implement panoramic camera-based video surveillance systems, such as the model described in Wang and Zhu (2012), where the authors propose a new approach, namely Multimodal Temporal Panorama (MTP), to accurately extracting and reconstructing moving vehicles in real-time, using a remote multimodal (audio/video) monitoring system. The system consists of a Panoramic View Image (PVI) for detecting vehicles using the concept of a 1D vertical detection line, an Epipolar Plane Image (EPI), generated from 1D epipolar lines along the vehicles' moving paths and corresponding audio signals collected at the vehicle detection point. This construction aims to palliate some common issues affecting these systems, such as occlusions, motion blur, and perspective changes.

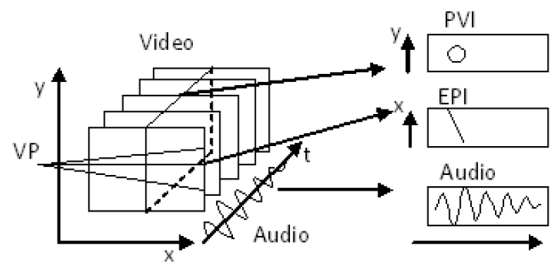


Figure 2.48: MTP representation used in the work by Wang and Zhu (2012).

As it can be noted in the works mentioned above, back in the 2000s, the most extended way of working with omnidirectional imaging-based video surveillance systems was to use expensive catadioptric sensors conveniently installed in room ceilings or building ledges that supplied an Omni-directional but somewhat distorted, image of the environment to be overlooked. Furthermore, because of this reason, they had to be completed by adding a PTZ camera that properly operated, supplied a high-definition image of the object detected by the Omni-directional sensor. Fortunately, nowadays, technology allows us to have cheaper and more powerful devices to obtain a high definition panoramic view of the areas under vigilance. These are the systems based on the modern 360° panoramic cameras. These devices often consist of a cluster of special cameras specifically placed and calibrated to cover an angle of 360 degrees in every direction, supplying a “spherical” image, representing the environment around the location of the cited camera (Figure 2.49).

A proper example of a system using this kind of technology can be found in Fan and Xu (2019), where the authors have developed a 360-degree environmental surveillance system designed to achieve continuous monitoring of the surrounding environment. The system consists of five fixed-focal-length cameras and one variable-focal-length camera image acquisition module capable of detecting movement in the frame by using pixel-level detectors than



Figure 2.49: Left, image of a Ladybug5+ 360° panoramic camera; right, 360° frame captured by the Ladybug5+.

will be used to track any moving object. The system, later, will identify the intruder through deep learning-based classifiers. Finally, we can find a similar work in Meng et al. (2018), where the authors illustrate the development of an underwater drone with a 360° panoramic camera acting as the “eye” of the drone. The system is intended to be used in underwater fish observation and classification. Just as it was done in the paper referred above, the fish identification task will be performed by several deep learning-based image classifiers powered by convolutional neural networks.

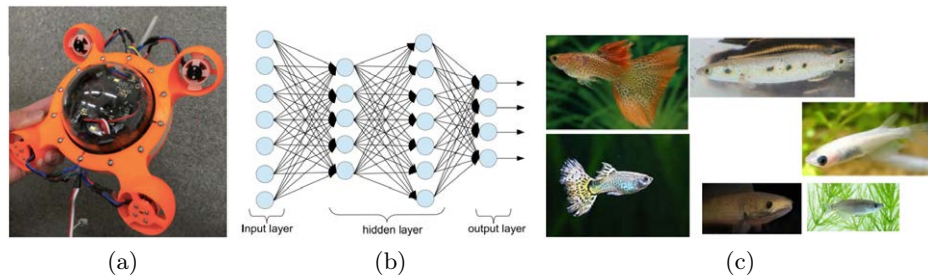


Figure 2.50: (a) Underwater drone. (b) Convolutional Neural Network (c) Fish classes observed and identified by the system developed in Meng et al. (2018).

2.5 Microcontrollers, microcomputers and low-cost hardware-based surveillance systems.

Computer vision-based automatic video surveillance systems usually require high computing power demanding tasks and, therefore, powerful image processing engines, which must be deployed in high-performance hardware devices that tend to be expensive and have high electric power demands. However, the high demand for these surveillance systems and the requirement for them to be as autonomous as possible in terms of both computing power and electric consumption have motivated in the last years the arising of a new trend

in Data processing. This trend is called *Edge Computing*, and it consists of a new distributed computing paradigm that moves computation and data storage as close as possible to the source of those data in order to save bandwidth and reduce response times. If we add to this paradigm the benefits represented by the reduction of the electric power needed and cheaper hardware devices, such as microcontrollers and SOC-based microcomputers, and we apply them to video surveillance systems, we will be talking about low-cost hardware-based video surveillance systems. However, the last and more advanced computer vision-based object detection and identification systems tend to rely upon deep learning techniques, which usually have large computational requirements supported by GPU acceleration, which is expensive and has high electric power demands. These facts force researchers to perform an optimisation work in both models and algorithms, representing the most challenging part. For instance, in Tong et al. (2014), the authors propose a low computation moving object detection method combined with video encoder oriented to optimise the moving object detection for video surveillance applications. The method described by the authors starts the object detection process at the same time the video is encoded to be processed or stored. This technique improves the efficiency of the detection process.

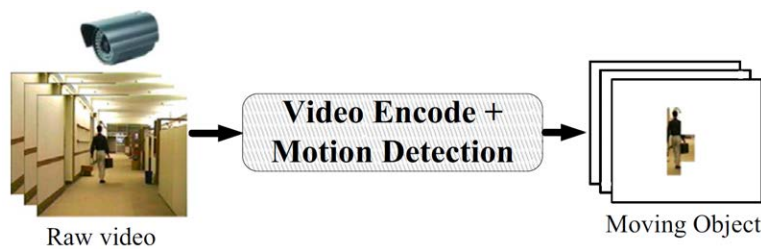


Figure 2.51: Moving object detection embedded in the video encoding phase performed in Tong et al. (2014).

As mentioned before, low-cost and power-efficient hardware devices are critical in designing object detection systems meant to be deployed in highly autonomous platforms such as Autonomous Vehicles. Hence, the research described in Angelov et al. (2017). In this work, the authors illustrate the design and implementation of a computationally efficient system for object detection on a moving platform that can be deployed on small, lightweight, low-cost and power-efficient hardware, intended to be mounted in an Unmanned Aerial Vehicle (UAV).

Because of their competitive price, versatility and a large amount of information available online referred to them, Raspberry Pi-type boards have tacitly established a new standard in low cost and power-efficient systems. For example, the work in Dziri et al. (2016) presents a new multi-camera real-

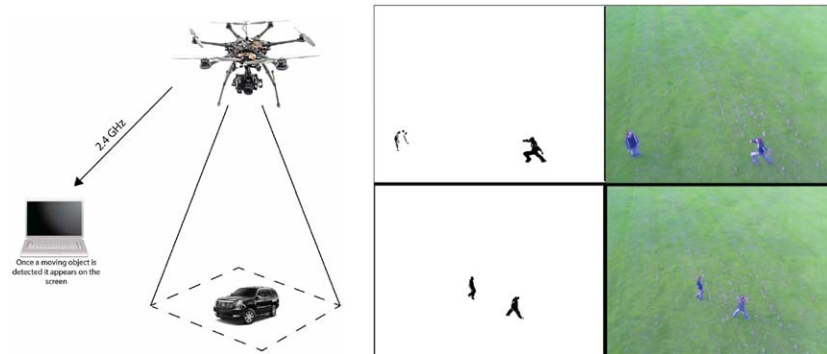


Figure 2.52: Left, Schematic representation of the system proposed in Angelov et al. (2017); right, real life experiment object detection results with the Angelov et al. (2017) proposal.

time multiple object detection system made from a tracking pipeline designed for fixed low-cost embedded smart cameras composed of a Raspberry-Pi board and a RaspiCam camera. This system is also capable of handling occlusions between objects.

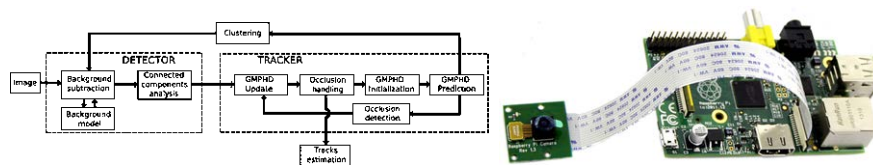


Figure 2.53: Left, smart camera composed by a Raspberry Pi and a RaspiCam used in Dziri et al. (2016); right, tracking pipeline powered by low-cost hardware developed in Dziri et al. (2016) proposal.

The world of portable and low energy consuming hardware devices is not constrained to microcontrollers and microcomputers. Field Programmable Gate Arrays (FPGAs) are a powerful type of electronic devices that stands out because of their reduced size and high efficiency, capable of executing the most complex calculus at high speeds. Moreover, with the appropriated optimisation and programming, it is possible to use them to train and test artificial neural networks with the consequent benefits they can bring to video surveillance systems. To illustrate this, we can focus on the work by Ortega-Zamorano et al. (2017), where the authors present an FPGA board implementation for the back-propagation neural networks algorithm. This implementation is done by using a multiplexing layer scheme, in which a single layer of neurons is physically implemented in parallel but can be reused any number of times in order to simulate multilayer architectures, representing an important step towards the FPGA implementation of deep

neural networks, one of the most successful existing models for prediction problems.

Getting back to environment perception using low profile hardware devices, in the existing literature, we can find several works presenting motion and proximity estimation and obstacle avoiding systems powered by very cheap and low power consuming devices as we can see in Dobrzynski et al. (2012). In this work, the authors present a new class of flexible compound-eye-like linear vision sensor dedicated to motion extraction and proximity estimation, namely Vision Tape (VT), consisting of an array of eight photodiodes attached to a flexible PCB that acts as mechanical and electrical support (Figure 2.54), capable of performing image acquisition and processing with an integrated microcontroller at a frequency of 1000 fps, even during bending of the sensor.



Figure 2.54: Left, image of the Vision Tape in curved configuration; right, examples of integration of the VT onto different substrates (Dobrzynski et al. (2012)).

The proposal described in Fung et al. (2014) also constitutes another example of the application of microcontrollers to object motion and position estimation, in this case, applied to weather prediction. Indeed, a compact and economic system is presented that measures cloud shadow motion vectors to estimate power plant ramp rates and provide short-term solar irradiance forecasts. This device is made from an array of luminance sensors and a high-speed data acquisition system to resolve cloud passages' progression across the sensor footprint. The data is processed by a microcontroller that uses a Cross-Correlation algorithm to determine the cloud motion vectors.

These microcontroller-based technologies can also be very useful when applied to a field in constant development such as smart cities and green cities, as it is referred to by the authors in Adnan et al. (2015). In the cited work, an initial development for an Energy-Saving Street Lighting (ESSL) system is developed. This development is based on a network of sensor nodes that can be used to control the brightness of the street light to save energy through an alternate switching method.

When running properly-optimised models and algorithms, microcontroller-based systems can present extremely efficient operation, allowing them to

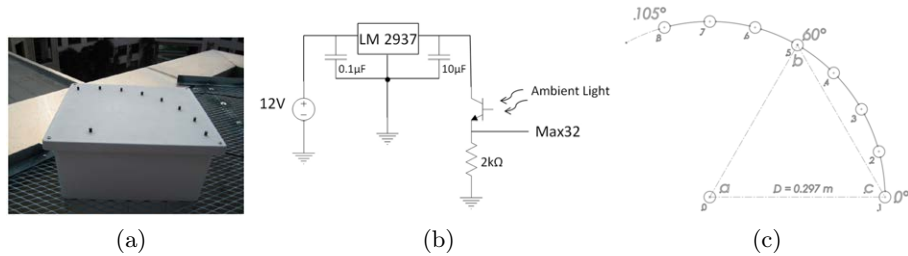


Figure 2.55: (a) CSS presenting a weather-proof enclosure. (b) Simplified schematic of the luminance sensor. (c) Sensor arrangement presenting the luminance sensors distribution through the enclosure. Fung et al. (2014).

perform at high accuracy levels, even when working with complex neural network models. These systems constitute a very attractive alternative in the design of highly efficient video surveillance systems, as is demonstrated by the authors in Ortega-Zamorano et al. (2016). In the cited work, an Arduino DUE board is used to deploy a novel self-organising map-based motion detection system in charge of processing the images supplied by a static conventional surveillance camera, resulting in a microcontroller-powered, highly accurate, automatic video surveillance system. (Figure 2.56).

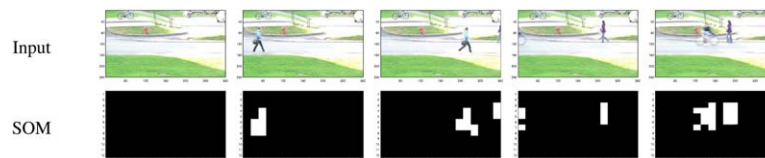


Figure 2.56: Motion detection examples for a pedestrian video with the SOM-based video surveillance system developed in Ortega-Zamorano et al. (2016).

One of the most important tasks automatic video surveillance systems require is object detection and identification. However, in order to achieve the best results possible, nowadays, these tasks often imply high computing power demanding tasks and massively-trained deep neural network-based models running in powerful GPU devices. Hence, the importance of rigorous research on designing neural network-based video surveillance systems that can be deployed in low-cost hardware devices. This circumstance leads us towards the four surveillance systems supporting this PhD Thesis, described in detail in the following four chapters of this memory.

The first one was entitled *Motion detection with low-cost hardware for PTZ cameras* and published in the *Integrated Computer-Aided Engineering (ICAE)* Journal. It belongs to the area of automatic video surveillance

systems. Consequently, it consists of an automatic motion detection system for PTZ camera video streams, supported by a multilayer perceptron and deployed in a Raspberry Pi Microcomputer.

The second work presented in this thesis was published at the *International Joint Conference on Neural Networks (IJCNN) 2018* Conference in Rio de Janeiro, Brasil. Entitled *Deep learning-based anomalous object detection system powered by microcontroller for PTZ cameras*, it describes the design and implementation of a deep learning-based video surveillance system, capable of detecting stationary or in motion foreground anomalous objects in video streams supplied by a PTZ camera, also deployed in a Raspberry Pi Board.

The third original research work developed for this PhD Thesis was published in the *Integrated Computer-Aided Engineering* Journal and proposed a deep learning-based automatic video surveillance system for 360° panoramic camera video streams, once again optimised to be deployed in a Raspberry Pi board. It was entitled *Deep learning-based video surveillance system managed by low-cost hardware and panoramic cameras*.

It is convenient to recall that the work developed in this thesis is oriented to the creation and implementation of neural network-based automatic surveillance systems. However, these kinds of systems' performance often rely on our abilities for developing innovative neural network models, so the last original work described in this thesis and entitled *Image Clustering Using a Growing Neural Gas with Forbidden Regions*, has a stronger theoretical orientation than the other three. Published at the *IJCNN 2020* celebrated in Glasgow, Scotland (UK), it consists of a system capable of predicting the position of different creatures inhabiting oceanic environments by performing a clustering operation of the actual sightings of those animals by using a novel Growing Neural Gas (GNG) variation with forbidden regions that we named the *Forbidden Regions Growing Neural Gas* or FRGNG.



UNIVERSIDAD
DE MÁLAGA

Part II

Research work



UNIVERSIDAD
DE MÁLAGA

Chapter 3

Motion detection with low cost hardware for PTZ cameras

*Justice does not help those who slumber
but helps only those who are vigilant.*

Mahatma Gandhi

ABSTRACT: This chapter presents our first research work published in the journal *Integrated Computer-Aided Engineering (ICAE)* in the year 2018. This work describes the design and implementation of a motion detection-based video surveillance system for PTZ cameras, adequately adapted to be deployed in a Raspberry Pi 3 board. The foundation of this work relies on specific processing of the video stream coming from the PTZ camera and utilising three types of classifiers, which will be in charge of movement detection.

Experimental results in both accuracy and speed measurements illustrate this system's suitability to be used as a movement detector for PTZ camera video streams.

Title	Motion detection with low cost hardware for PTZ cameras
Authors	Jesús Benito-Picazo, Enrique Domínguez, Esteban J. Palomo, Ezequiel López-Rubio, Juan Miguel Ortiz-de-Lazcano-Lobato
Journal	Integrated Computer-Aided Engineering
Year	2018
Impact Factor	5,264
JCR categories	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE (21/132 (Q1)) ENGINEERING, MULTIDISCIPLINARY (7/86 (Q1)) COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS (17/105 (Q1))
Status	Published
DOI	https://doi.org/10.3233/ICA-180579
Cite	Benito-Picazo, J., Domínguez, E., Palomo, E. J., López-Rubio, E., and Ortiz-De-Lazcano-Lobato, J. M. (2018b). Motion detection with low cost hardware for ptz cameras. <i>Integrated Computer-Aided Engineering</i> , 26(1):21–36

Chapter 4

Deep learning-based anomalous object detection system powered by microcontroller for PTZ cameras

Today, our very survival depends on our ability to stay awake, to adjust to new ideas, to remain vigilant and to face the challenge of change.

Martin Luther King Jr

ABSTRACT: The second work supporting this PhD Thesis was presented in the *International Joint Conference on Neural Networks (IJCNN)* as part of the *IEEE World Congress on Computational Intelligence (WCCI)* celebrated in July 2018 in Rio de Janeiro, Brazil. This work constituted the first step towards developing the most ambitious part of this thesis, presenting a novel video surveillance system for PTZ cameras, featuring a novel anomalous object detection and identification system based on convolutional neural networks specially optimised to be deployed in low-cost hardware System-On-Chip boards. Experimental results attest to this proposal's suitability to serve as a low-cost anomalous foreground object detection and classification system for PTZ camera video streams.

Title	Deep learning-based anomalous object detection system powered by microcontroller for PTZ cameras
Authors	Jesús Benito-Picazo and Enrique Domínguez and Esteban J. Palomo and Ezequiel López-Rubio and Juan Miguel Ortiz-de-Lazcano-Lobato
Conference	International Joint Conference on Neural Networks, (IJCNN 2018)
Year	2018
GGs Rating	B
CORE Rating	A
Status	Published
DOI	https://doi.org/10.1109/IJCNN.2018.8489437
Cite	Benito-Picazo, J., Dominguez, E., Palomo, E. J., Lopez-Rubio, E., and Ortiz-De-Lazcano-Lobato, J. M. (2018a). Deep learning-based anomalous object detection system powered by microcontroller for ptz cameras. In <i>Proceedings of the International Joint Conference on Neural Networks</i> , volume 2018-July

Chapter 5

Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras

The superior man, when resting in safety, does not forget that danger may come.

Confucius

ABSTRACT: The third work supporting this PhD Thesis was published in the *Integrated Computer-Aided Engineering (ICAE)* journal in 2020. The paper in question presents a more advanced deep-learning-based anomalous object detection and classification algorithm than the one described in Chapter 4. The cited algorithm is intended to be used to implement a novel video surveillance system for panoramic 360° cameras, presenting important improvements in both the image acquisition device and the potential detection generator functions. More precisely, aiming to get over the mechanical issues intrinsic to the PTZ cameras, in this work, panoramic cameras have been used. Furthermore, two new multivariate homoscedastic distributions have been added to improve the potential detection generator described in the work described in Chapter 4.

Results yielded by the experiments reveal that the system illustrated in this work outperforms one of the most popular object detection and classification algorithms in both accuracy and speed when deployed in a Raspberry Pi 3 Model B, presenting a suitable alternative for building low-cost automatic video surveillance systems driven by anomalous detection and identification deep learning-based models.

Title	Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras
Authors	Jesús Benito-Picazo, Enrique Domínguez, Esteban J. Palomo and Ezequiel López-Rubio
Journal	Integrated Computer-Aided Engineering
Year	2020
Impact Factor	4,706
JCR categories	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE (25/137 (Q1)) ENGINEERING, MULTIDISCIPLINARY (10/91 (Q1)) COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS (15/109 (Q1))
Status	Published
DOI	https://doi.org/10.3233/ICA-200632
Cite	Benito-Picazo, J., Domínguez, E., Palomo, E. J., and López-Rubio, E. (2020a). Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras. <i>Integrated Computer-Aided Engineering</i> , 27(4):373–387

Chapter 6

Image Clustering Using a Growing Neural Gas with Forbidden Regions

The Milky Way is nothing else but a mass of innumerable stars planted together in clusters.

Galileo Galilei

ABSTRACT: The last work supporting this PhD Thesis was presented in the *International Joint Conference on Neural Networks (IJCNN)* as part of the *IEEE World Congress on Computational Intelligence (WCCI)* celebrated in Glasgow, Scotland, UK, in July 2020. The work in question follows a parallel direction in surveillance systems with respect to the rest of this thesis's works because it is not focused on detecting and identifying anomalous objects in a certain scene or the motion detection in the video feed supplied by some surveillance camera. Moreover, the system described in this chapter presents a surveillance system oriented to monitor individuals' behaviour from the data corresponding to the locations those individuals were seen at. Thus, in this work, a novel unsupervised learning model based on the Growing Neural Gas is described, namely Forbidden Regions Growing Neural Gas (FRGNG), designed to keep all of its neuron prototype vectors out of a particular set of convex barriers as they move along the input space through the training of the network. More specifically, this new model is employed to monitor the migration behaviour of some marine animal species by performing clustering tasks, describing the topological relations between the different clusters, over some biogeographical data supported by the sightings of those animals around certain geographical areas whose forbidden regions are represented by the mainland zones of those geographical areas.

Experimental results attest to the competitiveness of the FRGNG.

Title	Image Clustering Using a Growing Neural Gas with Forbidden Regions
Authors	Jesús Benito-Picazo, Antonio Díaz Ramos, Esteban J. Palomo and Enrique Domínguez
Conference	International Joint Conference on Neural Networks, (IJCNN 2020)
Year	2020
GGs Rating	B
CORE Rating	A
Status	Published
DOI	https://https://doi.org/10.1109/IJCNN48605.2020.9207700
Cite	Benito-Picazo, J., Palomo, E. J., Dominguez, E., and Ramos, A. D. (2020b). Image clustering using a growing neural gas with forbidden regions. In <i>Proceedings of the International Joint Conference on Neural Networks</i>

Conclusion



UNIVERSIDAD
DE MÁLAGA

Chapter 7

Conclusion and future research lines

Thoroughly conscious ignorance is the prelude to every real advance in science.

James Clerk Maxwell

ABSTRACT: This chapter contains the conclusions extracted from the four works supporting this PhD thesis and several future research lines motivated by the experimental results obtained after testing the different models developed on them. The goal is to summarise the main contributions from this research period and formulate some ideas for keeping them evolving towards more powerful and robust systems.

7.1 Conclusion

This chapter presents the results achieved after four years of research on artificial intelligence. Four years dedicated to study, develop, test and optimise novel artificial neural networks-based surveillance systems to be deployed in low-cost hardware devices. Different mathematical models have been used throughout this work, including the Multilayer Perceptron (MLP), the Self-Organising Map (SOM) neural networks family, and the deep Convolutional Neural Networks (CNNs) integrated into the popular and powerful Deep Learning (DL) techniques. Therefore, this research is supported by four papers published in high impact factor journals and international conferences that attest to its relevance and novelty. Besides, during this thesis's development, some additional works were published in moderate impact factor journals and international conferences, presenting intermediate developments

of the four works that actually support this thesis. It is worth noting that the four works mentioned above are not isolated from each other. On the contrary, they constitute what could be categorised as a comprehensive automatic video surveillance system for low-cost hardware, consisting of different parts that handle tasks such as motion detection in image streams from surveillance cameras, detection and identification of objects in image streams from PTZ and 360° cameras, and prediction of the behaviour of individuals by estimating their current position in a natural environment from previous observations.

The multilayer perceptron is one of the first artificial neural network models to be used in machine learning and still one of the most popular models in the artificial neural networks field, both academically and for real-world applications, because of its simplicity and its ability to approximate functions. This network is also the first to be studied and used in this thesis to design and implement automatic video surveillance systems for low-cost hardware.

Indeed, in Chapter 3 we have illustrated the construction of a real-time motion detector for video streams supplied by Pan-Tilt-Zoom (PTZ) cameras and hosted by a low-cost piece of hardware. The system relies on an algorithm that processes a sequence of images coming from a moving PTZ camera, comparing subsequent frames to detect changes that could reveal the presence of possible moving foreground objects. Starting by taking two consecutive frames, the algorithm transforms the second one using a mathematical model and divides the two of them into a set of stripes. Next, it calculates the Mean Squared Error (MSE) error when comparing one stripe from one frame with its equivalent from the subsequent transformed frame. The numbers yielded by these comparisons will be used as the inputs of an adequately trained classifier who will decide whether a foreground moving object is present on the scene. In order to make this detector more reliable and practical, one different mathematical model has been designed for each type of movement performed by the PTZ camera. Those models have also been simplified by considering every movement to be executed at a constant speed and only one movement at a time.

The system has been trained separately in a desktop computer and successfully deployed in a Raspberry Pi 3 Model B microcomputer, obtaining a foreground object detector that will be the “powerhouse” of a budget and power-efficient automatic video surveillance system.

Three types of well-known machine learning models have been used as classifiers in charge of pointing out whether there are foreground objects in the scene: a Multilayer Perceptron (MLP), a Support Vector Machine (SVM) and a K-Nearest Neighbours (KNN) algorithm implementation. Also, three different measures have been proposed to classify the system’s performance: Accuracy, F-measure and the Area Under the ROC Curve (AUC).

Experimental results reveal that the numbers obtained for every classifier are different but almost equally good. Regarding classification accuracy, the KNN based classifier stands out from its competitors, whilst the SVM based classifier achieves the best balance between speed and accuracy. With processing speeds higher than 24 frames per second for every classifier type, the system developed in Chapter 3 presents a suitable proposal for a cheap and power-efficient automatic video surveillance system that operates in real-time.

As we stated before, several artificial neural network-based models were used in the research process required to support this thesis. Deep Learning (DL) is a popular machine learning technique based on multi-layered artificial neural networks. Especially useful for object classification tasks on images, convolutional neural networks are among the most used models in the research developed in this thesis, just as in work described in Chapter 4. This paper illustrates the construction of a low power demanding automatic video surveillance system for PTZ camera video streams using low-cost hardware devices. Featuring a deep learning-based anomalous foreground object localisation and classification algorithm, the core of this system integrates two well-differentiated parts. The first part is a detection and localisation engine, constituted by a potential detection generator based on a mixture of Gaussian and random distributions. The second part is a convolutional neural network-based classifier in charge of identifying the object appearing in the window framed by the potential detection generator.

The operating of the system starts with the acquisition of one picture from the video stream supplied by the PTZ camera. Next, the potential detection generator engine selects a certain number of portions of the image whose size and location will be determined by the cited gaussian-random mixture function, specially designed for that purpose. As indicated above, the built-in potential detection generator function consists of a gaussian-random mixture with two parts. The first part consists of a Gaussian distribution that helps the generator to preserve some memory about the locations of the previously detected objects. The second part consists of a random distribution that provides some novelty in the potential detection generation, so new objects entering the scene can be detected. Once the potential detection generator has guessed some areas of the frame where objects could be potentially detected, those areas are fed to a previously trained convolutional neural network, which will be in charge of identifying the objects framed in those areas and deciding whether they are considered as anomalous. This decision will be made by looking for them in a list of objects previously considered as anomalous for that environment. The program will emit an alert in case any of the detected objects is found in the cited list. Again, in order to preserve one of the main targets of this thesis, all the system has been optimised to be deployed in a cheap and low energy consuming hardware

device like the Raspberry Pi 3 Model B, by using small pre-trained convolutional neural networks and special software libraries in order to leverage the multi-core structure of the cited hardware. It is important to remark that extra ventilation was needed to prevent hardware from being burnt out due to work overload.

Experimental results reveal that the system can detect several static or in motion anomalous objects in up to half a second approximately, confirming this proposal's suitability to build a cheap and energy-efficient approach to automatic video surveillance systems.

Convolutional neural network-based systems have been very recurrent in the development of this research. Also, the use of PTZ cameras because of their versatility and ease of use. However, after some research, it was decided to try other image capture devices that, given the characteristics of the object detection and identification algorithms developed in this thesis, could offer a complete view of the environment to be overlooked. Thus, 360° cameras were considered as the next image acquisition devices to continue our research since they can supply a panoramic view of the environment they are installed in by capturing a spherical image of the real scenario. Consequently, they have been used as the primary image capture device in work described in Chapter 5, where the authors present an improvement to work developed in Chapter 4, incorporating the use of 360° panoramic cameras and a new potential detection generator based on three new homoscedastic probabilistic distributions, in order to build a new anomalous foreground object detector to be deployed in a Raspberry Pi 3 Model B.

Indeed, this work describes a new and improved automatic video surveillance system based on the foreground anomalous object detector mentioned above. The system's operation starts with acquiring a picture from the 360° panoramic camera stream. Subsequently, it sets a certain number of potential detections all over the image, following the positions and sizes generated by one of the three multivariate homoscedastic distributions designed for that purpose. Cited distributions consist of three mixtures of one random distribution and a triangular, a gaussian and a Student-t distribution, respectively. Once again, every area framed inside each of the potential detections will be fed to a previously trained convolutional neural network, specially optimised to require low computational resources. This network will recognise the objects present in the area enclosed in the potential detection to look for anomalous objects within. When deployed in Raspberry Pi-type hardware devices, the resulting system overtakes one of the most popular and influential systems in the current state of the art in terms of location, classification capabilities and speed.

As it was mentioned at state of the art in Chapter 2, deep learning is not the only artificial neural network-based technology to be successfully used in surveillance tasks. There are other artificial neural network-based models

which, because of their operating and features, can be used for surveillance tasks in systems different from the ones presented above. Thus, it was considered to investigate different types of neural networks other than the deep convolutional networks and their utility when designing systems that allow performing surveillance tasks at any level. Eventually, some of the surveillance tasks mentioned above may involve clustering processes intended to make regression upon any data. One of the many tasks scientists can use those data for consists of making observations on a group of individuals to extrapolate their geographical location through time, so predictions on the behaviour of that group of individuals can be made. Along these same lines, Chapter 6 describes a Growing Neural Gas variation, namely Forbidden Regions Growing Neural Gas (FRGNG), that can perform clustering tasks revealing the topological relations between those clusters, keeping its neuron prototypes out of a set of polygonal convex barriers. In the case of the work presented in Chapter 6, this network is used to perform clustering tasks on certain biogeographical data coming from the sightings of individuals belonging to various marine animal species around the coasts of specific geographical locations in order to monitor and possibly predict their migratory habits. In this context, as none of the marine species individuals can be sighted on land, the mainland areas of these locations would be considered as the forbidden regions the neurons of the FRGNG are not allowed to “invade” on their drift over the training process of the network.

Experimental results reveal that, even without parameter optimisation, the FRGNG presented in this thesis is a good alternative against the Forbidden Region Self Organising Feature Map (FRSOM) in its optimised-parameter version, presented as the most advanced competitor in state of the art.

The rest of the most relevant works developed throughout this thesis consist of conference papers¹ presenting early development stages of some of the neural network-based systems cited above, attesting step by step the complexity and difficulties faced by the author through all the process.

7.2 Future research lines

Throughout the development of any significant and rigorous research period, several future research ideas emerge motivated by the challenges faced as part of the problem resolution processes and the results obtained in the performed experiments.

Considering that most of the works presented to support this thesis use either deep convolutional neural networks, multilayer perceptrons or Self-Organising Map-based models, future research goes along three different

¹A small extract from all of the cited conference papers can be found in the appendix

lines: deployment optimisation of multilayer perceptron-based object detectors in low-cost hardware, deep learning-based video surveillance systems optimised to be deployed in low-cost hardware devices and Self-Organising Map-based models consisting of evolved versions of the Forbidden Regions Growing Neural Gas.

Starting with the design and development of multilayer perceptron-based movement detection systems, we find it very interesting to keep using these networks to build video surveillance systems because of their speed and versatility. Thus, if we remember the work developed in Chapter 3, the presented algorithm takes two consecutive frames, n and $n + 1$, from a PTZ camera video stream and transforms the frame $n + 1$ using a mathematical model so it can be compared to the frame n in order to generate a vector of numbers that would be fed to a classifier. This classifier would be in charge of deciding whether any motion was detected in the scene. In this context, it seems reasonable to think that movements performed by a surveillance camera set in a “monitoring” configuration had more sense to be at a constant low speed. Therefore, in order to simplify the cited mathematical model, it was considered that the pan, tilt and zoom movements of the PTZ camera were supposed to be executed at a constant speed, dismissing the possibility that the camera could perform an accelerated movement. Furthermore, just for simplicity, this mathematical model also does not contemplate the possibility that all the PTZ camera movements for reaching a certain position might be done at the same time, reducing the time needed by the camera to aim to the desired coordinates.

- Thus, in order to implement a more powerful, fast and complete movement detection system, future work suggests researchers undertaking the design of a new mathematical model in charge of fixing that possible biasing by preparing the system for a PTZ camera capable of performing all the necessary movements to reach a specific location at a non-constant speed and at the same time. This way, a faster and more accurate movement detector would be obtained.
- The other possible future research that could improve the performance of the system described in Chapter 3 is related to the optimal usage of the Raspberry Pi 3 Model B device as this is the hardware it is designed to be deployed in. The current implementation of the algorithm supporting this system does not present any multi-core specialised techniques, so the results obtained by the experiments with the current mathematical model and the three different classifiers employed have been obtained by using only one of the four cores supplied by the Raspberry Pi 3 Model B CPU. Therefore, future research lines for this movement detector must consider a new multi-core specific implementation that would take advantage of all the Raspberry Pi’s

computing power that would potentially accelerate the system's performance up to $4x$ the current speed. This improvement in the system would imply a multi-threading mutex-based implementation in order to force all the processing parts to run in parallel. Of course, this new use of the hardware would lead to the development and installation of a new additional ventilation system to prevent a possible malfunction of the hardware due to overheating causes.

Deep learning-based systems are very popular and, despite their current high development status, still, a very promising technology in machine learning, as illustrated in works like the one presented in Chapter 4 of this PhD Thesis. In this article, a video surveillance system for detecting anomalous objects in scenes shot by a PTZ camera is developed. Therefore, the research detailed in this paper leads us to propose some possible future research lines.

- The first research line aims to obtain some open-source video footage from moving PTZ cameras containing one or more objects considered anomalous for the scene, so it constitutes an adequate dataset to perform a battery of more complete tests to the developed system. This task is not easy because most of the existing PTZ camera footage belongs to security companies and corresponds to public and private monitored environments. The most recommended way, which is also the most expensive, to obtain PTZ footage, is to get a PTZ camera of our own to shoot videos of previously prepared and controlled scenarios, using these videos to test this system and the future systems based on it. This process will lead us to create a new dataset that will be published to be shared with the scientific community.
- In this work, it has been used a pre-trained convolutional neural network designed by the Microsoft ELL Team, which is able to identify objects from 1000 categories. However, the number of objects considered anomalous depends highly on the type of environment being monitored, so there is no point in considering a fixed amount of anomalous object categories. Furthermore, an over-dimensioned network demands unnecessarily large computing resources that slow down the processing speed achieved by the low-cost hardware they are meant to be deployed in. Therefore, the second future research line pursues the improvement of the object detection model by creating smaller and more specialised networks that can classify the adequate amount of anomalous object categories required for each environment. This way, the object recognition will be more robust, and at the same time, the processing speed in frames per second would increase, obtaining a more effective system.

It is essential to remark that not all of the potential improvements that

can be made to a deep learning-based anomalous object detection system ought to go through the improvement or optimisation of the neural network it relies on. Indeed, the results achieved by the research developed in Chapter 5 suggest some future research lines that are worth to explore. In the cited chapter, it is detailed an automatic video surveillance system for anomalous object detection, powered by panoramic 360° cameras and deployed in low-cost hardware. It presents important improvements with respect to the system designed in the article described in Chapter 4, such as the use of a 360° panoramic camera, providing all the system with a complete spherical vision, which allows, without processing speed penalty, the monitoring of the entire environment at a glance. The second improvement presented by this work is the design of two new potential detection generator distributions, including Student-t and triangular distribution mixtures, which bring versatility and more accuracy to the object detection tasks. Obtained results demonstrate this proposal's suitability as it outperforms one of the most popular object detection and classification systems of the state of the art when installed in low-cost hardware. However, experimental results also suggest several potential developments based on this work that can be grouped into three future research lines:

- The work presented in Chapter 5 uses a panoramic 360° camera to produce a spherical video stream from the environment that is meant to be under vigilance. However, it is quite challenging to find panoramic camera free usable footage from environments with anomalous objects. Consequently, all the tests performed in this work have been performed over a panoramic 360° video stream where anomalous objects have been introduced artificially, using video edition, in order to obtain the most controlled environment possible. This future line of development aims to get a panoramic camera of our own and prepare some controlled real environments with anomalous objects inside them to create a public dataset. This way, it could be shared with the scientific community to help other researchers design and test their models.
- The convolutional neural network the system relies on is one of the pre-trained networks from the Microsoft ELL Team, and once again, it is quite oversized, making the system slower. Hence, the second future research and development line consists of designing a new network, smaller and specifically trained to classify just a certain amount of categories that are considered anomalous for each environment. This way, we would get a smaller network that would improve our video surveillance system by making it more accurate and faster when deployed in a Raspberry Pi board-based hardware.
- The third future research line inspired by this work consists of optimising the parameters of the three multivariate homoscedastic dis-

tributions that constitute the foundations of the probabilistic model utilised to implement the potential detection generator. Indeed, the parameters corresponding to the cited probability distributions, such as the spread parameter, σ , and the mixing weight of the mixtures, q , were adjusted empirically after a trial and error process. However, in no way they have passed through a systematic optimisation process. Therefore, in order to get the highest performance from this system, the *Picasso* node from the Spanish Supercomputing Network (RES) will be used to perform a systematic optimisation process for both parameters.

The last batch of future research lines motivated by the different works developed in this thesis comes from the work described in Chapter 6, where a GNG-based model, namely Forbidden Regions Growing Neural Gas (FRGNG), is developed for monitoring and predicting animal species migration behaviour. Therefore, the FRGNG is a novel GNG-based neural network designed to perform clustering tasks, preserving the topological relations between the different clusters over data distributions with regions where no data can be observed. This Forbidden Regions Growing Neural Gas is used in this paper to monitor and predict some marine animal species' migration behaviour by clustering the biogeographical data corresponding to different sightings of these animals around the coast of several geographical areas. Needless to say that the forbidden areas where no marine animals can be sighted are the mainland zones of those geographical environments.

Considering the performance of the mathematical model presented in Chapter 6, the future research lines inspired by this work could follow four different paths.

- First, it has been considered the possibility of performing tests in other types of environments, where the FRGNG could be employed to predict observation patterns in other events, such as high accident-concentration zones in industrial environments, people movement patterns in crowded recreational events (such as musical shows or sports events), the evolution of infectious agents in human population and any other type of event eligible to be treated with forbidden areas clustering techniques.
- One of the most positive results obtained from the experiments performed on the FRGNG is its competitiveness with respect to the FRMOM network, which is the best known to date model from state of the art for clustering tasks with forbidden regions. However, it is very important to remark that the results achieved by the FRGNG in the tests performed were obtained with its parameters set to default values, whilst the values obtained by the FRMOM, in the work where it is described, were obtained after a very specific and rigorous parameter optimisa-

tion process, supported by the Spanish Supercomputing Network node *Picasso*, located at the University of Málaga facilities. Therefore, the second future research line inspired by this work would consist of an optimisation process intended to perform a meticulous tuning of the FRGNG parameters. This optimisation process would be supported by the *Picasso* supercomputer.

- The FRGNG model presented in Chapter 6 can perform clustering tasks from biogeographic data describing the topological relations between the different clusters keeping its neuron prototypes out of some barriers that we know as forbidden regions. However, this model is limited to bi-dimensional data distributions. Therefore, the last and most ambitious of the possible potential developments derived from this work will consist of designing and constructing a tri-dimensional version of the Forbidden Regions Growing Neural Gas, namely Forbidden Regions Growing Neural Gas 3D (FRGNG3D). This model will have the ability to perform clustering tasks in similar conditions as the FRGNG but with tri-dimensional data distributions, which will bring generality to our model, easing its utilisation in a broader range of potential applications.

7.3 Final reflection

The process of developing a PhD Thesis is the last and highest milestone in any researcher's academic career. For most people, this journey begins with a large dose of illusion, hope, unlimited ambition for knowledge and, why not, with a little uncertainty. In the case of the person who writes these lines, the time spent developing this thesis has been the most intense, inspiring and exciting period of his academic life.

Sometimes, scientists tend to be a little individualist. However, one word that is always hovering over the scientific community is *collaboration*. Furthermore, as one starts to progress in its research tasks, it is easy to notice that, nowadays, any researcher who works isolated from the scientific community is unavoidably condemned to failure. Consequently, publishing our work in conferences and journals must turn into a regular habit, as they are the cornerstone of the collaboration between researchers, so crucial for the progress of science.

These collaboration skills acquire a new dimension when regularly co-operating with other institutions such as universities and research institutes. Thus, some researchers, such as the one that is writing these thoughts, spend time staying in other countries' universities in order to share research practices between work teams. Not surprisingly, the time spent in research stay abroad was one of the most motivating and prolific in the development of

this thesis.

It has been a time of auto-discovering, a time for letting creativity fly and reach places where one never thought it could be and a time for pushing our capabilities beyond our supposed limitations. However, it has also been a time to be humble by being aware of how vast science and technology are and how limited is one's own knowledge and expertise. Of course, it has been a time to accept guidance from those who once were at the same point as we are and now are generous enough to help us do our best. Also, it is the right thing to thank them for their help and advice.

All in all, it has been a long journey. A time for looking through a door towards the vast knowledge that is out there for us. A time for hard work and sacrifice; and a time for meeting extraordinary people with the highest intellectual capabilities imaginable. And now that this period draws to a close, struggling between nostalgia and excitement, one can only imagine himself crossing that door and facing the new challenges science has reserved us.



UNIVERSIDAD
DE MÁLAGA

Appendices



UNIVERSIDAD
DE MÁLAGA

Appendix A

Publications Summary

ABSTRACT: This appendix presents a list of tables that summarises the information associated with the published works during the development of this thesis. For journals, it has been considered the *JCR* ranking whilst for the conferences the *GGG Conference Rating* published in 2018 and the *CORE 2020*, have been considered. The section starts with the information related to the four articles supporting this thesis by published works. Subsequently, it includes some information about additional works that, even though they were not selected to support this PhD Thesis, are worth to be included in this document as they correspond to intermediate steps of the works that were finally selected.

A.1 Works supporting this PhD Thesis

Title	Deep learning-based anomalous object detection system powered by microcontroller for PTZ cameras
Authors	Jesús Benito-Picazo and Enrique Domínguez and Esteban J. Palomo and Ezequiel López-Rubio and Juan Miguel Ortiz-de-Lazcano-Lobato
Conference	International Joint Conference on Neural Networks, (IJCNN 2018)
Year	2018
GGG Rating	B
CORE Rating	A
Status	Published
DOI	https://doi.org/10.1109/IJCNN.2018.8489437
Cite	Benito-Picazo et al. 2018a

Title	Motion detection with low cost hardware for PTZ cameras
Authors	Jesús Benito-Picazo, Enrique Domínguez, Esteban J. Palomo, Ezequiel López-Rubio, Juan Miguel Ortiz-de-Lazcano-Lobato
Journal	Integrated Computer-Aided Engineering
Year	2018
Impact Factor	5,264
JCR categories	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE (21/132 (Q1)) ENGINEERING, MULTIDISCIPLINARY (7/86 (Q1)) COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS (17/105 (Q1))
Status	Published
DOI	https://doi.org/10.3233/ICA-180579
Cite	Benito-Picazo et al. 2018b

Title	Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras
Authors	Jesús Benito-Picazo, Enrique Domínguez, Esteban J. Palomo and Ezequiel López-Rubio
Journal	Integrated Computer-Aided Engineering
Year	2020
Impact Factor	4,706
JCR categories	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE (25/137 (Q1)) ENGINEERING, MULTIDISCIPLINARY (10/91 (Q1)) COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS (15/109 (Q1))
Status	Published
DOI	https://doi.org/10.3233/ICA-200632
Cite	Benito-Picazo et al. 2020a

Title	Image Clustering Using a Growing Neural Gas with Forbidden Regions
Authors	Jesús Benito-Picazo, Antonio Díaz Ramos, Esteban J. Palomo and Enrique Domínguez
Conference	International Joint Conference on Neural Networks, (IJCNN 2020)
Year	2020
GGs Rating	B
CORE Rating	A
Status	Published
DOI	https://https://doi.org/10.1109/IJCNN48605.2020.9207700
Cite	Benito-Picazo et al. 2020b

A.2 Additional Publications

Title	Motion detection by microcontroller for panning cameras
Authors	Jesús Benito-Picazo, Ezequiel López-Rubio, Juan Miguel Ortiz-de-Lazcano-Lobato, Enrique Domínguez, Esteban J. Palomo
Conference	International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC 2017)
Year	2017
GGG Rating	- Work in Progress
CORE Rating	- National: Spain
Status	Published
DOI	https://doi.org/10.1007/978-3-319-59773-7_29
Cite	Benito-Picazo et al. 2017b

Title	Unsupervised Color Quantization with the Growing Neural Forest
Authors	Esteban José Palomo and Jesús Benito-Picazo and Ezequiel López-Rubio and Enrique Domínguez
Conference	International Work-Conference on Artificial and Natural Neural Networks (IWANN 2017)
Year	2017
GGG Rating	B -
CORE Rating	B
Status	Published
DOI	https://doi.org/10.1007/978-3-319-59147-6_27
Cite	Palomo et al. 2017

Title	Growing Neural Forest-Based Color Quantization Applied to RGB Images
Authors	Jesús Benito-Picazo and Ezequiel López-Rubio and Enrique Domínguez
Journal	International Journal of Computer Vision and Image Processing
Year	2017
Impact Factor	–
JCR categories	Not Indexed in JCR
Status	Published
DOI	https://doi.org/10.4018/IJCVIP.2017070102
Cite	Benito-Picazo et al. 2017a

Title	Deep Learning-Based Security System Powered by Low Cost Hardware and Panoramic Cameras
Authors	Jesús Benito-Picazo and Enrique Domínguez and Esteban J. Palomo and Ezequiel López-Rubio
Conference	International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC 2019)
Year	2019
GGG Rating	- Work in Progress
CORE Rating	- National: Spain
Status	Published
DOI	https://doi.org/10.1007/978-3-030-19651-6_31
Cite	Benito-Picazo et al. 2019



UNIVERSIDAD
DE MÁLAGA

Appendix B

Resumen en Español

*Cortas sentencias vienen de larga
experiencia.*

Miguel de Cervantes Saavedra

El cerebro humano es la máquina de aprendizaje más compleja, potente y versátil jamás conocida. Por ello, muchos científicos de diversas disciplinas están fascinados por sus estructuras y métodos de procesamiento de información. Debido a la calidad y cantidad de información que se extrae del sentido de la vista, la imagen es uno de los principales canales de información que utiliza el ser humano. Sin embargo, la enorme cantidad de imágenes de vídeo que se generan hoy en día dificulta el procesamiento manual de esos datos de vídeo con la suficiente rapidez. Por ello, los sistemas de visión por computador representan una herramienta fundamental en la extracción de información de las imágenes digitales, así como un gran reto para científicos e ingenieros.

El objetivo principal de esta tesis doctoral es la detección y clasificación automática de objetos en primer plano mediante el análisis de imágenes digitales, utilizando técnicas basadas en redes neuronales artificiales, específicamente diseñadas y optimizadas para ser desplegadas en dispositivos hardware de bajo coste. Este objetivo se complementará con el desarrollo de métodos de estimación del movimiento de individuos mediante el uso de aprendizaje no supervisado y modelos basados en redes neuronales artificiales.

Estos objetivos se han abordado a través de un trabajo de investigación ilustrado en una serie de cuatro publicaciones que han sido seleccionadas para apoyar esta tesis. La primera fue publicada en la revista *Integrated Computer-Aided Engineering (ICAE)* en 2018 y consiste en un sistema de detección de movimiento basado en redes neuronales para cámaras Pan-Tilt-Zoom (PTZ) desplegado en una placa Raspberry Pi. El segundo se publicó en el congreso *International Joint Conference on Neural Networks (IJCNN)*

en 2018 y consiste en un sistema de videovigilancia automático basado en aprendizaje profundo para cámaras PTZ desplegado en hardware de bajo coste. El tercero se publicó en la revista *Integrated Computer-Aided Engineering* en 2020 y consiste en un sistema de detección y clasificación de objetos anómalos en primer plano para cámaras panorámicas, basado en aprendizaje profundo y desplegado también en hardware de bajo coste. Por último, el cuarto trabajo fue publicado en el congreso *International Joint Conference on Neural Networks* en 2020 y consiste en un algoritmo de estimación de la posición de individuos basado en redes neuronales para entornos con regiones prohibidas.

Los resultados obtenidos por el autor en estas publicaciones avalan el trabajo de cuatro años de investigación que se resume en esta memoria de tesis doctoral.

B.1 Introducción

El cerebro humano es la más poderosa, completa y versátil máquina de aprender jamás conocida. Como consecuencia, hoy en día científicos de distintas disciplinas quedan fascinados por sus estructuras y mecanismos de procesamiento de información. La ciencia computacional es una de dichas disciplinas, en tanto en cuanto utiliza el cerebro humano como inspiración para el diseño de muchos de los algoritmos y métodos de aprendizaje computacional más potentes que existen en la actualidad.

Debido a la forma en que el cerebro humano procesa la información proporcionada por el sentido de la vista y la gran cantidad de información que puede extraer de dicha fuente, la imagen es uno de los principales canales de información utilizados por los seres humanos. No obstante, en muchas ocasiones la enorme cantidad de información visual generada dificulta la extracción de información de una imagen a una velocidad tal que dicha tarea resulte útil. Es por esta razón que los sistemas de visión por computador constituyen una herramienta esencial para la extracción de información de una imagen y, al mismo tiempo, uno de los mayores retos para científicos e ingenieros.

Para que un sistema automático extraiga información de una imagen, esta imagen debe ser capturada y almacenada convenientemente en formato digital. A continuación, el sistema debe tener la capacidad de identificar los objetos presentes en la escena y registrar sus trayectorias. Por lo tanto, será fundamental diferenciar el fondo de la imagen de los objetos que se encuentran en primer plano, mediante un proceso llamado segmentación.

Sin embargo, a causa de factores ambientales y técnicos, la capacidad de esos algoritmos para detectar objetos en primer plano no es eficaz al 100%, por lo que, en los últimos años, ha surgido una nueva generación de algoritmos basados en redes neuronales artificiales capaces de realizar la seg-

mentación de objetos en una imagen, alcanzado nuevos niveles de velocidad y precisión gracias a nuevos modelos de redes y a las mejoras de rendimiento logradas mediante su adaptación a hardware de alto rendimiento que permite el entrenamiento rápido de redes con enormes cantidades de datos. El problema es que este nuevo hardware tiene un alto coste y un alto consumo energético, lo que introduce la necesidad de optimizar los algoritmos permitiendo la creación de una nueva generación de sistemas portátiles baratos y eficientes con Inteligencia Artificial (IA) incorporada.

El principal objetivo que se persigue en esta tesis es realizar la detección e identificación de objetos en primer plano, estáticos o en movimiento, en una escena, mediante el análisis de imágenes digitales. Todo ello utilizando técnicas basadas en redes neuronales artificiales, especialmente diseñadas y optimizadas para ser desplegadas en dispositivos de hardware de bajo coste. Este trabajo se va a centrar en tres áreas diferentes:

- El estudio y la utilización de técnicas de aprendizaje no supervisado tales como los mapas autoorganizados como una poderosa herramienta para ser utilizada en tareas de aprendizaje automático, como la estimación de la ubicación de objetos o la compresión de datos.
- El estudio de redes neuronales poco profundas como una primera aproximación a la detección de movimiento en una escena utilizando dispositivos de hardware de bajo coste.
- El estudio y la utilización de redes neuronales convolucionales profundas optimizadas para realizar la detección y clasificación de objetos en primer plano, utilizando dispositivos hardware de bajo coste, con el objetivo de construir sistemas de videovigilancia portátiles y de bajo consumo de energía.

La metodología a seguir para la realización de la investigación que comprende esta tesis viene dictada en gran medida por la naturaleza de dicha investigación. Dadas las exigencias de las aplicaciones a las cuales van destinadas los algoritmos capaces de detectar objetos en primer plano, se debe establecer un equilibrio muy importante en el diseño de los mismos, entre precisión y velocidad de ejecución puesto que se les va a exigir una respuesta rápida que permita tomar decisiones en una escala temporal que se aproximará al tiempo real. Además debemos disponer de mecanismos que permitan cuantificar la bondad de los resultados obtenidos y si el tiempo empleado por los algoritmos para obtenerlos es aceptable.

Esta tesis está estructurada en tres bloques diferentes: el primero detalla los antecedentes y el estado del arte que ha sido utilizado por el autor como punto de partida para el proceso de investigación. El segundo bloque reúne los trabajos más importantes que han sido publicados como resultado del proceso de investigación. El tercer y último bloque de esta tesis presenta las

conclusiones obtenidas del proceso de investigación y los posibles trabajos futuros motivados por dichas conclusiones.

B.2 Estado del arte

En el capítulo dedicado a estado del arte se explican los conceptos básicos, la teoría y el trabajo relacionado con esta tesis doctoral. En primer lugar, se enumeran los fundamentos de los sistemas de detección y clasificación de objetos en primer plano (Dasiopoulou et al. (2005)). También se abordan los diversos problemas que surgen y se proporciona un completo estado del arte, donde se enumeran diferentes técnicas utilizadas para dar solución a estos problemas, tales como los algoritmos de ventana deslizante (Glumov et al. (1995); Li and Lee (2005); Cheng et al. (2014)), las características de Haar (Viola and Jones (2001)), los Histogramas de Gradientes Orientados (HOG) (Dalal and Triggs (2005)) o los Modelos de Partes Deformables (DPM) (Felzenszwalb et al. (2008)). Además, se presenta una introducción a las redes neuronales artificiales (Schmidhuber (2015)) explicando los diferentes tipos utilizados en esta tesis, que serán el Perceptrón Multicapa (Gardner and Dorling (1998)), los Mapas Autoorganizados (Kohonen (1990)) y las redes neuronales convolucionales como base de los sistemas de Aprendizaje Profundo o Deep Learning (Lecun et al. (2015)) aplicados a detección e identificación de objetos.

La segunda parte del estado del arte está dedicada a los fundamentos de la videovigilancia. En ella se hace referencia a la fase de segmentación y los problemas existentes en la fase de modelización de fondo. Además, se presenta una clasificación de los diferentes sistemas automáticos de videovigilancia basados en aprendizaje profundo (Zhang et al. (2019); Liang (2019); Bang et al. (2019); Maeda et al. (2019); Luo et al. (2019); Wang and Bai (2018); Shen et al. (2019)), según los modelos de aprendizaje en los cuales se basan, así como según los dispositivos utilizados para la captura de imágenes y secuencias de vídeo, tales como cámaras PTZ (Dimou et al. (2016); Kim et al. (2019); Son et al. (2019)) o cámaras panorámicas (Boult et al. (2004); Wang and Zhu (2012); Fan and Xu (2019)).

La última parte del capítulo dedicado al estado del arte describe el paradigma de computación distribuida *Edge Computing* y su conveniencia para el diseño de sistemas portátiles con un alto nivel de autonomía energética, gracias al ahorro de ancho de banda y la reducción de tiempos de respuesta que dicho paradigma supone. También se realiza una descripción detallada de diferentes tipos de sistemas de vigilancia basados en hardware de bajo coste Angelov et al. (2017); Dziri et al. (2016); Ortega-Zamorano et al. (2017); Benito-Picazo et al. (2017).

Una de las tareas más importantes que requieren los sistemas automáticos de videovigilancia es la detección e identificación de objetos, pero para

conseguir los mejores resultados posibles, hoy en día estos trabajos implican a menudo procesos que exigen una gran potencia de cálculo y modelos basados en redes neuronales profundas masivamente entrenadas que se ejecutan en potentes GPUs, lo que pone de relieve la importancia de una investigación rigurosa sobre cómo diseñar sistemas de videovigilancia basados en redes neuronales que puedan desplegarse en dispositivos de hardware de bajo coste. Esto nos lleva a los cuatro sistemas de vigilancia que sustentan esta tesis doctoral.

B.3 Trabajos de investigación que apoyan esta tesis

El primero de los trabajos, fue publicado en la revista *Integrated Computer-Aided Engineering*, que ocupa un lugar en el primer cuartil del JCR, y se titula *Motion detection with low cost hardware for PTZ cameras* o, en español, *Detección de movimiento con hardware de bajo coste para cámaras PTZ*. El trabajo descrito en este artículo pertenece a la rama de los sistemas automáticos de videovigilancia y consiste en un sistema automático de detección de movimiento para cámaras PTZ, desplegado en un microordenador Raspberry Pi. Para ello se proponen tres modelos matemáticos de movimiento (uno por cada tipo de movimiento ejecutado por la cámara PTZ) y tres métodos de clasificación, que se utilizan comúnmente en la inteligencia artificial: Un perceptrón multicapa, el algoritmo k-vecinos más cercanos (KNN) y una máquina de vectores de soporte (SVM).

El funcionamiento del sistema será el siguiente: primero, la cámara PTZ aporta fotogramas a un ordenador de escritorio que los procesa fuera de línea para estimar los parámetros de los modelos de detección de movimiento y para realizar el entrenamiento de los clasificadores binarios que se utilizarán. Finalmente, en el momento del despliegue, el microcontrolador funciona de forma autónoma obteniendo fotogramas de la cámara PTZ contabilizando las detecciones de movimiento, que de manera opcional, pueden ser enviadas al ordenador de escritorio.

Con la intención de obtener un sistema barato y energéticamente eficiente, los autores de este trabajo han elegido una Raspberry Pi 3 modelo B como plataforma hardware. En dicha plataforma se ha desplegado un sistema de software compuesto por un módulo de preprocesamiento que implementa los modelos matemáticos referidos anteriormente y tres clasificadores binarios para cada modelo de movimiento. Dichos clasificadores binarios serán un perceptrón multicapa, un algoritmo de k-vecinos más cercanos y una máquina de vectores de soporte.

Después de un proceso de validación cruzada, los resultados indican que es posible obtener buenos niveles de rendimiento de acuerdo con varias medidas bien conocidas, siendo el algoritmo KNN el que destaca en términos de precisión y siendo el clasificador SVM el que obtiene un mejor equilibrio

entre velocidad y precisión. Las pruebas de velocidad indican también que el sistema de detección de movimiento aquí propuesto muestra tiempos de entrenamiento aceptables y cuando se trata de probar el procesamiento de vídeo en tiempo real, siempre alcanza velocidades de procesamiento superiores a 24 fps sin importar el clasificador que se esté utilizando.

El segundo trabajo que apoya esta tesis fue publicado en la edición de 2018 del congreso *International Joint Conference on Neural Networks (IJCNN)* celebrado en Río de Janeiro, Brasil. Este congreso tiene calificación *CORE A* en el ranking *Computing Research and Education Association of Australasia (CORE)*. Titulado *Deep learning-based anomalous object detection system powered by microcontroller for PTZ cameras*, o en español, *Sistema de detección de objetos anómalos basado en aprendizaje profundo y alimentado por microcontrolador para cámaras PTZ*, este trabajo describe el diseño y la implementación de un sistema de videovigilancia basado en aprendizaje profundo, capaz de detectar objetos anómalos estacionarios o en movimiento en primer plano, en flujos de vídeo suministrados por una cámara PTZ, desplegado en un microordenador Raspberry Pi.

Este sistema utiliza una Red Neuronal Convolutiva (CNN) para detectar y caracterizar los objetos presentes en la escena. Además, se ha propuesto un modelo matemático para encuadrar un número fijo de áreas en la imagen proveniente de la cámara, que se utilizarán para alimentar la CNN y realizar un seguimiento de los posibles objetos anómalos encontrados en dichas áreas. Estas *detecciones potenciales*, se van a generar siguiendo una distribución de probabilidad consistente en una mixtura entre una distribución aleatoria y una gaussiana. Consideramos que un determinado objeto es anómalo en un determinado entorno cuando su presencia debe generar una alerta en cualquier sistema de vigilancia instalado en el citado entorno.

El algoritmo de detección de objetos asociado al modelo descrito anteriormente es el siguiente:

1. El conjunto de detecciones activas se inicializa al conjunto vacío.
2. Un nuevo fotograma es tomado del flujo de vídeo procedente de la cámara.
3. Todas las detecciones activas se actualizan según las ecuaciones del modelo matemático. Todas las detecciones actualizadas que caen fuera del fotograma son descartadas y se tornan inactivas.
4. Se genera un conjunto de detecciones potenciales aleatorias y para cada una de ellas, la ventana asociada del fotograma es suministrada a una red neuronal convolutiva. Si la salida resultante revela que es probable que se haya encontrado un objeto, entonces la muestra se inserta en el conjunto de detecciones activas, con un peso que es proporcional a la probabilidad de que un objeto esté realmente allí.

5. Ir al paso 2.

La arquitectura del sistema consta de tres partes: un programa que acepta un flujo continuo de imágenes procedentes de un sistema emulador de cámara PTZ, un programa que implementa el modelo matemático de generación de detecciones potenciales y una red neuronal convolucional que se encargará de detectar y caracterizar los objetos contenidos en cada una de las ventanas generadas por el generador de detecciones potenciales en cada fotograma. Con el fin de facilitar el trabajo y aprovechar la arquitectura multinúcleo del hardware donde se va a implantar, a saber, una Raspberry Pi 3 modelo B, se ha utilizado el framework Microsoft Cognitive Toolkit asistido por la biblioteca Microsoft Embedded Learning Library.

Los experimentos consistieron en contar el número de objetos detectados por el sistema en un vídeo con objetos anómalos procedente de una cámara PTZ. Para ello se realizaron 10 pasadas de 360° a dicho vídeo para un número de ventanas aleatorias que va del 1 al 10 y para cada uno de los dos modelos matemáticos de generación de detecciones potenciales considerados en este documento: la mixtura entre una distribución aleatoria y una gaussiana y una distribución puramente uniforme que se ha utilizado como control. La red neuronal utilizada para el clasificador fue una CNN particular preentrenada diseñada por el equipo de la Biblioteca de Aprendizaje Embebido de Microsoft (ELL), para el conjunto de datos del Desafío de Reconocimiento Visual a Gran Escala 2012 (ILSVRC2012).

Las pruebas de velocidad en la Raspberry Pi consistieron en contar el número de objetos detectados por el sistema realizando 10 pasadas de reconocimiento al vídeo de 360° para un número de ventanas aleatorias que va de 1 a 10. Estas operaciones se realizaron para la mixtura gaussiana-uniforme y la distribución uniforme.

Los resultados obtenidos a partir del proceso de experimentación revelan que, aunque el sistema no es capaz de realizar detección e identificación en tiempo real, sí que es capaz de detectar objetos en primer plano, que pueden estar o no en movimiento, en un tiempo razonable.

El tercer trabajo de investigación original que sustenta esta tesis de doctorado fue publicado en la revista *Integrated Computer-Aided Engineering* en 2020 y propone un sistema automático de videovigilancia basado en aprendizaje profundo y cámaras panorámicas de 360°, debidamente optimizado y una vez más, desplegado en una placa Raspberry Pi. El trabajo se titula *Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras* en inglés, y *Sistema de videovigilancia basado en aprendizaje profundo gestionado por hardware de bajo coste y cámaras panorámicas* en español. Como se ha dicho, en este trabajo se propone un nuevo sistema de videovigilancia para la detección de objetos anómalos en movimiento en imágenes panorámicas. Su principal diferencia con respecto a los demás sistemas de vanguardia radica en su algoritmo de generación probabilística de

ventanas candidatas, o *detecciones potenciales*, para la detección de objetos anómalos que utiliza tres nuevas distribuciones de probabilidad diferentes basadas en mixturas. Además, aunque incorpora un sistema de clasificación basado en redes neuronales de aprendizaje profundo, sigue siendo capaz de ser desplegado en una Raspberry Pi 3 modelo B para lograr un bajo consumo de energía y un bajo coste del hardware. Una vez más, se entiende que un objeto es anómalo si no está asociado a las clases de objetos comúnmente encontrados en la escena. En esta circunstancia, se debe disparar una alarma en el sistema de videovigilancia por cada objeto anómalo detectado.

La base del método de detección de objetos anómalos impulsado por este modelo es un conjunto compuesto por las detecciones que están activas. Este conjunto se asocia a aquellos objetos que han sido recientemente detectados por el dispositivo de vigilancia.

Como ha sido indicado anteriormente, en este trabajo se consideran tres distribuciones homocedásticas multivariantes para implementar el generador probabilístico de detecciones potenciales, a saber, Gaussiana, Student-t y triangular. Estas tres distribuciones han sido elegidas porque son distribuciones multivariantes unimodales, y sus funciones de densidad de probabilidad son relativamente fáciles de evaluar, lo que acelera el cálculo.

El razonamiento detrás de este modelo es que la búsqueda de objetos debe dirigirse hacia aquellas áreas del fotograma entrante donde hay detecciones registradas previamente. Esto se gestiona mediante la distribución multivariante homocedástica. Sin embargo, las otras regiones del fotograma también deben ser consultadas para buscar objetos, a un ritmo menor, lo que se gestiona mediante la distribución uniforme.

A la luz de lo anterior, se puede definir un algoritmo para detectar objetos anómalos con la ayuda de una cámara panorámica. El algoritmo se detalla a continuación:

1. El conjunto de detecciones se inicializa al conjunto vacío.
2. Se carga el siguiente fotograma de la cámara panorámica.
3. Se actualiza el conjunto de detecciones activas. Las detecciones que tras actualizarse quedan fuera de la escena serán borradas porque ya no están activas.
4. De forma aleatoria se genera un conjunto de detecciones potenciales utilizando la distribución de probabilidad correspondiente. Entonces, se localiza la ventana asociada a cada detección potencial y se ajusta su tamaño para que cumpla con el formato requerido por la red neuronal convolucional (CNN). Después, la ventana cuyo tamaño se ha ajustado se suministra a la CNN. Si el vector de salida indica una probabilidad alta de que se haya detectado un objeto, se añade la muestra generada

al conjunto de detecciones activas y se asocia dicha muestra con la probabilidad de que la detección sea fiable.

5. Volver al paso 2.

La detección y clasificación de los objetos en primer plano en las imágenes digitales suele requerir el procesamiento de grandes cantidades de información en períodos cortos. En circunstancias normales, estos trabajos requerirían el uso de una arquitectura de hardware de alto rendimiento integrada por ordenadores rápidos con potentes GPU, de modo que todos los cálculos necesarios se realicen a tiempo para que el sistema pueda llevar a cabo los trabajos con la suficiente rapidez y precisión.

Sin embargo, hay algunas ocasiones en que las condiciones ambientales hacen muy difícil o simplemente inviable la instalación de un sistema de videovigilancia automático que dependa de un hardware de alto rendimiento. Esto llevó a los autores de este trabajo a explorar la posibilidad de diseñar y poner en práctica un sistema capaz de realizar la detección y clasificación de objetos en primer plano en imágenes digitales, pero a una pequeña fracción del precio y del consumo de energía eléctrica que tienen los sistemas tradicionales basados en CNN. Este sistema presentaría una arquitectura que integraría un generador de detecciones potenciales basado en una distribución homocedástica multivariante y un módulo de clasificación basado en una CNN, convenientemente optimizado para lograr de forma local resultados aceptables cuando se despliega en dispositivos hardware de bajo coste.

Teniendo en cuenta la reducida potencia de cálculo del hardware en el que se va a desplegar, los autores han considerado, para el módulo de clasificación de objetos, una red neuronal convolucional basada en la VGG-16 preentrenada por el equipo de Microsoft ELL optimizada de forma que pueda aprovechar la arquitectura multinúcleo de la Raspberry Pi.

La arquitectura del software que controla el sistema consiste en un programa compuesto de dos módulos diferentes. El primero es un módulo que suministra un flujo continuo de imágenes tomadas por una cámara panorámica. El segundo módulo está dedicado a la identificación de los potenciales objetos anómalos que pueden aparecer en la escena observada por la cámara de vigilancia de 360° mediante la red neuronal convolucional que se encargará de procesar la información encontrada en cada una de las ventanas generadas por el generador de detecciones potenciales.

La elección del hardware es un tema crítico cuando se trata de aplicaciones de aprendizaje profundo basadas en microcontroladores. Con lo que el equipo de trabajo se decidió por una Raspberry Pi 3 Modelo B, por su equilibrio entre precio y potencia de cálculo.

Los experimentos consistieron en contar el número de objetos detectados por el sistema realizando 10 pasadas de reconocimiento a 300 fotogramas panorámicos de 1920x960 a partir de seis vídeos de 360° modificados artifi-

cialmente introduciendo respectivamente 1, 2, 3, 5, 7 y 10 objetos en movimiento considerados como anómalos. Dichos objetos pertenecen a categorías que aparecen en el dataset ILSVR2012, a saber: “aegyptian cat”, “golden retriever”, “soccer ball”, “sunglasses”, “laptop”, “sombrero”, “bald eagle”, “banana”, “wall clock” y “chainsaw”.

A continuación se realizaron experimentos para comprobar la velocidad de funcionamiento del sistema en una Raspberry Pi 3 modelo B, que consistieron en contar el número de objetos detectados por el sistema después de realizar 5 pasadas de reconocimiento a 10 fotogramas de vídeo 360° para un número de detecciones potenciales que va desde 1 a 10. Todas estas pruebas fueron realizadas para cada una de las tres mixturas consideradas para la implementación del generador de detecciones potenciales.

Con el objetivo de completar esta sección se incluyó una comparación con uno de los sistemas de detección y reconocimiento de objetos más popular en el estado del arte: El sistema *Tiny-YoloV3* de Joseph Redmon y Ali Farhadi. Para ello, hemos entrenado y probado *Tiny-YoloV3* con nuestro conjunto de datos realizando 10 pasadas de detección con cada uno de los vídeos de 300 fotogramas que conforman el conjunto de datos, es decir, vídeos que contienen 1, 2, 3, 5, 7 y 10 objetos en movimiento que hemos considerado como anómalos en los entornos representados en los vídeos citados. Después de las dichas pruebas, se pueden destacar los siguientes resultados:

- En primer lugar, entre todos los modelos de generación de ventanas presentados en esta investigación, la mixtura triangular-uniforme aparece como la mejor en términos de velocidad y precisión de detección porque, aunque su velocidad de procesamiento no es muy diferente de la de los otros modelos, su rendimiento en la detección de objetos es mejor bajo los parámetros que hemos utilizado para este estudio.
- En segundo lugar, los nuevos modelos de generación de detecciones potenciales descritos en el modelo parecen funcionar mejor que un modelo basado en la distribución uniforme pura.
- En tercer lugar, aunque el sistema de videovigilancia descrito en este documento no es capaz de detectar objetos en tiempo real, sí es capaz de hacerlo a una velocidad máxima de 2 fotogramas por segundo cuando se despliega en una Raspberry Pi 3 modelo B. Por estas razones creemos que nuestra propuesta está justificada en términos de autonomía y relación precio/rendimiento, ya que puede desplegarse en un hardware que cuesta aproximadamente 25\$.
- Por último, los resultados obtenidos en la comparación con el sistema *Tiny-YoloV3* concluyen que la velocidad de procesamiento del *Tiny-YoloV3* en una Raspberry Pi 3 Modelo B es más de 3 veces más lenta

que la velocidad presentada por el sistema basado en el algoritmo probabilístico presentado en este trabajo. Este hecho tiene una notable relevancia para ilustrar el rendimiento de nuestro sistema frente a uno de los sistemas de detección más populares del estado del arte como es Tiny-YoloV3.

Es conveniente recordar que el trabajo desarrollado en esta tesis está orientado a la creación de sistemas de vigilancia automática basados en redes neuronales. Sin embargo, el rendimiento de este tipo de sistemas a menudo depende de nuestras capacidades para desarrollar modelos innovadores de redes neuronales, por lo que el último trabajo original descrito en esta tesis y titulado *Image Clustering Using a Growing Neural Gas with Forbidden Regions* o *Agrupamiento en imágenes utilizando un Gas Neuronal Creciente con Regiones Prohibidas*, tiene una orientación teórica más fuerte que los otros tres, y consiste en un sistema capaz de estimar la posición de diferentes criaturas que habitan en ambientes oceánicos, realizando una operación de agrupamiento de los avistamientos reales de esos animales utilizando una novedosa variación del Gas Neuronal Creciente (GNG) que denominamos *Gas Neuronal Creciente con Regiones Prohibidas* o (FRGNG).

En este trabajo, la base matemática del nuevo algoritmo del FRGNG estará constituida por el conocido algoritmo del GNG. En esencia, el FRGNG está diseñado como un GNG que tiene la capacidad de evitar las regiones prohibidas en el momento de la creación de una nueva neurona y en el momento en que los pesos sinápticos de una neurona existente se modifican como resultado del proceso de entrenamiento. Esta nueva característica ha sido diseñada en base a la capacidad de evitar regiones prohibidas de una variación del mapa autoorganizado llamada *Forbidden Region Self-Organising Map* (FRSOFM), desarrollada por el Dr. Antonio Díaz Ramos y sus colaboradores Ramos et al. (2019), capaz de mantener los prototipos de sus neuronas fuera de determinadas zonas a lo largo de todo el proceso de entrenamiento de la red.

Los cinco conjuntos de datos utilizados para la realización de los experimentos están compuestos por datos biogeográficos que reflejan las coordenadas de latitud y longitud de los avistamientos reales de animales marinos de las categorías “Tiburón Azul”, “Pez Dorado-Delfín”, “Tiburón Tigre”, “Delfín Común” y “Rorcual Común”. Por razones obvias, ningún individuo perteneciente a estas especies marinas puede ser avistado en tierra, por lo que las zonas continentales de las diferentes zonas geográficas constituirán las regiones prohibidas en nuestros experimentos.

El proceso experimental consistió en entrenar una instancia del modelo FRGNG para un número final de 4, 16, 36 y 64 neuronas. Nótese el empleo de la palabra “final” para referirnos al número de neuronas del FRGNG porque este modelo es una red que crece a lo largo del proceso de entrenamiento. Como competidor del modelo del FRGNG presentado en este trabajo, se ha

considerado el modelo FRSOFM anteriormente referido, para el que los experimentos también incluyeron el entrenamiento de una red FRSOFM para 4, 16, 36 y 64 neuronas con topología cuadrada. Los valores para los parámetros del entrenamiento de ambas redes se han mantenido inalterados a lo largo de todo el proceso de experimentación.

Cada algoritmo ha sido probado realizando un proceso de validación cruzada de 10 folds para cada modelo, conjunto de datos y número de neuronas. Mediante este proceso se han obtenido los valores medios de tres medidas de rendimiento de calidad de agrupación comúnmente utilizadas: El Índice de Davies-Bouldin (DBI), el Error Cuadrático Medio (MSE) y el Índice de Dunn. Dichos valores ilustran cómo la calidad del clustering y la fidelidad de representación de la topología lograda por los dos modelos son similares. Sin embargo, el modelo FRSOFM a menudo parece ganar cuando se trata de las medidas del índice DBI y Dunn, mientras que el modelo FRGNG presenta mejores valores de MSE. Este es un comportamiento muy interesante dado que los parámetros del modelo FRGNG no se han optimizado, mientras que los parámetros para el entrenamiento del modelo FRSOFM sí fueron optimizados para todos los conjuntos de datos. Estos resultados postulan el FRGNG como un modelo matemático de gran interés para el agrupamiento de datos o *clustering* y la representación de las relaciones topológicas entre los diferentes clusters.

B.4 Conclusiones y trabajo futuro

B.4.1 Conclusiones

En este capítulo se presentan los resultados obtenidos tras cuatro años de investigación dedicados a la inteligencia artificial. Cuatro años dedicados a estudiar, desarrollar, probar y optimizar nuevos sistemas de vigilancia basados en redes neuronales artificiales para su despliegue en dispositivos de hardware de bajo coste. A lo largo de este trabajo se han utilizado diferentes modelos matemáticos, entre ellos el perceptrón multicapa, la familia de redes neuronales de los mapas autoorganizados y las redes neuronales convolucionales profundas integradas en las populares y potentes técnicas de aprendizaje profundo. Esta investigación está respaldada por cuatro trabajos publicados en revistas de alto factor de impacto y congresos internacionales que atestiguan su pertinencia y novedad. Además, durante el desarrollo de esta tesis, también se han publicado otros trabajos adicionales en revistas con un factor de impacto moderado y conferencias internacionales, presentando desarrollos intermedios de los cuatro trabajos que apoyan dicha tesis.

Cabe destacar que los cuatro trabajos mencionados no están aislados entre sí. Por el contrario, constituyen lo que podría catalogarse como un sistema automático e integral de videovigilancia para hardware de bajo coste,

compuesto por diferentes partes que se encargan de tareas como la detección de movimiento en flujos de vídeo procedentes de cámaras de vigilancia, la detección e identificación de objetos en flujos de vídeo procedentes de cámaras PTZ y 360°, y la predicción del comportamiento de individuos mediante la estimación de su posición actual en un entorno natural a partir de observaciones previas.

El perceptrón multicapa es el primer modelo que se ha estudiado y utilizado en esta tesis para el diseño e implementación de sistemas automáticos de videovigilancia para hardware de bajo coste. Así, en el Capítulo 3 hemos ilustrado la construcción de un detector de movimiento en tiempo real para flujos de vídeo suministrados por cámaras PTZ desplegado en un hardware de bajo coste. El sistema se basa en un algoritmo que procesa una secuencia de imágenes procedentes de una cámara PTZ en movimiento, comparando fotogramas consecutivos para detectar cambios que podrían revelar la presencia de posibles objetos en movimiento en primer plano. El modelo ha sido entrenado por separado en un ordenador de escritorio y se ha desplegado con éxito en un microcomputador Raspberry Pi 3 Modelo B, obteniendo un detector de objetos en primer plano que será el motor de un sistema automático de videovigilancia económico y de bajo consumo.

Los resultados experimentales revelan que el sistema presenta una propuesta adecuada para un sistema automático de videovigilancia barato y de bajo consumo que funciona en tiempo real.

El aprendizaje profundo es una técnica popular de aprendizaje máquina basada en redes neuronales artificiales de múltiples capas. En el trabajo descrito en el Capítulo 4 se ilustra la construcción de un sistema automático de videovigilancia de bajo consumo para flujos de vídeo de cámaras PTZ utilizando aprendizaje profundo optimizado para dispositivos hardware de bajo coste. Con un algoritmo de localización y clasificación de objetos anómalos en primer plano basado en el aprendizaje profundo, el núcleo de este sistema integra dos partes bien diferenciadas. La primera parte es un motor de detección y localización, constituido por un generador de detecciones potenciales basado en una mixtura de distribuciones gaussianas y aleatorias. La segunda parte es un clasificador basado en una red neuronal convolucional, encargado de identificar el objeto que aparece en la ventana enmarcada por el generador de detecciones potenciales.

De nuevo, todo el sistema ha sido optimizado para ser desplegado en un dispositivo de hardware barato y de bajo consumo de energía como la Raspberry Pi 3 Modelo B, utilizando pequeñas redes neuronales convolucionales preentrenadas y bibliotecas de software especiales para aprovechar la estructura multinúcleo de la Raspberry Pi.

Los resultados experimentales revelan que el sistema es capaz de detectar varios objetos anómalos estáticos o en movimiento en una escena a velocidades de un fotograma cada dos segundos, lo que confirma la idoneidad de esta

propuesta para la construcción de sistemas automáticos de videovigilancia baratos y de bajo consumo de energía.

Gracias a su versatilidad y facilidad de uso, las cámaras PTZ han sido un tipo de dispositivo muy recurrente en esta tesis. Sin embargo, después de algunos trabajos, se decidió probar otros dispositivos de captura de imágenes que pudieran ofrecer una imagen completa del entorno que se monitoriza para ser procesada. Así, se consideraron las cámaras de 360° como los siguientes dispositivos de adquisición de imágenes para continuar la investigación. En consecuencia, se han utilizado como el principal dispositivo de captura de imágenes en el trabajo descrito en el Capítulo 5, donde los autores presentan una mejora del trabajo desarrollado en el Capítulo 3, incorporando el uso de cámaras panorámicas de 360° y un nuevo generador de detecciones potenciales basado en tres nuevas distribuciones probabilísticas homocedásticas, para construir un nuevo detector de objetos anómalos en primer plano que se desplegará en una Raspberry Pi 3 modelo B.

El sistema resultante supera a uno de los sistemas más populares e influyentes del estado del arte actual, en términos de capacidad de localización y clasificación, y de rendimiento, cuando se despliega en dispositivos de hardware tipo Raspberry Pi.

Tal y como fue mencionado en el estado del arte, en el Capítulo 2, el aprendizaje profundo no es la única tecnología basada en redes neuronales artificiales que se utiliza con éxito en tareas de vigilancia. Así pues, se consideró la posibilidad de investigar otros tipos de redes neuronales distintas de las redes convolucionales profundas y su utilidad al diseñar sistemas que permitan realizar tareas de vigilancia a cualquier nivel. Con el tiempo, algunas de las tareas de vigilancia mencionadas anteriormente pueden implicar procesos de agrupación para hacer una regresión sobre cualquier conjunto de datos. Una de las muchas tareas para las que los científicos pueden utilizar esos datos consiste en realizar observaciones sobre un grupo de individuos a fin de extrapolar su ubicación geográfica a través del tiempo, de modo que se puedan hacer predicciones sobre su comportamiento. En esta línea, el Capítulo 6 describe una variación del Gas Neuronal Creciente, a saber, llamado Gas Neuronal Creciente con Regiones Prohibidas (o *FRGNG*), capaz de realizar tareas de agrupación que revelen las relaciones topológicas entre esos grupos o *clusters*, manteniendo sus prototipos de neuronas fuera de un conjunto de barreras poligonales convexas. En el caso concreto del trabajo presentado en el Capítulo 6, esta red se utiliza para realizar tareas de clustering sobre ciertos datos biogeográficos procedentes de los avistamientos de individuos pertenecientes a diferentes especies animales marinas alrededor de las costas de determinadas localizaciones geográficas, con el fin de monitorizar y posiblemente predecir sus hábitos migratorios.

Los resultados experimentales revelan que, incluso sin optimización de parámetros, el FRGNG presentado en esta tesis es una buena alternativa

frente al Mapa Autoorganizado con Regiones Prohibidas (o *FRSOM*) en su versión de parámetros optimizados, presentado como el competidor más avanzado en el estado del arte.

El resto de los trabajos más relevantes desarrollados en este período consisten en ponencias en conferencias internacionales que presentan las primeras etapas de desarrollo de algunos de los sistemas citados, atestiguando paso a paso la complejidad y las dificultades encontradas a lo largo de todo el proceso de investigación.

B.4.2 Trabajo Futuro

A lo largo de cualquier período de investigación significativo y riguroso, surgen inevitablemente ideas para futuras investigaciones motivadas por los desafíos enfrentados por los investigadores como parte de los procesos de resolución de problemas y los resultados obtenidos en los experimentos realizados.

Teniendo en cuenta las características de los trabajos presentados para apoyar esta tesis, la investigación futura se divide en tres líneas diferentes:

- Optimización del despliegue en hardware de bajo coste de detectores de movimiento basados en perceptrones multicapa.
- Sistemas de videovigilancia basados en aprendizaje profundo optimizados para ser desplegados en dispositivos de hardware de bajo coste.
- Desarrollo de versiones evolucionadas del Gas Neuronal Creciente con Regiones Prohibidas (FRGNG).

Comenzando con el diseño y desarrollo de sistemas de detección de movimiento basados en el perceptrón multicapa, nos parece muy interesante seguir desarrollando este tipo de redes para construir sistemas de videovigilancia debido a su velocidad y versatilidad.

- Por lo tanto, con el fin de implementar un sistema de detección de movimiento más potente, rápido y completo, los trabajos futuros sugieren el diseño de un nuevo modelo matemático preparando el sistema para el uso de una cámara PTZ capaz de realizar todos los movimientos necesarios para llegar a un lugar específico a una velocidad no constante y al mismo tiempo. De esta forma se obtendría un detector de movimiento más rápido y preciso.
- El otro posible futuro desarrollo que podría mejorar el rendimiento del sistema descrito en el Capítulo 3 debe considerar una nueva implementación específica para múltiples núcleos que aprovecharía toda la potencia de computación suministrada por el dispositivo Raspberry

Pi, acelerando el rendimiento del sistema hasta casi cuadruplicar la velocidad actual. Por supuesto, este nuevo uso del hardware llevaría al desarrollo e instalación de un nuevo sistema de ventilación adicional para prevenir un posible sobrecalentamiento del hardware.

Los sistemas basados en el aprendizaje profundo son una tecnología muy popular y, a pesar de su alto estado de desarrollo actual, siguen siendo una tecnología muy prometedora en el aprendizaje computacional, como puede comprobarse en trabajos como el presentado en el Capítulo 4 de esta tesis. En él se desarrolla un sistema de videovigilancia para la detección de objetos anómalos en escenas filmadas por una cámara PTZ, lo que hace que este trabajo también nos lleve a proponer algunas líneas de investigación futuras.

- La primera línea de investigación tiene por objeto la elaboración de vídeos filmados por medio de una cámara PTZ en movimiento que contengan uno o más objetos considerados anómalos para la escena. De esta forma se obtendría un conjunto de datos adecuado de cámaras PTZ con objetos anómalos que sería compartido con la comunidad científica.
- La segunda línea de investigación futura persigue la mejora del modelo de detección de objetos mediante la creación de redes más pequeñas y especializadas que puedan clasificar la cantidad adecuada de categorías de objetos anómalos que requiere cada entorno. De esta manera, el reconocimiento de objetos será más potente y al mismo tiempo, la velocidad de procesamiento en fotogramas por segundo aumentaría, obteniendo un sistema más eficaz.

En el trabajo recogido en el Capítulo 5 se detalla un sistema automático de videovigilancia para la detección de objetos anómalos alimentado por cámaras panorámicas de 360° y desplegado en hardware de bajo coste. Este sistema presenta importantes mejoras respecto al sistema diseñado en el artículo descrito en el Capítulo 4, como el uso de una cámara panorámica de 360°, que proporciona a todo el sistema una visión esférica completa. No obstante los resultados experimentales sugieren varios posibles desarrollos que pueden agruparse en tres futuras líneas de investigación:

- La primera consiste en la elaboración propia de vídeos panorámicos de entornos reales con objetos anómalos en su interior, para crear un conjunto de datos público. De esta manera, podría compartirse con la comunidad científica para ayudar a otros investigadores a diseñar y probar sus modelos.
- La segunda línea de investigación y desarrollo futura consiste en el diseño de una nueva red, más pequeña y específicamente entrenada para

clasificar sólo cierta cantidad de categorías que se consideran anómalas para cada entorno. De esta manera, obtendríamos una red más pequeña que mejoraría nuestro sistema de videovigilancia haciéndolo más preciso y rápido cuando se despliega en un dispositivo hardware de bajo coste.

- La tercera línea de investigación futura inspirada en este trabajo consiste en la optimización de los parámetros de las tres distribuciones homocedásticas multivariantes que constituyen los fundamentos del modelo probabilístico utilizado para implementar el generador de detecciones potenciales. Por lo tanto, con el fin de obtener el máximo rendimiento de este sistema, se utilizará el nodo *Picasso* de la RES (Red Española de Supercomputación) para realizar un proceso de optimización sistemática de dichos parámetros de manera oportuna.

Las últimas líneas de investigación motivadas por los diferentes trabajos desarrollados en esta tesis provienen del trabajo descrito en el Capítulo 6, donde se desarrolla un modelo basado en el Gas Neuronal Creciente, llamado Gas Neuronal Creciente con Regiones Prohibidas, para el seguimiento y la predicción del comportamiento migratorio de las especies animales.

Teniendo en cuenta el rendimiento del modelo matemático presentado en el Capítulo 6, las futuras líneas de investigación inspiradas en este trabajo podrían seguir cuatro caminos diferentes.

- En primer lugar, se ha considerado la posibilidad de realizar pruebas en otro tipo de entornos, en los que el FRGNG podría emplearse para estimar patrones de observación en otros tipos de eventos, como las zonas de alta concentración de accidentes en entornos industriales, los patrones de movimiento de personas en eventos recreativos multitudinarios (como espectáculos musicales o eventos deportivos), la evolución de los agentes infecciosos en la población humana y cualquier otro tipo de evento susceptible de ser tratado con técnicas de agrupamiento con zonas prohibidas.
- Es muy importante recordar que los resultados obtenidos por el FRGNG en los ensayos realizados se obtuvieron con sus parámetros fijados a valores por defecto. Por lo tanto, la segunda línea de investigación futura inspirada en este trabajo consistiría en un proceso de optimización destinado a realizar una meticulosa puesta a punto de los parámetros del FRGNG, apoyada por el supercomputador *Picasso*, con el fin de obtener los mejores resultados para este modelo.
- El modelo FRGNG es capaz de realizar tareas de clustering a partir de datos biogeográficos manteniendo los prototipos de sus neuronas fuera de algunas barreras que conocemos como regiones prohibidas.

Sin embargo, este modelo se limita a distribuciones de datos bidimensionales. Así pues, el último y más ambicioso de los posibles desarrollos derivados de este trabajo consistirá en el diseño y construcción de una versión tridimensional del Gas Neural Creciente con Regiones Prohibidas, a saber, un Gas Neural Creciente con Regiones Prohibidas 3D o FRGNG3D. Este modelo tendrá la capacidad de realizar tareas de agrupación en condiciones similares a las del FRGNG, pero con distribuciones de datos tridimensionales, lo que aportará generalidad a nuestro modelo facilitando su utilización en una gama más amplia de posibles aplicaciones.

Bibliography

*Imagination is the Discovering Faculty,
pre-eminently. It is that which penetrates
into the unseen worlds around us, the
worlds of Science.*

Ada Byron Lovelace

- Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Computers in Biology and Medicine*, 100:270 – 278.
- Adeli, H. and Wu, M. (1998). Regularization neural network for construction cost estimation. *Journal of Construction Engineering and Management*, 124(1):18–23.
- Adnan, L., Yussoff, Y., Johar, H., and Baki, S. (2015). Energy-saving street lighting system based on the waspmote mote. *Jurnal Teknologi*, 76(4):55–58.
- Allebosch, G., Van Hamme, D., Veelaert, P., and Philips, W. (2019). Robust pan/tilt compensation for foreground-background segmentation. *Sensors*, 19(12):27.
- Angelov, P., Sadeghi-Tehran, P., and Clarke, C. (2017). AURORA: Autonomous real-time on-board video analytics. *Neural Comput. Appl.*, 28(5):855–865.
- Ansari, A. H., Cherian, P. J., Caicedo, A., Naulaers, G., De Vos, M., and Van Huffel, S. (2019). Neonatal seizure detection using deep convolutional neural networks. *International journal of neural systems*, 29(4).
- Antoniades, A., Spyrou, L., Martin-Lopez, D., Valentin, A., Alarcon, G., Sanei, S., and Took, C. C. (2018). Deep neural architectures for mapping scalp to intracranial eeg. *International journal of neural systems*, 28(8).

- Arriola, Y. and Carrasco, R. A. (1990). Integration of multi-layer perceptron and markov models for automatic speech recognition. In *IEE Conference Publication*, pages 413–420.
- Atkinson, P. M. and Tatnall, A. R. L. (1997). Introduction neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4):699–709.
- Bang, S., Park, S., Kim, H., and Kim, H. (2019). Encoder–decoder network for pixel-level road crack detection in black-box images. *Computer-Aided Civil and Infrastructure Engineering*, 34(8):713–727.
- Bauer, H. U. and Villmann, T. (1997). Growing a hypercubical output space in a self-organizing feature map. *IEEE transactions on neural networks*, 8(2):218–26.
- Benito-Picazo, J., Dominguez, E., Palomo, E. J., Lopez-Rubio, E., and Ortiz-De-Lazcano-Lobato, J. M. (2018a). Deep learning-based anomalous object detection system powered by microcontroller for ptz cameras. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2018-July.
- Benito-Picazo, J., Domínguez, E., Palomo, E. J., and López-Rubio, E. (2019). *Deep Learning-Based Security System Powered by Low Cost Hardware and Panoramic Cameras*, volume 11487 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Benito-Picazo, J., Domínguez, E., Palomo, E. J., and López-Rubio, E. (2020a). Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras. *Integrated Computer-Aided Engineering*, 27(4):373–387.
- Benito-Picazo, J., Domínguez, E., Palomo, E. J., López-Rubio, E., and Ortiz-De-Lazcano-Lobato, J. M. (2018b). Motion detection with low cost hardware for ptz cameras. *Integrated Computer-Aided Engineering*, 26(1):21–36.
- Benito-Picazo, J., López-Rubio, E., and Domínguez, E. (2017a). Growing neural forest-based color quantization applied to RGB images. *International Journal of Computer Vision and Image Processing*, 7(3):13–25.
- Benito-Picazo, J., López-Rubio, E., Ortiz-de-Lazcano-Lobato, J. M., Domínguez, E., and Palomo, E. J. (2017b). Motion detection by microcontroller for panning cameras. In de Vicente, J. M. F., Sánchez, J. R. Á., de la Paz López, F., Toledo-Moreo, F. J., and Adeli, H., editors, *Biomedical Applications Based on Natural and Artificial Computing -*

- International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2017, Corunna, Spain, June 19-23, 2017, Proceedings, Part II*, volume 10338 of *Lecture Notes in Computer Science*, pages 279–288. Springer.
- Benito-Picazo, J., López-Rubio, E., Ortiz-De-lazcano lobato, J. M., Domínguez, E., and Palomo, E. J. (2017). *Motion detection by micro-controller for panning cameras*, volume 10338 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Benito-Picazo, J., Palomo, E. J., Dominguez, E., and Ramos, A. D. (2020b). Image clustering using a growing neural gas with forbidden regions. In *Proceedings of the International Joint Conference on Neural Networks*.
- Blackmore, J. and Miikkulainen, R. (1993). Incremental grid growing: encoding high-dimensional structure into a two-dimensional feature map. In *IEEE International Conference on Neural Networks*, pages 450–455. IEEE.
- Boult, T., Gao, X., Micheals, R., and Eckmann, M. (2004). Omni-directional visual surveillance. *Image and Vision Computing*, 22(7):515–534.
- Chen, C., Li, S., Qin, H., and Hao, A. (2016). Robust salient motion detection in non-stationary videos via novel integrated strategies of spatio-temporal coherency clues and low-rank analysis. *Pattern Recognition*, 52:410 – 432.
- Cheng, G., Han, J., Guo, L., Qian, X., Zhou, P., Yao, X., and Hu, X. (2013). Object detection in remote sensing imagery using a discriminatively trained mixture model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 85:32–43.
- Cheng, M.-M., Zhang, Z., Lin, w.-y., and Torr, P. (2014). Bing: Binarized normed gradients for objectness estimation at 300fps. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3286–3293.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, volume I, pages 886–893.
- Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V. ., and Strintzis, M. G. (2005). Knowledge-assisted semantic video object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(10):1210–1224.

- Dimou, A., Medentzidou, P., García, F. A., and Daras, P. (2016). Multi-target detection in cctv footage for tracking applications using deep learning techniques. In *Proceedings - International Conference on Image Processing, ICIP*, volume 2016-August, pages 928–932.
- Ding, C., Bappy, J. H., Farrell, J. A., and Roy-Chowdhury, A. K. (2017). Opportunistic image acquisition of individual and group activities in a distributed camera network. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3):664–672.
- Ding, C., Song, B., Morye, A., Farrell, J., and Roy-Chowdhury, A. (2012). Collaborative sensing in a distributed PTZ camera network. *IEEE Transactions on Image Processing*, 21(7):3282–3295.
- Dobrzynski, M. K., Pericet-Camara, R., and Floreano, D. (2012). Vision tape-a flexible compound vision sensor for motion detection and proximity estimation. *IEEE Sensors Journal*, 12(5):1131–1139.
- Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York.
- Dziri, A., Duranton, M., and Chapuis, R. (2016). Real-time multiple objects tracking on raspberry-pi-based smart embedded camera. *Journal of Electronic Imaging*, 25:041005.
- Déniz, O., Bueno, G., Salido, J., and De La Torre, F. (2011). Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 32(12):1598–1603.
- Fan, Q. and Xu, Y. (2019). A robust target recognition and tracking panoramic surveillance system based on deep learning. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 11342.
- Felzenszwalb, P., McAllester, D., and Ramanan, D. (2008). A discriminatively trained, multiscale, deformable part model. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*.
- Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645.
- Ferone, A. and Maddalena, L. (2014). Neural background subtraction for pan-tilt-zoom cameras. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(5):571–579.
- Freund, Y. and Schapire, R. E. (1995). *A decision-theoretic generalization of on-line learning and an application to boosting*, volume 904 of *Lecture*

- Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).*
- Fritzke, B. (1994). Growing cell structures - A self-organizing network for unsupervised and supervised learning. *Neural Networks*, 7(9):1441–1460.
- Fritzke, B. (1995a). Growing Grid - a self-organizing network with constant neighborhood range and adaptation strength. *Neural Processing Letters*, 2(5):9–13.
- Fritzke, B. (1995b). A growing neural gas network learns topologies. *Advances in Neural Information Processing Systems*, 7:625–632.
- Fung, V., Bosch, J. L., Roberts, S. W., and Kleissl, J. (2014). Cloud shadow speed sensor. *Atmospheric Measurement Techniques*, 7(6):1693–1700.
- Gandhi, T. and Trivedi, M. M. (2004). Motion analysis for event detection and tracking with a mobile omnidirectional camera. *Multimedia Systems*, 10(2):131–143.
- Gardner, M. W. and Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15):2627–2636.
- Glumov, N. I., Kolomiyetz, E. I., and Sergeyev, V. V. (1995). Detection of objects on the image using a sliding window mode. *Optics and Laser Technology*, 27(4):241–249.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.
- Hua, C., Wang, H., Lu, S., Liu, C., and Khalid, S. M. (2019). A novel method of building functional brain network using deep learning algorithm with application in proficiency detection. *International journal of neural systems*, 29(1).
- Hua, C., Wang, H., Lu, S., Liu, C., and Khalid Syed, M. (2018). A novel method of building functional brain network using deep learning algorithm with application in proficiency detection. *International Journal of Neural Systems*.
- Jin, K. H., McCann, M. T., Froustey, E., and Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522.
- Kazmierczak, H. (1978). Contour and object extraction from image data. *Proceedings of SPIE - The International Society for Optical Engineering*, 130:53–60.

- Kim, D., Kim, K., and Park, S. (2019). Automatic ptz camera control based on deep-q network in video surveillance system. In *ICEIC 2019 - International Conference on Electronics, Information, and Communication*.
- Kohli, N., Yadav, D., and Noore, A. (2018). Face verification with disguise variations via deep disguise recognizer. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2018-June, pages 17–24. Cited By :12.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480.
- Komagal, E. and Yogameena, B. (2018). Foreground segmentation with ptz camera: a survey. *Multimedia Tools and Applications*, 77:22489–22542.
- Koziarski, M. and Cyganek, B. (2017). Image recognition with deep neural networks in presence of noise - dealing with and taking advantage of distortions. *Integrated Computer-Aided Engineering*, 24:337–349.
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Leung, W. F., Leung, S. H., Lau, W. H., and Luk, A. (1991). Fingerprint recognition using neural network. In *Neural Networks for Signal Processing*, pages 226–235.
- Li, S. and Lee, M. C. (2005). An improved sliding window method for shot change detection. In *Proceedings of the Seventh IASTED International Conference on Signal and Image Processing, SIP 2005*, pages 464–468.
- Liang, X. (2019). Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with bayesian optimization. *Computer-Aided Civil and Infrastructure Engineering*, 34(5):415–430.
- Lienhart, R. and Maydt, J. (2002). An extended set of haar-like features for rapid object detection. In *IEEE International Conference on Image Processing*, volume 1, pages I/900–I/903.
- Lin, L., Wang, X., Yang, W., and Lai, J. . (2015). Discriminatively trained and-or graph models for object shape detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(5):959–972.
- Liu, N., Zhang, M., Li, H., Sun, Z., and Tan, T. (2016). Deepiris: Learning pairwise filter bank for heterogeneous iris verification. *Pattern Recognition Letters*, 82:154–161. Cited By :83.
- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234(November 2016):11–26.

- Luo, X., Li, H., Yang, X., Yu, Y., and Cao, D. (2019). Capturing and understanding workers' activities in far-field surveillance videos with deep action recognition and bayesian nonparametric learning. *Computer-Aided Civil and Infrastructure Engineering*, 34(4):333–351.
- Maeda, K., Ogawa, T., Haseyama, M., , and Takahashi, S. (2019). Convolutional sparse coding-based deep random vector functional link network for distress classification of road structures. *Computer-Aided Civil and Infrastructure Engineering*, 34(8):654–676.
- Martinetz, T. and Schulten, K. (1991). A "neural-gas" network learns topologies. *Artificial neural networks*, 1:397–402.
- McCann, M. T., Jin, K. H., and Unser, M. (2017). Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95.
- McCulloch, N., Ainsworth, W. A., and Linggard, R. (1988). Multi-layer perceptrons applied to speech technology. *British Telecom technology journal*, 6(2):131–139.
- Meng, L., Hirayama, T., and Oyanagi, S. (2018). Underwater-drone with panoramic camera for automatic fish recognition based on deep learning. *IEEE Access*, 6:17880–17886.
- Meng, Z., Hu, Y., and Ancey, C. (2020). Using a data driven approach to predict waves generated by gravity driven mass flows. *Water*, 12.
- Micheloni, C., Rinner, B., and Foresti, G. (2010). Video analysis in pan-tilt-zoom camera networks. *IEEE Signal Processing Magazine*, 27(5):78–90.
- Mita, T., Kaneko, T., and Hori, O. (2005). Joint haar-like features for face detection. In *Proceedings of the IEEE International Conference on Computer Vision*, volume II, pages 1619–1626.
- Ortega-Zamorano, F., Jerez, J. M., Gómez, I., and Franco, L. (2017). Layer multiplexing fpga implementation for deep back-propagation learning. *Integrated Computer-Aided Engineering*, 24(2):171–185.
- Ortega-Zamorano, F., Molina-Cabello, M. A., López-Rubio, E., and Palomo, E. J. (2016). Smart motion detection sensor based on video processing using self-organizing maps. *Expert Systems with Applications*, 64:476 – 489.
- Palomo, E. J., Benito-Picazo, J., López-Rubio, E., and Domínguez, E. (2017). Unsupervised color quantization with the growing neural forest. In

- Rojas, I., Joya, G., and Català, A., editors, *Advances in Computational Intelligence - 14th International Work-Conference on Artificial Neural Networks, IWANN 2017, Cadiz, Spain, June 14-16, 2017, Proceedings, Part II*, volume 10306 of *Lecture Notes in Computer Science*, pages 306–316. Springer.
- Palomo, E. J. and Lopez-Rubio, E. (2017). The growing hierarchical neural gas self-organizing neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 28(9):2000–2009.
- Perdana, A. B. and Prahara, A. (2019). Face recognition using light-convolutional neural networks based on modified vgg16 model. In *2019 International Conference of Computer Science and Information Technology, ICOSNIKOM 2019*. Cited By :3.
- Rafiei, M. H. and Adeli, H. (2017). A novel machine learning-based algorithm to detect damage in high-rise building structures. *The Structural Design of Tall and Special Buildings*, 26(18):e1400.
- Rafiei, M. H. and Adeli, H. (2018). Novel machine-learning model for estimating construction costs considering economic variables and indexes. *Journal of Construction Engineering and Management*, 144(12):04018106.
- Rafiei, M. H., Khushefati, W., Demirboga, R., and Adeli, H. (2017). Supervised deep restricted boltzmann machine for estimation of concrete. *ACI Materials Journal*.
- Ramos, A. D., López-Rubio, E., and Palomo, E. J. (2019). The forbidden region self-organizing map neural network. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–11.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection.
- Sajid, H., Cheung, S.-C. S., and Jacobs, N. (2016). Appearance based background subtraction for PTZ cameras. *Signal Processing: Image Communication*, 47:417 – 425.
- Sato, Y., Hashimoto, K., and Shibata, Y. (2008). A new networked surveillance video system by combination of omni-directional and network controlled cameras. In Takizawa, M., Barolli, L., and Enokido, T., editors, *Network-Based Information Systems*, pages 313–322, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117.

- Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June-2015, pages 815–823. Cited By :4492.
- Scotti, G., Marcenaro, L., Coelho, C., Selvaggi, F., and Regazzoni, C. S. (2005). Dual camera intelligent sensor for high definition 360 degrees surveillance. *IEE Proceedings - Vision, Image and Signal Processing*, 152(2):250–257.
- Shen, J., Xiong, X., Xue, Z., and Bian, Y. (2019). A convolutional neural network-based pedestrian counting model for various crowded scenes. *Computer-Aided Civil and Infrastructure Engineering*, 34(10):897–914.
- Son, K. ., Yildirim, M. E., Park, J. ., and Song, J. . (2019). Flood detection by using fcn-alexnet. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 11041.
- Song, K.-T. and Tai, J.-C. (2006). Dynamic calibration of pan-tilt-zoom cameras for traffic monitoring. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(5):1091–1103.
- Sun, Y., Wang, X., and Tang, X. (2016). Hybrid deep learning for face verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):1997–2009. Cited By :63.
- Tang, J., Deng, C., and Huang, G. . (2016). Extreme learning machine for multilayer perceptron. *IEEE Transactions on Neural Networks and Learning Systems*, 27(4):809–821.
- Tong, L., Dai, F., Zhang, D., Wang, D., and Zhang, Y. (2014). Encoder combined video moving object detection. *Neurocomputing*, 139:150–162.
- Troxel, S. E., Rogers, S. K., and Kabrisky, M. (1988). Use of neural networks in psri target recognition. pages 593–600.
- Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages I511–I518.
- Wang, P. and Bai, X. (2018). Regional parallel structure based cnn for thermal infrared face identification. *Integrated Computer-Aided Engineering*, 25:1–14.
- Wang, T. and Zhu, Z. (2012). Real time moving vehicle detection and reconstruction for improving classification. In *Proceedings of IEEE Workshop on Applications of Computer Vision*, pages 497–502.

- Wei, Y., Tian, Q., Guo, J., Huang, W., and Cao, J. (2019). Multi-vehicle detection algorithm through combining harr and hog features. *Mathematics and Computers in Simulation*, 155:130–145.
- Wen, X., Shao, L., Xue, Y., and Fang, W. (2015). A rapid learning algorithm for vehicle classification. *Information Sciences*, 295:395–406.
- Yagi, Y. (1999). Omnidirectional sensing and its applications. *IEICE Transactions on Information and Systems*, E82-D(3):568–579.
- Zhang, A., Wang, K. C. P., Fei, Y., Liu, Y., Chen, C., Yang, G., Li, J. Q., Yang, E., and Qiu, S. (2019). Automated pixel-level pavement crack detection on 3d asphalt surfaces with a recurrent neural network. *Computer-Aided Civil and Infrastructure Engineering*, 34(3):213–229.
- Zhang, L., Chu, R., Xiang, S., Liao, S., and Li, S. Z. (2007). *Face detection based on multi-block LBP representation*, volume 4642 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Zhu, Q., Avidan, S., Yeh, M. ., and Cheng, K. . (2006). Fast human detection using a cascade of histograms of oriented gradients. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1491–1498.

*The cosmos is within us.
We are made of star-stuff.
We are a way for the universe to know itself.*

Carl Sagan



UNIVERSIDAD
DE MALAGA