Classic versus deep learning approaches to address computer vision challenges: A study of faint edge detection and multispectral image registration

By

YEHONATAN NATI OFIR

Department of Computer Science and Mathematics
KINGSTON UNIVERSITY LONDON

A dissertation submitted in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science, Engineering, and Computing.

JULY 6, 2021

Supervisor:

Professor Jean-Christophe Nebel
Kingston University, London

Email: J.Nebel@kingston.ac.uk

Abstract

COMPUTER Vision involves many challenging problems. While early work utilized classic methods, in recent years solutions have often relied on deep neural networks. In this study, we explore those two classes of methods through two applications that are at the limit of the ability of current computer vision algorithms, i.e., faint edge detection and multispectral image registration. We show that the detection of edges at a low signal-to-noise ratio is a demanding task with proven lower bounds. The introduced method processes straight and curved edges in nearly linear complexity. Moreover, performance is of high quality on noisy simulations, boundary datasets, and real images. However, in order to improve accuracy and runtime, a deep solution was also explored. It utilizes a multiscale neural network for the detection of edges in binary images using edge preservation loss. The second group of work that is considered in this study addresses multispectral image alignment. Since multispectral fusion is particularly informative, robust image alignment algorithms are required. However, as this cannot be carried out by single-channel registration methods, we propose a traditional approach that relies on a novel edge descriptor using a feature-based registration scheme. Experiments demonstrate that, although it is able to align a wide field of spec-

tral channels, it lacks robustness to deal with every geometric transformation. To that end, we developed a deep approach for such alignment. Contrarily to the previously suggested edge descriptor, our deep approach uses an invariant representation for spectral patches via metric learning that can be seen as a teacher-student method. All those pieces of work are reported in five published papers with state-of-the-art experimental results and proven theory. As a whole, this research reveals that, while traditional methods are rooted in theoretical principles and are robust to a wide field of images, deep approaches are faster to run and achieve better performance if, not only sufficient training data are available, but also they are of the same image type as the data on which they are applied.

Acknowledgement

THE author wishes to thank his wife Shir and his family for the support over the years of work. I would like to express my sincere thanks to my supervisor Professor Jean-Christophe Nebel. The author also wishes to thank the co-authors of his publications without whose support this work could not have been done: Professor Ronen Basri, Doctor Sharon-Duvdevani Bar, and Professor Yosi Keller.

Contents

Abstract **iii**

Acknowledgement **v**

Author's declaration **xiv**

List of Publications **xv**

Glossaries **xvii**

1 Introduction **1**
1.1 Computer vision 2
1.2 Faint edge detection 3
1.3 Multispectral registration 4
1.4 Deep learning in computer vision 5
1.5 Aims and objectives 7
1.5.1 Classic detection of faint edges using hierarchical partitioning 8
1.5.2 CNN-based detection of faint edges relying on edge preservation loss 8
1.5.3 Classic registration of multispectral images using a novel edge descriptor . 9

1.5.4 Deep registration of multispectral images relying on invariant descriptor learning 9
1.5.5 Deep learning solutions versus classic methods 9
1.6 Scientific contribution 10
1.6.1 Classic detection of faint edges 10
1.6.2 CNN-based detection of faint edges . . 12
1.6.3 Classic registration of multispectral images 14
1.6.4 Deep registration of multispectral images 15
1.6.5 Comparison between classic and deep learning-based approaches 17
1.7 Thesis outline 19

2 Literature review **20**
2.1 Faint edge detection 20
2.2 Multispectral registration 21
2.3 Deep learning solutions versus classic methods 22
2.3.1 Classic methods 23
2.3.2 Comparison based on specific applications 24
2.3.3 Energy, computing resources and associated hardware requirements 27
2.3.4 Adversarial deep learning 29
2.3.5 Black box effect and development process 30
2.4 Literature review summary 31

3 Faint edge detection **33**

4 Multispectral image registration **66**

5 Differences between classic and deep approaches **78**
5.1 Introduction 78
5.2 Comparison of deep learning and classic methods for faint edge detection 80
5.3 Comparison of deep learning and classic methods for multispectral image registration 86
5.4 Conclusion 89

6 Conclusions **91**
6.1 Summary of contributions 91
6.2 Discussion 92
6.3 Closing remarks 93

List of Figures

1.1 Rectangle-Partition-Tree of an image. My classic method for detection of faint edges search for every curve between every two boundary points $\forall p_1, p_2$ the best concatenation of sub-curves by breaking point p_3. This search for curves is done recursively in a bottom-up dynamic-programming-like approach. For an image with N pixels, the complexity is $O(N^{1.5})$ in the full mode and $O(N \log N)$ in the optimized mode. 10
1.2 My FED-CNN network of a U-Net Architecture is the multi scale CNN that I used to mimic the hierarchical tree approach of my classic algorithm. This deep neural network was trained by an edge preservation loss. 12
1.3 Multispectral patches and their corresponding edge maps that are part of their edge descriptors. Although the original patches are not correlated their edge maps attain a significant similarity. Therefore, this kind of descriptor is more invariant to different spectra than descriptors of inter-spectral alignment. 15

1.4 My learning architecture for a deep invariant descriptor. This can be seen as a pseudo-Siamese network or as a teacher-student scheme. The visible color patch is forwarded through a pre-trained classification network that was trained on the CIFAR10 dataset. Its corresponding Near-Infrared (NIR) patch is being used to train the infrared network to produce a similar invariant representation. 16

5.1 Simulation of faint edges detection in noisy image. The strict F-score graph along the different signal-to-noise ratios from 0 to 2. Our methods, FED-CNN and FastEdges which are geared for faint edges detection, achieve best performance. 81

5.2 Simulation results of my classic FED with the noisy BSDS-500 images [46]. Left: Precision vs. Recall (PR) of contour detection by various algorithms. Right: Performance table. ODS refers to the F-Measure at the optimal threshold across the entire dataset, OIS to the best per image F-measure, and AP to the area under the PR curve. 84

5.3 Evaluation of registration error across simulated scaling transformations. Left: error of the translation parameters when solving scales from 0.9 to 1.1. Right: error of the scaling parameter across the same range of scalings between the cross-spectral images. The translation error is around 1 pixels while the scaling error is negligible. 87

List of Tables

1.1 Topics of the papers presented at CVPR 2019 and their respective percentage. 6

2.1 A table comparing the advantages of deep learning versus classic methods as reported by the literature. 26

5.1 F-score of the methods at SNR 1 and 2. At both SNRs my methods achieve the best scores. . . . 81

5.2 Run time in milli-seconds of the different methods of edge detection. My runtime is very close to Canny's time and is order of magnitude faster than FastEdges. I achieve this improvement mainly by running our network on a GPU, which is a hardware optimization. The advantage of CNN approaches over classic methods like FastEdges [51] is that they are easily implemented and accelerated on a GPU. In addition, due to the simplicity of FED-CNN, its runtime on a CPU is also faster than FastEdges [51]. 82

5.3 Classification accuracy of the different methods and architectures averaged on all noisy levels. My approach of multi-scale preprocessing by IDCNN [72] and classification by resnet 20 [69] achieves the highest accuracy. 85
5.4 Quantitative PSNR(dB) and SSIM results of denoising on the noisy natural images from the dataset of [46]. My approach, using IDCNN for denoising achieves the high perceptual score of SSIM [74]. In addition, my distortion score of PSNR is also competitive relative to the state-of the art approach of DnCNN [78]. My edge preserving auxiliary loss improve the performance of IDCNN in denoising in both measurements. The highest SSIM scores are highlighted. 85
5.5 The average PSNR(dB) and SSIM results of different methods on the BSD68 dataset [61]. The highest SSIM scores are highlighted. . . . 86
5.6 Error in pixels of multi-spectral registration when searching for translation only. My deep method is compared to edge descriptors approach, correlation of Canny [14], correlation of Sobel [24], maximization of mutual-information and LGHD [1]. As can be seen (in bold), my deep-algorithm achieves the highest accuracy. 87
5.7 Comparison of the features of DL and classic approaches observed in this study. 89

Author's declaration

HEREBY, I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Yehonatan Nati Ofir July 6, 2021

List of Publications

1. Ofir N., Galun M., Nadler B., Basri R. (2016). Fast detection of curved edges at low SNR. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (pp. 213-221).

2. Ofir N., Galun, M., Alpert, S., Brant A., Nadler B., Basri R. (2019). On detection of faint edges in noisy images. IEEE transactions on pattern analysis and machine intelligence, 42(4), 894-908.

3. Ofir N., Keller Y. (2021, January). Multi-scale Processing of Noisy Images using Edge Preservation Losses. In 2021 25th IEEE International Conference on Pattern Recognition. (ICPR) IEEE.

4. Ofir N., Silberstein, S., Rozenbaum D., Duvdevani Bar S. (2018, October). Registration and fusion of multi-spectral images using a novel edge descriptor. In 2018 25th IEEE International Conference on Image Processing. (ICIP) IEEE, (pp. 1857-1861).

5. Ofir N., Silberstein, S., Levi H., Rozenbaum D., Duvdevani Bar S. (2018, October). Deep Multi-Spectral Registration Using Invariant Descriptor Learning. In 2018

25th IEEE International Conference on Image Processing. (ICIP) IEEE, (pp. 1238-1242).

Glossaries

1. **ANN** Artificial Neural Network
2. **BSDS** Berkely Segmentation Dataset
3. **CIFAR** Canadian Institute For Advanced Research
4. **CNN** Convolutional Neural Network
5. **CV** Computer Vision
6. **CPU** Central Processing Unit
7. **CVPR** Computer Vision and Pattern Recognition Conference
8. **DL** Deep Learning
9. **FED** Faint Edge Detection
10. **FED-CNN** Faint Edge Detection Convolutional Neural Network
11. **FFT** Fast Fourier Transform
12. **FLOPS** Floating Operations Per Second
13. **GAN** Generative Adverserial Networks
14. **GPU** Graphic Processing Unit

15. **HED** Holistically Edge Detection
16. **IR** Infrared
17. **IDCNN** Image Denoisin Convolutional Neural Network
18. **IDCNN-E** Image Denoisin Convolutional Neural Network Trained with Edge Loss
19. **ML** Machine Learning
20. **MWIR** Middle Wave Infrared
21. **NIR** Near Infrared
22. **NLP** Natural Language Processing
23. **PAC** Probably Approximatley Correct
24. **RANSAC** Random Sample Consensus
25. **RPT** Rectangle Partition Tree
26. **SIFT** Scale Invariant Feature Transform
27. **SLAM** Simultaneous Localization and Mapping
28. **SNR** Singal to Noise Ratio
29. **SOTA** State Of The Art
30. **TPT** Triangle Partition Tree
31. **USD** United States Dollar
32. **VGG** Visual Geometry Group
33. **YOLO** You Only Look Once

Chapter 1

Introduction

COMPUTER vision and image processing involve many challenging problems. While early methods utilized 'classic' approaches, work in the last decade has focused on deep neural network architectures to solve them. Here, 'classic' refers to approaches that are not learning-based, such as engineered feature descriptors, theoretic-based algorithms, search methods and usage of theoretically proven thresholds. In this study, I explore the differences between classic and deep learning (DL) approaches in order to gain better insight regarding which is more suitable for a given imaging modality and what are their associated constraints. Its focus is on the DL part of the non-classic methods, i.e., I do not cover machine learning algorithms that are not deep. This is motivated by the fact that, DL has become the leading and most reported machine learning approach. Actually, currently, around 25 percent of the papers presented at computer-vision conferences take advantage of DL. Moreover, a session dedicated to DL has become the norm on the program of a variety of scientific venues. For example, in CVPR 2019, 25% of the published papers were assigned to the Deep Learning subject area.

In order to perform that investigation, I have focused on two computer vision tasks that are at the limit of the ability of current computer vision algorithms: faint edge detection in noisy images and multispectral registration of images.

Edge detection is one of the earliest problems that has been tackled by image processing and computer vision [24, 45, 15]. Although many approaches have been proposed to address this task, they still fail to detect edges when they are faint and the images are noisy as shown in [51, 50]. Those limitations are particularly problematic as these kinds of edges can be found in most imaging domains including satellite, medical, low-light, and even natural images.

With the development of multi-sensor cameras that capture images from different modalities, multispectral image alignment has become a very important computer vision task. Indeed, robust alignment between the different image channels forms the basis for informative image fusion, and for data fusion. Examples for image fusion and its applications are described in detail in [66]. E.g., object detection can be derived from color visible to infrared given an accurate multispectral alignment. Moreover, cross-spectral alignment cannot be carried out by single-channel registration methods like scale-invariant-feature-transform (SIFT) [41, 12].

1.1 Computer vision

Computer vision is a multi-disciplinary field that is related to artificial intelligence and image processing. It deals with how a computer can gain an understanding of, usually, the real world from images and videos. In a sense, computer vision is the opposite of computer graphics. Computer graphics transforms data into images while computer vision translates images into relevant data. From an engineering perspective, computer vision attempts to mimic the ability of the human visual system. Subdomain of computer vision includes object recognition, scene understanding, object tracking, image alignment, 3D reconstruction, motion estimation, and image restoration.

The field of computer vision is a well-studied area with a plethora of related works. Early methods were focused on low level tasks such as edge detection [15], noise removal [10] and optical flow computing [42]. More advanced works addressed complex problems like image alignment [12], image processing using wavelets [48] and object detection using machine learning classifiers [17]. In the next sub-sections, I will briefly introduce the technical domains that are at the core of this piece of research, i.e., faint edge detection, multispectral registration and deep learning.

1.2 Faint edge detection

Edge detection is the problem of identifying the pixels in an image where there are image intensity discontinuities. In other words, edges are composed of the pixels where image intensity levels change slightly to sharply. Typically, these edges are organized as a set of curved lines in the image. Edges can be modeled as either sharp step edges or a smooth transition between two segments with a significant difference in intensity levels. Edge detection is fundamental to computer vision and image processing as it is the basis for higher-level computer vision tasks like feature extraction, feature detection, and motion estimation.

Unfortunately, all the methods for regular edge detection are not geared to handle faint edges and noisy images [14]. To that end, researchers addressed the challenging task of faint edge detection in noisy images. [23] introduced a method to detect straight edges in nearly linear complexity by matched filters. My work, [51], was the first to address the problem for curved edges that also worked in nearly linear complexity. I utilized a dynamic programming approach and a binary multiscale partitioning of the image to compute the curved matched filter in practical time. In [50] introduced theoretical results of threshold limitation, concluding on how faint an edge can be and still be detected. I was first to introduce a Convolutional Neural Network (CNN) approach for the task, improving the accuracy of the previous methods on noisy simulations [52]. Because of the versatility of neural networks, my approach also delivered a high-quality performance for other vision tasks involving noisy images. I proposed methods for noisy image classification and natural image denoising using edge preservation losses and multi-scale CNN. Recent works introduced sub-linear approaches for faint edge detection [28, 73].

1.3 Multispectral registration

Image registration and alignment is a process of transforming two or more images into one coordinate system. In standard image registration, the input data are multiple images captured by the same sensor. Multimodal, or multispectral when the modalities are different spectral channels, image registration also permits processing images from different sensors. Here, registration is essential in order to integrate those heterogeneous data and gray levels into a single system or image. It is often used in medical imaging, panoramas creation, and multi-sensor cameras. The registration methods can be classified into two groups, i.e., intensity-based like [42] and feature-based like [12]. Registration is carried out using transformation models that can be as simple as only dealing with translation and scaling and as complex as integrating affine, projection, and nonlinear transformations. This process can be performed on either the spatial or the frequency domain.

Image registration allows producing image fusion, image panoramas, matching stereo, and recover an object's shape. While early method relied on Fast-Fourier-Transform [58] to solve complex geometric relations, more advanced methods used complex engineered descriptors for the alignment [12]. However, all those methods for single-channel registration fail when attempting to register multi-modal data. To that end, a group of works developed a unique method for cross-spectral alignment.

1.4 Deep learning in computer vision

In recent years, deep learning techniques have been revolutionizing computer vision. Analysis of the program of recent computer vision conferences like CVPR 2019 shows that deep learning has become a leading topic among the published papers, see Table 1.1. Deep learning is a specific type of machine learning method that has only started being particularly successful around 2010. That was made possible thanks to, in particular, the availability of GPUs computing and large datasets

Topic	Percentage
Deep Learning	24.4
Recognition	22.1
Face, Gesture, and Body Pose	10.8
Low-Level Vision	10.6
Image and Video Synthesis	8.6
Vision and Language	8.2
Segmentation, Grouping, and Shape	7.6
3D from Multiview and Sensors	7.6

Table 1.1: Topics of the papers presented at CVPR 2019 and their respective percentage.

[63]. The domain DL originates from, i.e., artificial neural networks (ANNs), was founded in the 40's [47]. Deep learning allows computational models with the architecture of multiple hidden layers and non-linear activations to process and learns a representation of data [35]. In other words, deep learning is the class of machine learning algorithms that uses multiple layers to extract high-level features from raw input.

One of the earlier ANN algorithms, back from the 50's, is the Perceptron [60], an algorithm for supervised learning of binary classifier based on a single artificial neuron. The Perceptron introduced supervised learning of a classifier based on a linear prediction combining weights and a feature vector. In 1989, the idea of using back-propagation to train a neural network for handwritten recognition was proposed [36]. Afterward, deep methods introduced a neural network for the classification of handwritten digits [37]. Advanced work presented the classification of a natural image to pre-known classes [33]. Then, classification formed to basics for deep learning-based object detection [40]. Recently, generative adversarial networks (GANs) were introduced to create naturally looking artificial images [57]. Deep and machine learning can be useful to real-life applications going from plant identification [38] to brain-seizure detection [21].

A recent study [56] compared traditional approaches to deep learning-based ones. They ”laid down many arguments why traditional CV techniques are still very much useful even in the age of DL”. They claim that, although DL has changed the limits of the ability of computer vision in such a way that more problems can be addressed and sometimes even solved, knowledge of classic computer vision should be maintained since it still has advantages. For instance, classic algorithms are often more robust to variability in the image, and the theory behind them can inform us about the nature of the problem. One should note that the nature of that study is quite unlike mine. Not only does it focus on a very different set of challenges, i.e., panorama generation, 3D vision and Simultaneous Localization And Mapping (SLAM), but it is based on surveys and opinions instead of experiments.

1.5 Aims and objectives

The first aim of this study is to advance fields of computer vision, more specifically faint edge detection and multispectral image alignment, by developing novel (both classic and deep learning-based) approaches. Its second aim is to gain an insight into the respective strengths and weaknesses of classic and deep approaches when addressing computer vision challenges. Those aims will be achieved by completing the following technical objectives:

1.5.1 Classic detection of faint edges using hierarchical partitioning

Detection of faint edges in noisy images is a challenging task that requires a tailored method as methods for regular edge detection fail at low Signal-to-Noise Ratios (SNRs). I propose to address this problem by applying a matched filter to curved edges. A matched filter smooths the noise along the edge and maximizes the contrast across the edges. As curved edge locations in the image are unknown a priori, it is intended to apply a smart search for curved edges which should offer nearly linear complexity in the image pixels. This search is expected to be performed by multiscale binary partitioning of the image. This study will also investigate theoretical questions such as: how faint can an edge be while still be detectable?

1.5.2 CNN-based detection of faint edges relying on edge preservation loss

As in many applications, neural networks have proved to outperform classic solutions, I also intend to investigate the faint edge detection challenge by exploiting a multi-scale neural network. While improving detection accuracy is of prime interest, the potential for faster processing using GPUs is also attractive. By performing this study, I also wish to understand fundamental questions regarding CNN-based approaches such as: How easy an approach can be transformed to other vision tasks? How relevant is edge information to computer vision tasks? Are edge detection maps and gradients informative for tasks involving noisy images? Can faint edge detection be carried out by a CNN?

1.5.3 Classic registration of multispectral images using a novel edge descriptor

As multispectral image alignment can form the basis for image fusion, it allows taking advantage of different spectra to analyze the content of a scene. I propose to approach this challenging task by aligning images using feature point calculation of geometry alignment based on an edge descriptor that is invariant to spectra. Experiments should reveal the limitation of the proposed descriptor both in terms of the spectral field of alignment and types of geometric transformations.

1.5.4 Deep registration of multispectral images relying on invariant descriptor learning

By definition, manually engineered descriptors are suboptimal. Therefore, the design of an invariant descriptor using deep learning schemes is an attractive proposition. It is proposed to extend the previous classic method by generating automatically a descriptor using metric learning taking advantage of a pseudo-Siamese network that can be seen as a teacher-student architecture. An important aspect of this study is also the understanding if such descriptors can be learned from minimal data acquisition and labeling. I would like to address a fundamental question: Is deep learning an appropriate approach to conduct multispectral registration?

1.5.5 Deep learning solutions versus classic methods

Having proposed both classic and deep learning-based state-of-the-art solutions for diverse applications, I will be in a position to compare those two types of approaches and obtain some general insight into their strengths and limitations. In addition to accuracy and runtime speed, some more fundamental aspects will be studied. They include the complexity of their development, their theoretical background, their reliance on data and/or previous knowledge, and their robustness when applied in other similar domains and/or more extreme conditions.

1.6 Scientific contribution

In this thesis, novel ideas are presented leading to novel algorithms for faint edge detection and multispectral image alignment, and a comparison of classic and deep approaches.

1.6.1 Classic detection of faint edges

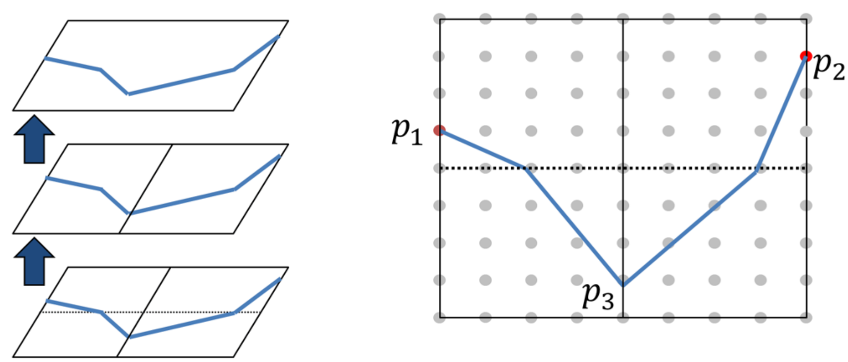


Figure 1.1: Rectangle-Partition-Tree of an image. My classic method for detection of faint edges search for every curve between every two boundary points $\forall p_1, p_2$ the best concatenation of sub-curves by breaking point p_3. This search for curves is done recursively in a bottom-up dynamic-programming-like approach. For an image with N pixels, the complexity is $O(N^{1.5})$ in the full mode and $O(N \log N)$ in the optimized mode.

The challenge I address is the detection of curved edges of arbitrary shapes at low SNRs. My approach is to average the noise and maximize the contrast using a matched filter. Even though the search space of such curves in an image is of exponential size, an efficient polynomial approach can detect edges sufficiently by designing a classic algorithm for computer vision. Indeed, an important issue is the ability to detect curved edges accurately in a practical run-time.

My classic approach for faint edge detection utilizes a matched filter. This filter smoothes the noise along the curve and maximizes the contrast across the edge and therefore allows detection of curved edges at low SNRs. However, as the edge locations are unknown apriori, I search for their location using dynamic programming like algorithm. My approach relies on a Rectangle-Partition-Tree (RPT) of the image pixels as described in Figure 1.1. At the bottom level, I scan every straight-line edge in a 5×5 tiling of the image. Every higher-level contains rectangles divided into 2 sub-rectangles. In Figure 1.1, p_1 is a boundary point on the left sub-rectangle, p_2 is a pixel on the right one. For every such pair $\forall p_1, p_2$ I search for the best middle pixel p_3 that maximizes the curve contrast. The full mode of this approach costs $O(N^{1.5})$ in terms of complexity for an image with N pixels. In my optimized mode, I select smartly the k best candidates of p_3, achieving an $O(N \log N)$ complexity, which is more practical for big images.

Given, the best curve candidates the algorithm found, I introduce a theoretical function for thresholding them. For a curve of length L and width w, where the algorithm search for K_L such curves, and the Gaussian noise level in the image is σ, the derived threshold is:

$$T(L, K_L) = \sigma\sqrt{\frac{2\ln K_L}{wL}}. \tag{1.1}$$

Since I derived a closed-form for K_L in my method, I can infer the minimal detectable contrast of a faint edge. By $L \to \infty$ I prove that the faintest edge that can be detected is a function of:

$$T_\infty = \Omega(\frac{\sigma}{\sqrt{w}}). \tag{1.2}$$

In conclusion, I developed a multiscale algorithm for the detection of faint curved edges in noisy images. My approach is nearly linear and utilizes a proven threshold and lower bound.

1.6.2 CNN-based detection of faint edges

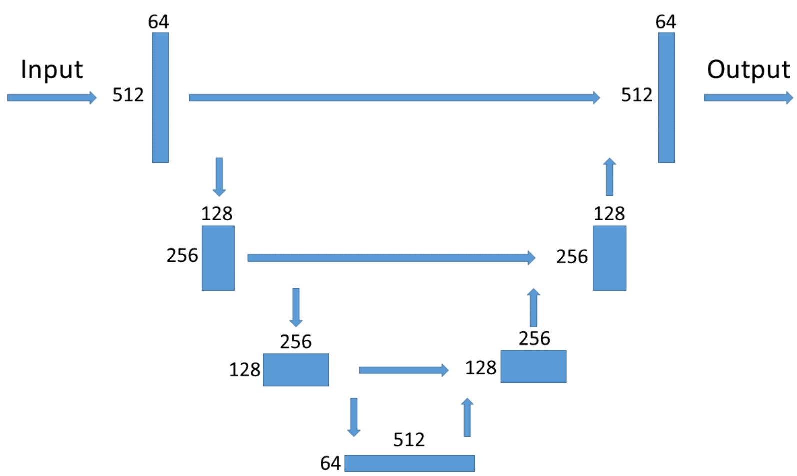


Figure 1.2: My FED-CNN network of a U-Net Architecture is the multi scale CNN that I used to mimic the hierarchical tree approach of my classic algorithm. This deep neural network was trained by an edge preservation loss.

The challenge I address here is similar to the one of the previous sub-section. However, I propose the detection of curved edges of arbitrary shape at low SNR using a deep neural network. I aim to carry out this procedure together with an improvement of run-time. Indeed, faint edge detection can be addressed by deep learning detecting curves of arbitrary shapes. The expectation was that the DL approach, as in many other applications, would outperform state-of-the-art classic methods in terms of accuracy.

To detect edges by the deep approach I need a multiscale Convolutional Neural Network that mimics my RPT filter. My Faint-Edge-Detection (FED) network that addresses this requirement is of U-Net architecture as described in Figure 1.2. Instead of deriving thresholds as in the classic scheme, I should derive loss functions to train my FED-CNN. To that end, I use an edge preservation loss which is based on dice-coefficients: Given an edge result of my network y, and edge label $y^{'}$ the loss is

$$Di(y, y^{'}) = -\frac{\sum_p y^{'}(p)y(p)}{\sum_p y^{'}(p) + \sum_p y(p)}. \tag{1.3}$$

The enumerator maximizes the shared edges while the denominator minimizes the false positive detections. Given my CNN architecture and edge preservation loss, I can solve other tasks of noisy images like image denoising. Given an Image-Denoising CNN, I can improve results by a similar edge preservation loss:

$$L_E = ||\frac{\partial}{\partial x}I_c - \frac{\partial}{\partial x}IDCNN(I_n)||_2^2. \tag{1.4}$$

In conclusion, my approach for multiscale processing of noisy images using edge-preservation losses achieved high-quality results in three tasks, faint edge detection, noisy image classification, and natural image denoising.

1.6.3 Classic registration of multispectral images

To gain information from multi-sensor cameras, a pre-processing step of image alignment is usually needed. Although single-channel registration is a well-studied problem, in the case of multi-channel, it can be very challenging. Multi-spectral alignment, even though an ill-posed problem, can be addressed by engineering features that provide invariant representation.

To that end, I need a method that is robust both for geometric distortions and for cross-spectral changes. My registration is carried out by a feature-based approach since feature correspondences are the basis for global geometry derivation.

My method utilized an iterative outlier rejection which is an improvement of the regular Random-Sample-Consensus (RANSAC). The features are calculated by a Harris Corner Detector. However, the core of my method is the proposed descriptor. My descriptor is based on edge representation, which is invariant to different spectra, but also to geometric relations. The proposed representation is based on Canny Edges, and on gradient orientations. The justification for this engineered descriptor can be seen in Figure 1.3. Although the original cross-spectral patches have no correlation between them, their edge images are of high similarity.

To conclude, I developed a feature-based robust alignment of multispectral images. That is invariant to a wide field of spectral channels (Visible and Middle Wave Infrared) and is also invariant to significant rotation, scaling and translation.

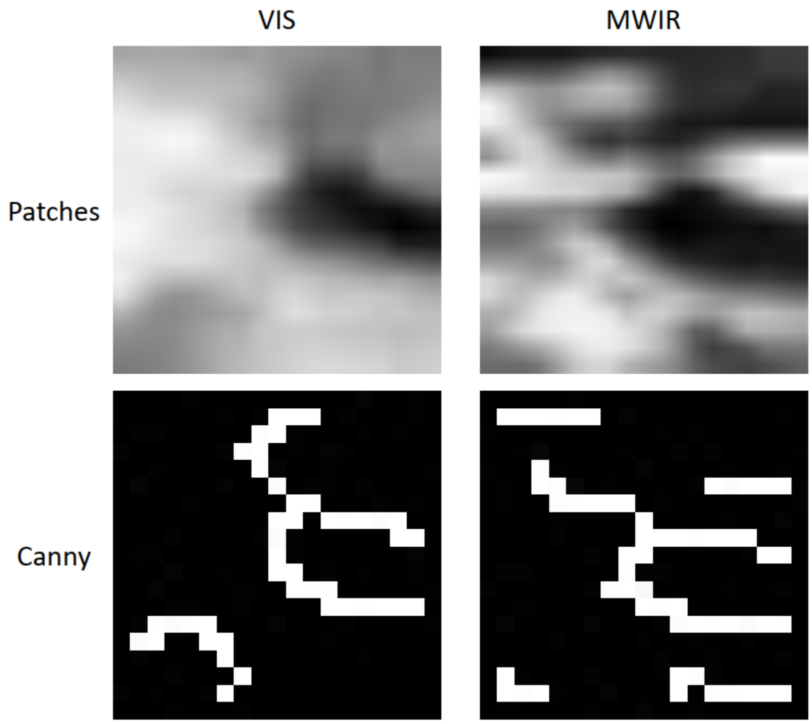


Figure 1.3: Multispectral patches and their corresponding edge maps that are part of their edge descriptors. Although the original patches are not correlated their edge maps attain a significant similarity. Therefore, this kind of descriptor is more invariant to different spectra than descriptors of inter-spectral alignment.

1.6.4 Deep registration of multispectral images

The challenge I address here is similar to the one of the previous sub-section. I propose to reuse my classic approach using the same feature-based method, however, replacing its feature representation by a CNN. The question I address is how to learn an invariant descriptor of multispectral feature points. It is proposed to exploit a CNN feature extraction that is learned out of a cross-spectral image dataset. A deep approach to the teacher-student scheme can produce a pseudo-Siamese network for es-

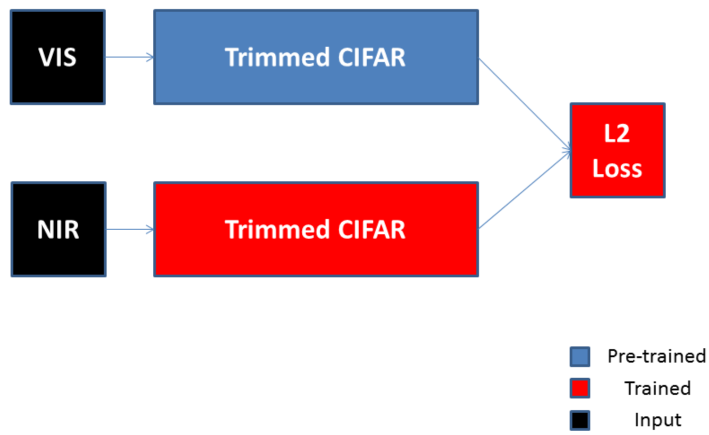


Figure 1.4: My learning architecture for a deep invariant descriptor. This can be seen as a pseudo-Siamese network or as a teacher-student scheme. The visible color patch is forwarded through a pre-trained classification network that was trained on the CIFAR10 dataset. Its corresponding Near-Infrared (NIR) patch is being used to train the infrared network to produce a similar invariant representation.

timating the metric between corresponding patches.

Figure 1.4 describes my learning scheme. This architecture can be seen as teacher-student, pseudo-Siamese, or metric learning. The color visible network is pre-trained to classify images from the CIFAR10 dataset. After this network is trimmed, it extracts a descriptor for every color visible corner patch. To obtain a similar network for infrared I trained the second network by l_2 loss. The infrared network is of the same architecture but with different weights. The input for this learning method is pairs of aligned cross-spectral patches from my dataset.

In comparison, relative to the classic approach, my deep descriptor is more robust to geometric relations and solves a narrower field of spectral channels.

1.6.5 Comparison between classic and deep learning-based approaches

An important contribution of this study is a comparative perspective between classic and deep learning-based approaches informed by my experiments on both faint edge detection and multispectral image alignment. It reveals that each one has its advantages and limitations.

Classic and DL-based approaches have many differences. DL approaches are based on data while classic ones often rely on theoretical models. Therefore, DL ones tend to achieve high performance when processing data similar to those of the training dataset, whereas classic assumptions often generalize to other imaging domains as well. While DL ones rely on loss heuristics and pre-defined CNN architectures, classic methods are based on proof and theoretic thresholds. In addition, DL approaches take advantage of CNNs seen as a black box, while in a classic one the developer is able to justify the algorithmic steps. In general, although DL methods have stretched the limits of computer vision and have won every context, classic-traditional approaches still have many advantages, including relevance to other imaging domains, theoretical foundations, a better understanding of the problems and their solutions, proven algorithms steps, which provides a degree of trust in the code that cannot be delivered by that relying on a black box.

My study has focused on faint edge detection and multispectral image registration which are two problems at the limit of the ability of computer vision. The conclusion of my work confirms that each type of approach, classic and deep-learning, has its advantages. Classic methods are robust to image type and modality and are based on theoretic foundations. Deep approaches are faster to develop given a relevant dataset and achieving higher accuracy on a test set. They are robust to geometry variabilities and different object appearances.

More specifically, as faint edge detection in the classic methods assumes both an edge and a noise model, I developed an algorithm that exploits these assumptions on edges. Although those models are based on simplifications, the proposed classic method can be applied to other edge domains like those of natural or medical images. In the deep approach, the assumptions are replaced by a dataset, and as my works show, the CNN training achieving better accuracy on the train and test sets. However, it is limited to other domain transitions. Similar lessons were gained from multispectral image alignment. As the classic method assumed an invariant property of spectral channels, I developed a descriptor that represents only the shared information of different spectra. Experimentally, this assumption delivered invariance to a wide field of spectral images. In the deep-based approach, I trained a CNN to be invariant by using a set of same and not same cross-spectral patch pairs. The training process produced a metric between such pairs of patches. Experiments suggested the CNN approach to be invariant to a narrower field of imaging domains but a wider field of geometric transformations. Here again, I achieved higher accuracy on images similar to the train and test sets.

1.7 Thesis outline

This thesis is divided into several chapters as follows:

Chapter 2 surveys previous works related to this thesis content. It presents early and recent approaches attempting to address the tasks discussed in this work.

Chapter 3 is about classic and deep approaches for FED. The first contribution introduces a classic solution related to the detection of faint edges. It consists of two publications, i.e., a conference and a journal one. In particular, they discuss existing methods to detect faint edges in nearly linear complexity. In addition, I present a deep approach for the detection of faint edges. The published paper also introduces similar approaches for noisy image processing using multiscale CNN trained by edge preservation losses.

Chapter 4 reports on classic and DL methods for multispectral alignment. The classic paper proposes a method to align and fuse cross-spectral images using a novel edge descriptor. This method is classic as it relies on an engineered alignment scheme. I also introduce a deep based method to create an invariant descriptor of multispectral image registration which is learned by a pseudo-Siamese networks.

Chapter 5 discusses the differences between classic and deep approaches that could be highlighted from work reported in the published papers.

Chapter 6 reports the conclusions of this thesis, it also contains a discussion and suggests future work.

Chapter 2

Literature review

SINCE this thesis by publication consists of five published papers, each of them including a comprehensive literature review, their content will not be duplicated in this section. Instead, I will refer to the relevant sections in those papers so that the reader can find easily the relevant material in those manuscripts.

2.1 Faint edge detection

Edge detection is a fundamental problem of computer vision and image processing. The most basic approach detects edge by gradient magnitude [24]. Marr and Hildreth detected edge locations by laplacian zero-crossing [45]. Canny extended the basic approaches by hysteresis gradient thresholding [15]. A recent group of works focused on a similar problem of boundary detection in natural images [46]. They introduced machine learning approaches to detect the boundaries [8, 29, 19]. As demonstrated on a standard dataset [46], their accuracy could be improved by deep learning methods [76, 77, 43].

Additional details on previous studies about edge detection are found in Chapter 3. More specifically, in paper [51] of Chapter 3, the previous works Section 2 provides a comprehensive literature review on the topic of edge detection, from early methods up to the state-of-the-art in FED. It shows that classic methods for regular edge detection are unable to detect accurately edges under the presence of faint signals and noisy images. More specifically, current methods do not accurately detect edges at low SNRs and are highly affected by the presence of noise. Consequently, they produce false-detections out of noise gradients and they are not sensitive enough to long curves with faint contrast.

2.2 Multispectral registration

Image alignment is a classic and useful problem of computer vision. Image registration allows producing image fusion, image panoramas, matching stereo, and recover an object's shape.

The field of multispectal image alignment is addressed by a group of works. A unique descriptor was invented for that task [3], however, my experiments showed that it was not accurate enough in multispectral alignment. I offered an edge-based descriptor [54], that together with my improved version of iterative Random-Sample-Consensus (RANSAC) [22], outperformed their performance. My engineered descriptor manages to align a wide field of spectral channels, color visible to middle-wave-infrared (MWIR). Later, [2] introduced classification CNN for pairs of cross-spectral patches. Their classifier attempts to decide if a pair of visible and infrared image patches capture the same real-world visual object or not. However, their approach did not produce a full end-to-end solution for image alignment. My deep method, [53], introduced a teacher-student scheme to learn a similarity measure between cross-spectral edges. My method is robust to complex geometric relations and achieves high accuracy registration of visible color to near-infrared (NIR). However, to deal with small geometric transformations between multimodal images, a deep learning-based study [6] proposed an end-to-end approach for computing multispectral optical flow using spatial transformer networks [30].

Details on previous studies about multispectral image alignment are found in Chapter 4. While the classic paper [54] focuses mainly on classic methods and fusion approaches, the emphasis of deep manuscript [53] is on DL-based algorithms. In papers [54] and [53], in the previous works Section 2, there is a detailed review on alignment methods, from early methods to most recent studies. These literature reviews reveal that methods for regular image alignment fail under the problem of multispectral image registration. Therefore, more work is needed using both deep and classic approaches to align cross-spectral images captured by different sensors.

2.3 Deep learning solutions versus classic methods

Early computer vision algorithms utilized classic methods, while in the last decade the subject faced a major transition of focus into deep learning-based approaches. First, I introduce classic methods, providing both definitions and examples. Second, I survey studies that have compared classic and deep in specific applications. Third, I report on the amount of energy and computing resources required by deep learning methods.

This is followed by a discussion on the computational time and the associated hardware requirement of DL during both training and test phases. Next, I introduce the issue of adversarial attacks on deep neural networks. Finally, I discuss the black-box effect of DL and the development procedure.

One can consider three types of approaches: classic, ML without DL, and DL. Some ML algorithms have similarities with the classic ones in the sense that their learned classifier or regressor are on the top of features extracted in a classic way. Alternatively, many ML approaches are like DL ones as their input is, e.g., the original image's pixels and they learn a whole function on them. For the algorithm comparison in this work, I will focus on the main two extremes, i.e., classic and DL. I will not consider ML that are not DL, since they comprise a large and heterogeneous group of methods and have a lot of overlap with the two other approaches.

2.3.1 Classic methods

In this thesis, classic or traditional approaches are defined as approaches that do not rely on machine learning. Classic methods are engineered algorithms that rely on theory and/or mathematical models and not directly on external data. This is fundamentally different from ML approaches that generate models by learning directly from data, setting algorithm parameters using methods like gradient descend. Classic approaches are designed by experiments and intuitions while ML methods are developed in a more specific way comprising model selection, objective function, and parameters learning. A typical example of such traditional algorithms is Canny edge detection [14] that uses hysteresis of gradients to identify curves in the image. Another classic example is the SIFT [41] descriptor, which is an engineered and handcrafted representation of an image interest point. SIFT is the basics of many high-level computer-vision methods such as panorama stitching and object detection.

2.3.2 Comparison based on specific applications

A recent study [56] compares traditional and deep learning algorithms. Their study investigates three computer vision tasks, i.e., panorama generation, 3D reconstruction, and Simultaneous Localization and Mapping (SLAM). They show that each approach has its advantages and limitations. They report that DL enables engineers to achieve greater accuracy on the reviewed tasks. Moreover, since the algorithms are trained and not developed, they require less expert analysis and manual fine-tuning than classic methods. They also mention that an additional advantage of DL techniques is their flexibility. For example, a CNN model can be retrained using a custom dataset for any related use case. A major new feature brought by DL has been the unification of feature extraction and classification within a single framework.

Despite those advantages, classic methods also have some of their own. The authors claim that in many situations, usage of DL is probably overkilled with a huge number of parameters like, e.g., in image denoising [78]. In addition, they feel that classic methods could have addressed the problem much more efficiently in fewer lines of code. Moreover, with respect to feature extraction and representation, they highlight that those classic descriptors are more general and therefore are suitable for many image types, whereas features learned from CNN are specific to the training set.

There are more studies that inform such a comparison. They focus on a single application and perform experiments to evaluate their difference in terms of accuracy. A recent publication [11] reports the comparison of a set of classic keypoint descriptors with their deep learning-based competitors [55], [18]. Following the evaluation of those descriptors under various geometric transformations and illumination conditions, they show that some combinations of classic keypoint detectors and descriptors outperform pre-trained deep models, demonstrating there is still a value in considering classic feature descriptors. In contrast to the expectation, the tested deep models did not outperform the classic approaches dramatically. This can be explained by the fact that for such an application feature representation seems quite intuitive and therefore easy to engineer by an expert. While they claim that DL helps for illumination variance mainly, as a whole, contrary to the conclusions seen in [56] which suggest the general superiority of DL techniques, here they do not improve significantly performance.

An additional study surveys traditional and DL methods for face recognition [71]. As it is generally accepted in the community like in boundary detection for example [76], they report that CNNs have become the standard since they deliver significant accuracy improvements. Since the state of the art for face recognition is dominated by DL-based approaches, this suggests that problems that are particularly hard to model like this one benefit greatly from DL when compared to approaches relying on classic feature extraction [11]. Thus, I can infer that conclusions regarding performance depend on the task that is considered. This study claims that improvement in CNN results may be easily achieved by extending the network's capacity and making the dataset bigger. However, these improvements are expensive and deep CNN is slow to train and deploy.

A practical study carried out a similar comparison for a mobile robot application [44], where the authors focus on the important problem of visual object detection. They evaluate two methods: classic feature extraction with a learned classifier versus object detection with a compact CNN named YOLO v3 [59]. While this study is far from being comprehensive, it can still inform my comparison. They found that the classic-based detector does not detect all the objects under varying geometries such as size and rotations, while they report that the tiny CNN-based detector deals with these variations outperforming the classic approach.

Table 2.1 summaries the outcomes of the mentioned comparison studies.

Subject	Deep learning	Classic
Features extraction [11]	High-Accuracy	High-Accuracy
Feature extraction robustness [11]	Lightning-Invariant	Image-Type-Invariant
Object detection [44]	High-Accuracy	Moderate-Accuracy
Object detection robustness [44]	Geometry-Invariant	Geometry-Limited
Face recognition [71]	High-Accuracy	Moderate-Accuracy

Table 2.1: A table comparing the advantages of deep learning versus classic methods as reported by the literature.

Although generally, DL approaches tend to perform better than classic ones, this still depends on the specific application of interest. For those where the features that are relevant are well understood and can be engineered to model nature, classic methods may still have a leading role to play [11]. Indeed, human intuition and expertise about the solution may not be easily learned by a deep neural network. However, where objects vary according to complex distributions, like people and faces, DL detection and recognition are much more robust as can be seen in [44] and [71]. In general, deep approaches are invariant to illumination and geometry, while classic ones are more robust to image type and modality. This may be explained by the fact that classic approaches use assumptions that are usually modality invariant, while deep approaches are designed to be flexible to the variance in the training set, for example, in terms of varying geometries and lightning conditions.

2.3.3 Energy, computing resources and associated hardware requirements

Many researchers have discussed the limitations of deep neural networks by considering other aspects than accuracy performance. [26] criticizes the training phase and its associated large energetic footprint. As better results are usually achieved by increasing a network's size, one may want to train larger and larger networks with tens of millions of parameters. In their study they claim that training a network with 65 million parameters, which is the size required for achieving accurate results on a typical DL framework such as PyTorch, requires 27kWh energy - this amount of energy is equivalent to lighting a led bulb for two months - and costs between 41 to 140 USD of Cloud computing. Another publication, which also discusses the resources needed for DL [67], mentions that, although large neural networks improve the accuracy of NLP algorithms, they rely on the availability of large computational devices. They report that the training of an NLP standard DL model like the one proposed in [9] requires 120 hours which can cost up to 180 USD of cloud computing and electricity.

Both studies reach the same conclusion that, while deep methods deliver high accuracy, they are not as efficient as their previous traditional methods in terms of energy and computing resources. Indeed classic methods rely on a much lower number of parameters, multipliers, and do not require a training phase.

A recent study [31] predicts the computational cost of DL models. They report that the average-case assumption is that the training time is linear in the number of floating-point operations in the model. This is a disadvantage of DL since it requires millions of parameters to work properly, whereas classic approaches not only do not require any training phase: they, and standard ML approaches, rely on a number of parameters of several orders of magnitude smaller. Note that, although in general there is no linear relationship between the number of parameters in a DL network and the number of FLOPS, they share a strong connection. This study also compares different hardware platforms from cloud computing to GPUs. The deeper the network, the more accurately the authors succeeded at predicting the training time required for convergence. They conclude that the required time increases with the batch-size while depending on the optimization method.

In summary, all these studies show that for DL networks to work properly, important resources are needed. While [26] discusses their huge footprint, [67] claims that DL methods are not as efficient as their previous traditional approaches. In addition, [31] predicts the training time required for a model, and since it needs to be large to work accurately, they require a huge amount of time on every potential computing device. In conclusion, if a developer decides to use a DL approach to address a problem, they should take into account not only accuracy but also the high energy consumption of the development process, and the power consumption associated with running it. Thus, if the accuracy achieved by a classic approach is sufficient, it may be preferable over its DL alternative.

2.3.4 Adversarial deep learning

One of the major problems of decision systems based on deep learning is that they can be fooled by adversarial attacks. They can classify incorrectly as a result of many types of attacks. Among them, one can mention one-pixel [68], spatial [75], and physical attacks [34]. A one-pixel attack is a process of changing the gray-level of single pixels to get a different and wrong decision. For example, an image of a dog with one defective pixel can be interpreted as a cat. Spatial attacks apply geometric warping of the image such that it looks the same but recognized as another object. For example, a tiny geometric distortion on one digit leads to be classified as another digit. A physical attack is a sticker that one can put in the real work on an object, a stop sign for example. Due to this computed sticker, the object is misidentified: for example, the stop sign will be detected as a speed limit sign. Although there are ways to defend deep networks, they are still not as secure as classic decision algorithms [4]. Indeed, although some classic approaches may not be secure, their attack cannot be performed using a systematic approach like these applied to CNNs. Thus, more hacking skills are generally required to fool them.

2.3.5 Black box effect and development process

In this subsection, I discuss two issues that have not been covered yet, i.e., the black-box effect and the development process. A problem of DL methods is the limited ability of humans to interpret them, the infamous black-box effect. Although recent work has investigated approaches to understand these black boxes by either information provision [65] or visualization [62], this research domain is still in its infancy. Moreover, many DL developers use ablation studies to understand their trained networks better like in style transfer for example [20]. Still, the question remains whether one can trust an algorithm that one does not fully understand. A typical example is the deployment of self-driving cars: this black box effect creates many legal problems as ethical issues are paramount in autonomous driving [39].

The availability of DL frameworks has dramatically changed the way algorithms are developed with respect to traditional approaches. While classic algorithms require expert knowledge, DL methods are able to learn from examples given an engineered model. On one hand, the development phase is faster and models can easily be recycled for new purposes, e.g., using transfer learning [70]. On the other hand, the development relies on the availability of large datasets [5]. For some applications, the generation of large enough training sets may be impossible. In some cases, the images may be biased on a specific type of examples, leading the trained CNN to be biased as well. In addition, as very often image labels are not available, the development process requires manual labeling of images, a task that is extremely labor-intensive.

2.4 Literature review summary

As this literature review shows that DL methods have managed to stretch the limits of what is possible in computer vision, they are clearly very successful and relevant. They have also changed the software development process which is faster in many cases. However, DL methods still have their limitations with respect to the classic approaches. They rely on big datasets and large energy consumption. Moreover, they are often not able to generalize the learning examples to other imaging types and domains. Thus, classic algorithms have many advantages: they can be better understood by humans, they do not require large datasets, and the assumptions on which they are designed are frequently relevant to many test cases. Moreover, they are usually more secure and harder to be hacked.

This chapter is derived from the literature review of faint edge detection, multispectral registration, and comparisons between classic and deep learning approaches. These problems are challenging and regular methods do not solve them accurately. In edge detection, DL stretched the limits of F-measure achieved on datasets. However, the deep methods require a large training time, and a large footprint with respect to their classic predecessors. The first edge detection methods were based on theory only, they were fast to develop and to run. More advanced methods utilized complex engineering and basic machine learning. The classic methods are more explainable, while the deep methods achieve higher accuracy. In image registration, engineered methods are still state-of-the-art. The power and invariance of these descriptors are not easy to achieve with deep learning approaches. Still, a modern method is often a combination of the two, i.e., deep learning and classic approaches. Regarding the general comparison between the classic and DL methodologies, there are key differences that are valid for all problems. Deep learning methods require energy and resources to train, but they are usually faster to develop. Classic methods may be based on theory foundations, require significant development time, and are often faster to run.

Chapter 3

Faint edge detection

THIS chapter introduces classic and deep manuscripts addressing the ill-posed problem of faint edge detection in noisy images and from arbitrary shapes. In the piece of research reported in [51], detection of faint curved edges is carried out by a binary hierarchical partitioning of the image pixels. This work is the first to introduce a nearly linear complexity algorithm for detection of faint curved edges. Moreover, it provides a proven lower bound on how faint an edge can be and still be detected by the method. The journal paper of [50] surveys all the nearly linear methods for that problem from a straight-lines efficient algorithm to the proposed fast curved edge detection.

To improve the accuracy of FED and robustness to edge arbitrary shapes, I continued the FED effort by using DL techniques. My deep manuscript introduces a CNN-based approach to detect faint curved edges [52]. This CNN performs multiscale processing of noisy images and it mimics the classic hierarchical structures presented in the classic FED papers. This CNN, which is trained by an edge preservation loss, achieves the best results on simulations of noisy images and step curved edges. Moreover, due to the nature of deep learning approaches, the proposed scheme can be used to address similar problems. This paper also reports highly accurate results with a similar CNN on noisy image classification and natural image denoising.

These FED papers where I am the primary author are included in this Chapter:

1. Ofir N., Galun M., Nadler B., Basri R. (2016). Fast detection of curved edges at low SNR. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (pp. 213-221).

2. Ofir N., Galun, M., Alpert, S., Brant A., Nadler B., Basri R. (2019). On detection of faint edges in noisy images. IEEE transactions on pattern analysis and machine intelligence, 42(4), 894-908.

3. Ofir N., Keller Y. (2021, January). Multi-scale Processing of Noisy Images using Edge Preservation Losses. In 2021 25th IEEE International Conference on Pattern Recognition. (ICPR) IEEE.

Papers removed for copyright reasons

Chapter 4

Multispectral image registration

IN the current chapter I introduce my classic and DL works regarding multispectral image alignment. The classic paper uses an engineered descriptor, while the DL one takes advantage of learned representation. One should note that both approaches are feature-based.

The contribution for classic multispectral image alignment relies on feature-based registration [54]. The descriptor used in this scheme is engineered to be invariant to different spectra. The proposed descriptor is edge-based and contains information on edge pixels and gradient directions. It is invariant to spectral channels, translation, scaling, and small rotations. This work also introduces a fast method to fuse alignment images that can be applied to real-time multisensor cameras.

The previous contribution is extended by a deep learning-based approach. I improve the previous method by replacing the descriptor with CNN-based alignment [53]. The proposed invariant learned descriptor is a result of CNN training by metric learning that can be seen as a teacher-student scheme. The outcome of the training process is two pseudo-Siamese networks, one for color visible and the second for

NIR. The two networks produce similar representations of corresponding cross-spectral patches and far representations of non-corresponding patches. This forms the basics for multispectral feature-based image registration. As experiments show, this registration is more robust to geometric transformation, but less robust to wider spectral fields.

The corresponding manuscripts where I am the primary author are included in this Chapter:

1. Ofir N., Silberstein, S., Rozenbaum D., Duvdevani Bar S. (2018, October). Registration and fusion of multi-spectral images using a novel edge descriptor. In 2018 25th IEEE International Conference on Image Processing. (ICIP) IEEE, (pp. 1857-1861).

2. Ofir N., Silberstein, S., Levi H., Rozenbaum D., Duvdevani Bar S. (2018, October). Deep Multi-Spectral Registration Using Invariant Descriptor Learning. In 2018 25th IEEE International Conference on Image Processing. (ICIP) IEEE, (pp. 1238-1242).

Papers removed for copyright reasons

Chapter 5

Differences between classic and deep approaches

5.1 Introduction

In the last decade, deep learning methods have achieved great success in addressing computer vision problems [56]. My study shows that these approaches are indeed relevant for addressing challenges on the limit of processing ability. As could be seen in the previous chapters, deep learning approaches have delivered the best algorithmic accuracy on the corresponding training and test datasets. My experiment shows that the deep approaches for faint edge detection and multispectral image alignment are more accurate than the classic methods. However, shall I abandon the classic approaches? My study has shown that a first limitation of the deep based methods is their transfer to another imaging domain. While deep methods are accurate on images similar to the training set, classic schemes are more robust to different imaging domains and modalities [51], [54], [25]. This limitation can be explained by the probably-approximately-correct model (PAC) [27]. In a learning scheme, DL or ML, the learner selects a generalization function out of the hypothesis classes. In order for the problem to be learnable, the selected function should have a low generalization error. However, if the hypothesis class is extended to other imaging domains after training, the generalization error would increase. The PAC model has proved limitations on every learning scheme. Alternatively, models based on classic methods rely on simplifying assumptions. Consequently, these are more likely to fit a variety of imaging domains such as real natural images as in [7]. Indeed traditional schemes are developed in a different way than DL ones. Although a classic approach is developed to solve a specific scenario that is modeled with simplicity, the developed theory and/or heuristics associated with that simple model is still able to deliver good quality results on other imaging domains [51], [11].

An additional difference between the classic and deep learning approaches is related to the process of algorithm development, which affects the theoretic findings of a contribution [51], [11]. Classic methods require algorithms with new data structures, rigorous complexity analysis, and proven thresholds. On the other hand, deep approaches require different development mechanisms that include CNN architecture and loss functions. An advantage of such type of development process is that deep methods can be easily transferred to address similar problems. This is exemplified in my work [52] where the same architecture was used for faint edge detection, noisy image classification and natural image denoising. Similarly, [40] introduced the recycling of classification networks for object detection. Thus, a method that shows superiority in faint edge detection can easily be transformed to achieve the best quality in noisy image classification or natural image denoising. While a new dataset and new loss are required, the general CNN architecture can be reused, e.g., for multiscale processing of noisy images.

As mentioned in Chapter 2, there are several studies that compared classic and deep learning approaches through specific problems: [56] investigated those algorithms applied to panorama, 3D and SLAM, while [44] compared those approaches through the applications of robotic vision and visual object detection. Moreover, [26] and [67] discussed the limitations of DL approaches in terms of the large computing resources they require. This chapter starts by making the comparison according to my two applications of study: FED and multispectral alignment. Then, it goes further to generalize the highlighted differences between deep learning and classic methods.

5.2 Comparison of deep learning and classic methods for faint edge detection

Faint edge detection is a challenging problem that requires special approaches for addressing it. Chapter 3 introduced a classic method to address it using a hierarchical binary partitioning of the image pixels. In addition, it suggested a DL-based approach using a multiscale CNN. As each of those methods showed strengths and limitations, this section discusses them based on the outcomes of the specific experiments that were conducted.

First, accuracy performances of four methods, two from each type, are compared: Canny [14] - the classic approach

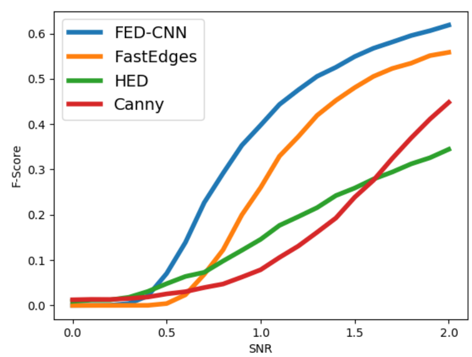


Figure 5.1: Simulation of faint edges detection in noisy image. The strict F-score graph along the different signal-to-noise ratios from 0 to 2. Our methods, FED-CNN and FastEdges which are geared for faint edges detection, achieve best performance.

Algorithm	SNR = 1	SNR = 2
FED-CNN	**0.4**	**0.62**
FastEdges	0.28	0.56
HED	0.14	0.34
Canny	0.08	0.45

Table 5.1: F-score of the methods at SNR 1 and 2. At both SNRs my methods achieve the best scores.

for edge detection - is used as a reference, Holistically Edge Detection (HED) [76] - a deep method based on the standard VGG-16 classification network - and two of my methods, i.e., Fast Edges [51] - my classic FED solution, and FED-CNN [52] - my deep learning method. The F-measure is used as evaluation metric as it is particularly suitable to access binary classifiers, such as edge detectors, as it provides a balance between precision and recall. It is the harmonic mean of the precision and recall: $\frac{2PR}{P+R}$, where precision is the ratio of $\frac{true-positive}{true-positive+false-positive}$, and recall is the ratio of

Algorithm	Run-Time (milliseconds)
FED-CNN	**10**
FED-CNN-CPU	**800**
FastEdges	2600
Canny	3

Table 5.2: Run time in milli-seconds of the different methods of edge detection. My runtime is very close to Canny's time and is order of magnitude faster than FastEdges. I achieve this improvement mainly by running our network on a GPU, which is a hardware optimization. The advantage of CNN approaches over classic methods like FastEdges [51] is that they are easily implemented and accelerated on a GPU. In addition, due to the simplicity of FED-CNN, its runtime on a CPU is also faster than FastEdges [51].

$\frac{true-positive}{true-positive+false-negative}$.

As it can be seen in Figure 5.1 and Table 5.1 from [52], my deep learning approach achieves the highest F-measure in this simulation experiment. It does so since the simulation domain area is the same domain as the one of the training set, i.e., binary images contaminated with Gaussian noise. My classic method also performs well, but not as well as my DL-based approach, although the simulations fit its model assumptions.

The second aspect of interest is runtime and computational complexity. As it can be seen in Table 5.2 from [52], the CPU runtime of my Intel i9 Sky-Lake processor for my CNN solution is lower than this of my hierarchical algorithm. Note that this speedup is explained by the reduction of computational complexity, from nearly linear [51] to linear [52]. Moreover, the DL algorithm can easily be accelerated using a GPU, in my case a GeForce gtx 1070.

Although the deep learning-based FED significantly outperforms my classic scheme in both accuracy and speed, the traditional approach has its clear advantages. Firstly, it has strong theoretic foundations. The algorithmic steps are described in [51], and its calculated complexity, C(N), can be expressed by the following Equations:

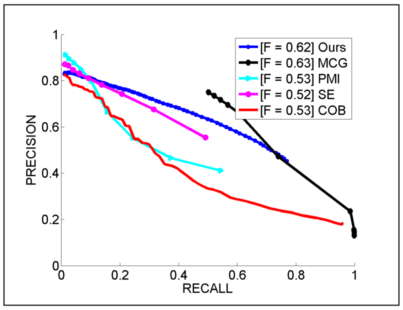
$$C(N) \leq 6N^{1.5}\left[\sum_{l=0}^{\infty} 2^{-l} + \sum_{l=1}^{\infty} 2^{-l}\right] = 18N^{1.5}, \qquad (5.1)$$

where N is the number of image pixels, and l denotes the hierarchical level. Moreover, Equation

$$T_{\infty} = \Omega(\frac{\sigma}{\sqrt{w}}). \qquad (5.2)$$

specifies its theoretic lower bound, where σ denotes the noise standard deviations and w the filter width. The reason that the minimal detectable contrast is lower-bounded is that the space of possible curves of the algorithm is exponential in curve length. The method search for an exponential number of curves takes polynomial time due to the dynamic programming approach that is used. This bound specifies how faint an edge can be and still be detected by the classic algorithms. Although this method assumes step edges with constant contrast and Gaussian noise, it achieves accurate results in other imaging domains. Figure 5.2 shows that it achieves competitive results on the noisy BSDS [46] dataset. Indeed, my faint edge detection method that was initially developed to address step edges and Gaussian noise models, see Chapter 3, shows superiority in addressing edge detection in noisy natural and medical images.

While this suggests that deep methods are not invariant to imaging domains, one may wonder if they are not robust in general. Actually, my study shows that they are highly flexible



Algorithm	ODS	OIS	AP
Ours	**0.60**	**0.62**	**0.51**
MCG	**0.63**	**0.63**	0.24
PMI	0.47	0.53	0.32
SE	0.52	0.52	**0.35**

Figure 5.2: Simulation results of my classic FED with the noisy BSDS-500 images [46]. Left: Precision vs. Recall (PR) of contour detection by various algorithms. Right: Performance table. ODS refers to the F-Measure at the optimal threshold across the entire dataset, OIS to the best per image F-measure, and AP to the area under the PR curve.

to geometric variations such as edge curvatures and geometric transformations, if the imaging domain remains similar to the one of the training set.

The last issue investigated here is the transition of an approach to different problems. As it can be seen in [52], my CNN architecture that was designed for FED can be easily used to perform noisy image classification and natural image denoising. Table 5.3 from [52] shows that the CNN applied before a classification network improves the classification accuracy on the CIFAR10 and CIFAR100 datasets [33]. Moreover, Tables 5.4 and 5.5 indicate that natural image denoising of high quality can be carried out by a multiscale CNN trained by an edge preservation loss.

Finally, regarding data and training requirements, an obvious advantage of the classic over the deep approach is that it does not require a dataset for the development process. The classic method is based on assumptions only, while the CNN approach learns from many samples of faint edges. This, not only simplifies the development process, but also constrains the performance of the deep learning-based approach as it works better on samples that are similar to the original dataset.

Algorithm	**CIFAR10**	**CIFAR100**
$resnet(IDCNN)$	**82.7**	**53.3**
$resnet_{noisy}$	77.5	46.0
$resnet_c$	34.1	16.9

Table 5.3: Classification accuracy of the different methods and architectures averaged on all noisy levels. My approach of multi-scale preprocessing by IDCNN [72] and classification by resnet 20 [69] achieves the highest accuracy.

Algorithm	$\sigma = \mathbf{15}$	$\sigma = \mathbf{25}$	$\sigma = \mathbf{50}$
IDCNN-E	31.00/**0.9**	28.86/**0.85**	25.95/**0.75**
IDCNN	30.80/0.89	28.73/0.84	25.93/**0.75**
DnCNN	31.74/**0.9**	29.89/**0.85**	25.69/0.71
BM3D	31.07/0.88	28.26/0.81	24.57/0.67

Table 5.4: Quantitative PSNR(dB) and SSIM results of denoising on the noisy natural images from the dataset of [46]. My approach, using IDCNN for denoising achieves the high perceptual score of SSIM [74]. In addition, my distortion score of PSNR is also competitive relative to the state-of the art approach of DnCNN [78]. My edge preserving auxiliary loss improve the performance of IDCNN in denoising in both measurements. The highest SSIM scores are highlighted.

While performance scores are essential when selecting an approach, the cost of its development is also important. There are main differences between the development processes of classic and deep FED. The classic approach requires planning, analysis, parameter optimization, and complex derivation of computational complexity and threshold. Moreover, it takes a significant amount of time to write the code and optimize it. Finally, good results are hard to get and require non-negligible effort. On the contrary, the development of the deep CNN-based method achieved almost all of these in an order of magnitude

Algorithm	$\sigma = \mathbf{15}$	$\sigma = \mathbf{25}$	$\sigma = \mathbf{50}$
IDCNN-E	30.78/**0.9**	28.61/**0.84**	25.78/**0.74**
IDCNN	30.51/0.89	28.53/**0.84**	25.73/**0.74**
DnCNN	31.73/**0.9**	29.16/**0.84**	26.23/0.71
DeepAM [32]	31.68/0.89	29.21/0.82	26.24/0.72
TRD [16]	31.42/0.88	28.91/0.81	25.96/0.70
MLP [13]	-	28.91/0.81	26.00/0.71
CSF [64]	31.24/0.87	28.91/0.81	-
BM3D	31.12/0.87	28.91/0.81	25.65/0.69

Table 5.5: The average PSNR(dB) and SSIM results of different methods on the BSD68 dataset [61]. The highest SSIM scores are highlighted.

faster. Even the training time, in this case, is negligible with respect to the development time of the classic approach. Finally, optimization is performed automatically by training and testing on a GPU. In addition, the nature of the development time is different: with DL approaches, it is devoted more to dataset creation and less to code writing. Consequently, classic methods need to be developed by computer vision experts, whereas DL algorithms can be developed by more general ML specialists.

5.3 Comparison of deep learning and classic methods for multispectral image registration

I addressed the problem of multispectral image alignment using a classic approach in Chapter 4 together with a deep learning method. As shown previously, each approach has its advantages. Table 5.6 from [53] shows that the deep approach achieves the highest accuracy in VIS to NIR alignment. It outperforms my classic approach [54] that relies on a unique handcrafted edge descriptor. It is also robust to geometric dis-

Algorithm	VIS-NIR
DL solution [53]	**0.03**
Handcrafted descriptor [54]	0.08
Canny	0.07
Sobel	0.07
Mutual Information	0.11
LGHD	0.21

Table 5.6: Error in pixels of multi-spectral registration when searching for translation only. My deep method is compared to edge descriptors approach, correlation of Canny [14], correlation of Sobel [24], maximization of mutual-information and LGHD [1]. As can be seen (in bold), my deep-algorithm achieves the highest accuracy.

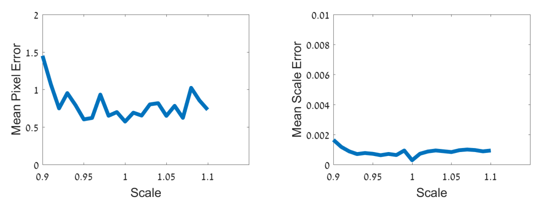


Figure 5.3: Evaluation of registration error across simulated scaling transformations. Left: error of the translation parameters when solving scales from 0.9 to 1.1. Right: error of the scaling parameter across the same range of scalings between the cross-spectral images. The translation error is around 1 pixels while the scaling error is negligible.

tortions as Figure 5.3 shows. Although the deep approach is accurate with respect to the traditional method, the handcrafted descriptor proved to be robust to other imaging domains. As demonstrated in [54], the classic approach aligns a wider field of spectral channels. While the deep only works on the VIS to NIR that it is trained on, the classic method aligns VIS to MWIR as well.

My research has proved that the feature descriptor can be engineered in a classic way, or learned by a CNN approach in a metric learning scheme. In terms of registration error, each descriptor has its advantages. While the classic descriptor is robust to spectral channels, the deep is robust to geometric variations. The deep learning approach showed that a descriptor can be useful even though I do not fully understand its meaning and what it exactly represents.

While registration error is a key element when comparing deep and classic approaches for multispectral image registration, other important aspects should also be considered. First, as the deep approach requires a forward pass of a CNN for every key point, the processing time of creating a feature descriptor is much slower than when using a classic approach. Second, while a classic approach does not require training resources, the deep learning method relies on the existence or creation of a valid multispectral database with a corresponding aligned image without which it could not operate. Moreover, its accuracy also depends on the level of information in the key-point features in that dataset. Third, both approaches have different hardware requirements: whereas the classic methods can be run easily on a standard CPU, real-time computing can only be achieved by the DL method if its execution takes place on a GPU. Not only is an expensive processing platform required, but also this prevents its usage on some embedded systems. Finally, in this specific case, there was a major difference regarding the development time that was needed to produce those two solutions. While the classic method was developed with much effort, once available, it could be quite rapidly transformed into its deep variant.

Feature/Approach	DL	Classic
Accuracy (Acc.)	**High**	Moderate
Acc. for other domains	Low	**Moderate**
Speed on CPU	Slow	Slow/**Fast**
Speed on GPU	Fast	/
Theoretical basis	Moderate	**High/Moderate**
Training dataset	Essential	**No**
Geometric variability	**Robust**	Weak
Development	**Fast**	Slow
Repurposing ability	**High**	Low

Table 5.7: Comparison of the features of DL and classic approaches observed in this study.

5.4 Conclusion

Table 5.7 summarizes the advantages I found in this study between deep and classic approaches. Although it reveals that the DL solutions achieve higher accuracy than the classic ones, the DL algorithms are not invariant to imaging domains as their usage is restricted by the nature of their training sets. Still, CNN showed high flexibility to complex geometric relations and objects. The theory behind DL is focused on the loss function and the CNN architecture, while in classic methods, the developer can prove an algorithm's thresholds and bounds. One also should note that due to the general nature of the CNN architecture, a CNN-based algorithm can easily be reused for other vision tasks.

The DL methods discussed in this chapter have intrinsic limitations. First, there is the black-box effect: I do not understand either the CNN filter for FED or the invariant descriptor for multispectral image registration that is learned by my teacher-student approach. This is a serious drawback even if the development of methods for interpreting and understanding deep neural networks is a very dynamic research area [49]. Moreover, the training phase of their development requires a large footprint with respect to the corresponding classic approaches. The classic methods also have their qualitative advantages. They proved more robust to image types and modalities in both problems. This is a consequence of their simplifying assumptions, and the fact to they are not example based.

Deep learning has revolutionized the way computer vision problems are addressed. It has stretched the limits of what is now possible and regularly achieved the best quantitative scores on benchmarks. Moreover, it has made the algorithm development a process of data collections and neural network training using loss function. However, traditional methods still have their strengths, e.g., a lot can be learned from their development process. Moreover, as they are mostly based on proven theories, their algorithms can be described to humans in detail. This transparency is important as it provides a sense of trust and, therefore, facilitates their acceptance in real-life applications. Thus, there is no doubt that trust and transparency are important challenges that need to be met to ensure the general adoption of DL solutions.

Chapter 6

Conclusions

6.1 Summary of contributions

THE study described in this thesis has achieved several contributions. I conducted a meaningful comparison between classic and deep learning approaches for addressing challenges in computer vision. The study was performed through the perspective of the two problems, faint edge detection, and multi-spectral image registration. I introduced a nearly linear complexity approaches for straight and curved edge detection. I developed theoretic thresholds and lower bounds related to the classic FED, addressing the question of how faint an edge can be and still be detected by my method. In addition, I introduced a CNN approach for FED. This approach delivered state-of-the-art results in my simulations. Moreover, the CNN that was used for FED showed its robustness to similar computer vision problems such as classification and noisy images and natural image denoising.

I also developed a classic feature-based approach for multi-spectral image alignment that achieved a high quality of registration under a wide field of spectral channels, as fusion results based on those alignments showed. In addition, I improved this classic approach by making it invariant to the descriptor by learning it using a teacher-student scheme.

As this whole study has given us insights into computer vision challenges and the ways to address them, I could conduct a comparison between classic and deep learning approaches. I revealed that deep methods are an order of magnitude faster to develop. In addition, they usually achieve the best accuracy on test sets, and are robust to certain variabilities such as geometry transformations and lightning conditions. However, they are understood mainly as a black-box and require a large footprint to be trained. Their classic alternatives are based on deeper theoretic foundations. Moreover, due to their simple assumptions, they tend to be more robust to image type and modality.

6.2 Discussion

I have reported in this thesis that classic and deep learning approaches for addressing challenges in computer vision have each their advantages and limitations. Each approach is robust and invariant to different properties of the addressed problem. Moreover, the process of development of each approach is very different and relies on distinct theoretic concepts.

My study shows that there is a lower bound on how faint an edge can be and still be detected. If the space of possible curved edges is large enough, there are low SNRs that cannot be detected as they are below the detection threshold. It means that they are less statistically significant than pure noise.

My research teaches us that it is possible to register multispectral images, even from a wide field of spectral channels.

Fusion can be carried out under some limited geometric transformations between the channels. Moreover, I have shown that a CNN architecture can help to improve both alignment accuracy and the robustness of alignment in complex relations.

This thesis emphasizes that deep learning methods are more accurate on test sets, and more invariant to object geometry and lightning conditions. On the other hand, as classic methods are able to handle other image types, they may be developed on simulation and still to be accurate on real images. I have also highlighted that the development process of each approach is different. While DL is faster to develop, it requires a large footprint. Contrarily, classic methods are harder to develop and require expert knowledge, however, they are associated with insights on the theoretical foundations of their approach.

6.3 Closing remarks

It has been a pleasure to work on this study tackling challenging problems, and comparing two main approaches in computer vision. I would also like to thank the reader for their patience. I hope they have found this thesis interesting and informative.

Bibliography

[1] C. Aguilera, A. D. Sappa, and R. Toledo. Lghd: A feature descriptor for matching across non-linear intensity variations. In *IEEE International Conference on Image Processing*, page 5. IEEE, Sep 2015.

[2] C. A. Aguilera, F. J. Aguilera, A. D. Sappa, C. Aguilera, and R. Toledo. Learning cross-spectral similarity measures with deep convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–9, 2016.

[3] C. A. Aguilera, A. D. Sappa, and R. Toledo. Lghd: A feature descriptor for matching across non-linear intensity variations. In *IEEE International Conference on Image Processing*, pages 178–181. IEEE, 2015.

[4] N. Akhtar and A. Mian. Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access*, 6:14410–14430, 2018.

[5] A. Alwosheel, S. van Cranenburgh, and C. G. Chorus. Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of choice modelling*, 28:167–182, 2018.

[6] M. Arar, Y. Ginger, D. Danon, A. H. Bermano, and D. Cohen-Or. Unsupervised multi-modal image regis-

tration via geometry preserving image-to-image translation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 13410–13419, 2020.

[7] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, May 2011.

[8] P. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 328–335, 2014.

[9] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[10] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Transactions on image processing*, 7(3):421–432, 1998.

[11] D. Bojanić, K. Bartol, T. Pribanić, T. Petković, Y. D. Donoso, and J. S. Mas. On the comparison of classic and deep keypoint detector and descriptor methods. In *International Symposium on Image and Signal Processing and Analysis*, pages 64–69. IEEE, 2019.

[12] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007.

[13] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *IEEE conference on computer vision and pattern recognition*, pages 2392–2399. IEEE, 2012.

[14] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[15] J. Canny. A computational approach to edge detection. In *Readings in Computer Vision*, pages 184–203. Elsevier, 1987.

[16] Y. Chen, W. Yu, and T. Pock. On learning optimized reaction diffusion processes for effective image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5261–5269, 2015.

[17] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE computer society conference on computer vision and pattern recognition*, volume 1, pages 886–893. IEEE, 2005.

[18] D. DeTone, T. Malisiewicz, and A. Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 224–236, 2018.

[19] P. Dollár and C. L. Zitnick. Structured forests for fast edge detection. In *IEEE International Conference on Computer Vision*, pages 1841–1848. IEEE, 2013.

[20] L. Du. How much deep learning does neural style transfer really need? an ablation study. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3150–3159, 2020.

[21] P. Fergus, A. Hussain, D. Hignett, D. Al-Jumeily, K. Abdel-Aziz, and H. Hamdan. A machine learning sys-

tem for automated whole-brain seizure detection. *Applied Computing and Informatics*, 12(1):70–89, 2016.

[22] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[23] M. Galun, R. Basri, and A. Brandt. Multiscale edge detection and fiber enhancement using differences of oriented means. In *IEEE International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[24] W. Gao, X. Zhang, L. Yang, and H. Liu. An improved sobel edge detection. In *Computer Science and Information Technology*, volume 5, pages 67–71. IEEE, 2010.

[25] J. F. Haddon. Generalised threshold selection for edge detection. *Pattern Recognition*, 21(3):195–203, 1988.

[26] K. Hao. Training a single ai model can emit as much carbon as five cars in their lifetimes. *MIT Technology Review*, 2019.

[27] D. Haussler. *Probably approximately correct learning*. University of California, Santa Cruz, Computer Research Laboratory, 1990.

[28] I. Horev, B. Nadler, E. Arias-Castro, M. Galun, and R. Basri. Detection of long edges on a computational budget: A sublinear approach. *SIAM Journal on Imaging Sciences*, 8(1):458–483, 2015.

[29] P. Isola, D. Zoran, D. Krishnan, and E. H. Adelson. Crisp boundary detection using pointwise mutual information. In *European Conference on Computer Vision*, pages 799–814. Springer, 2014.

[30] M. Jaderberg, K. Simonyan, A. Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.

[31] D. Justus, J. Brennan, S. Bonner, and A. S. McGough. Predicting the computational cost of deep learning models. In *IEEE International Conference on Big Data*, pages 3873–3882. IEEE, 2018.

[32] Y. Kim, H. Jung, D. Min, and K. Sohn. Deeply aggregated alternating minimization for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6419–6427, 2017.

[33] A. Krizhevsky, V. Nair, and G. Hinton. The cifar-10 dataset. *online: http://www. cs. toronto. edu/kriz/cifar. html*, 2014.

[34] A. Kurakin, I. Goodfellow, and S. Bengio. Adversarial examples in the physical world. *arXiv preprint arXiv:1607.02533*, 2016.

[35] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[36] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[37] Y. LeCun, C. Cortes, and C. Burges. Mnist handwritten digit database. *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, 2, 2010.

[38] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino. Deep-plant: Plant identification with convolutional neural net-

works. In *IEEE international conference on image processing*, pages 452–456. IEEE, 2015.

[39] P. Lin. Why ethics matters for autonomous cars. In *Autonomous driving*, pages 69–85. Springer, Berlin, Heidelberg, 2016.

[40] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[41] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[42] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. 1981.

[43] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool. Convolutional oriented boundaries: From image segmentation to high-level tasks. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):819–833, 2018.

[44] S. Manzoor, S.-H. Joo, and T.-Y. Kuc. Comparison of object recognition approaches using traditional machine vision and modern deep learning techniques for mobile robot. In *International Conference on Control, Automation and Systems*, pages 1316–1321. IEEE, 2019.

[45] D. Marr and E. Hildreth. Theory of edge detection. *Proc. R. Soc. Lond. B*, 207(1167):187–217, 1980.

[46] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring eco-

logical statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision*, volume 2, pages 416–423. IEEE, 2001.

[47] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[48] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin. Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters*, 6(12):300–303, 1999.

[49] G. Montavon, W. Samek, and K.-R. Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018.

[50] N. Ofir, M. Galun, S. Alpert, A. Brandt, B. Nadler, and R. Basri. On detection of faint edges in noisy images. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):894–908, 2019.

[51] N. Ofir, M. Galun, B. Nadler, and R. Basri. Fast detection of curved edges at low snr. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 213–221, 2016.

[52] N. Ofir and Y. Keller. Multi-scale processing of noisy images using edge preservation losses. In *International Conference on Pattern Recognition*, pages 1–8. IEEE, 2021.

[53] N. Ofir, S. Silberstein, H. Levi, D. Rozenbaum, Y. Keller, and S. D. Bar. Deep multi-spectral registration using invariant descriptor learning. In *IEEE International Conference on Image Processing*, pages 1238–1242. IEEE, 2018.

[54] N. Ofir, S. Silberstein, D. Rozenbaum, Y. Keller, and S. D. Bar. Registration and fusion of multi-spectral images using a novel edge descriptor. In *IEEE International Conference on Image Processing*, pages 1857–1861. IEEE, 2018.

[55] Y. Ono, E. Trulls, P. Fua, and K. M. Yi. Lf-net: learning local features from images. In *Advances in neural information processing systems*, pages 6234–6244, 2018.

[56] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh. Deep learning vs. traditional computer vision. In *Science and Information Conference*, pages 128–144. Springer, 2019.

[57] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[58] B. S. Reddy and B. N. Chatterji. An fft-based technique for translation, rotation, and scale-invariant image registration. *IEEE transactions on image processing*, 5(8):1266–1271, 1996.

[59] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[60] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[61] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *IEEE Computer Society Con-*

ference on Computer Vision and Pattern Recognition, volume 2, pages 860–867. IEEE, 2005.

[62] W. Samek, T. Wiegand, and K.-R. Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*, 2017.

[63] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[64] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2774–2781, 2014.

[65] R. Shwartz-Ziv and N. Tishby. Opening the black box of deep neural networks via information. *arXiv preprint arXiv:1703.00810*, 2017.

[66] T. Stathaki. *Image fusion: algorithms and applications*. Elsevier, 2011.

[67] E. Strubell, A. Ganesh, and A. McCallum. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*, 2019.

[68] J. Su, D. V. Vargas, and K. Sakurai. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841, 2019.

[69] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[70] L. Torrey and J. Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends:*

algorithms, methods, and techniques, pages 242–264. IGI global, 2010.

[71] D. S. Trigueros, L. Meng, and M. Hartnett. Face recognition: From traditional to deep learning methods. *arXiv preprint arXiv:1811.00116*, 2018.

[72] R. Unnikrishnan, C. Pantofaru, and M. Hebert. A measure for objective evaluation of image segmentation algorithms. In *IEEE Computer Vision and Pattern Recognition-Workshops*, pages 34–34. IEEE, 2005.

[73] Y.-Q. Wang, A. Trouvé, Y. Amit, and B. Nadler. Detecting curved edges in noisy images in sublinear time. *Journal of Mathematical Imaging and Vision*, 59(3):373–393, 2017.

[74] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[75] C. Xiao, J.-Y. Zhu, B. Li, W. He, M. Liu, and D. Song. Spatially transformed adversarial examples. *arXiv preprint arXiv:1801.02612*, 2018.

[76] S. Xie and Z. Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

[77] J. Yang, B. Price, S. Cohen, H. Lee, and M.-H. Yang. Object contour detection with a fully convolutional encoder-decoder network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–202, 2016.

[78] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn

for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.