UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

Herberth Birck Fröhlich

Evaluation of adhesion failures in composite laminated plates using deep learning-based object detection in shearography images

Florianópolis
18 de Junho de 2021

Herberth Birck Fröhlich

Evaluation of adhesion failures in composite laminated plates using deep learning-based object detection in shearography images

Tese submetida ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Engenharia Mecânica.
Orientador: Prof. Armando Albertazzi Gonçalves Jr., Dr. Eng.
Coorientadores: Analucia Vieira Fantin, Dra. e Bernardo Cassimiro Fonseca de Oliveira, Dr. Eng.

Florianópolis
18 de Junho de 2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Fröhlich, Herberth Birck
Evaluation of adhesion failures in composite laminated plates using deep learning-based object detection in shearography images / Herberth Birck Fröhlich ; orientador, Armando Albertazzi, coorientadora, Analucia Vieira Fantin, coorientador, Bernardo Cassimiro Fonseca de Oliveira, 2021.
99 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, , Programa de Pós-Graduação em Engenharia Mecânica, Florianópolis, 2021.

Inclui referências.

1. Engenharia Mecânica. 2. Materiais compósitos. 3. Detecção de defeitos. 4. Shearografia. 5. Aprendizagem profunda. I. Albertazzi, Armando. II. Fantin, Analucia Vieira. III. Oliveira, Bernardo Cassimiro Fonseca de IV. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Engenharia Mecânica. V. Título.

Herberth Birck Fröhlich

Evaluation of adhesion failures in composite laminated plates using deep learning-based object detection in shearography images

O presente trabalho em nível de Doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Leandro dos Santos Coelho, Dr.
Pontifícia Universidade Católica e Universidade Federal do Paraná (relator, videoconferência)

Prof. Robert H. Schmitt, Dr.-Ing.
Rheinisch-Westfälische Technische Hochschule Aachen (videoconferência)

Prof. Tiago Loureiro Figaro da Costa Pinto, Dr.
Universidade Federal de Santa Catarina (videoconferência)

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Engenharia Mecânica.

Documento assinado digitalmente
Paulo de Tarso Rocha de Mendonca
Data: 09/07/2021 16:39:09-0300
CPF: 149.822.691-49
Verifique as assinaturas em https://v.ufsc.br

Coordenação do Programa de
Pós-Graduação em Engenharia Mecânica

Documento assinado digitalmente
Armando Albertazzi Goncalves Junior
Data: 08/07/2021 21:42:19-0300
CPF: 232.495.015-49
Verifique as assinaturas em https://v.ufsc.br

Prof. Armando Albertazzi Gonçalves Jr., Dr. Eng.
Orientador

Florianópolis, 18 de Junho de 2021.

Aos meus irmãos, Aléxia e Alonso.

ACKNOWLEDGEMENTS

À Pamela Angst, por me acompanhar e incentivar nessa jornada acadêmica por tanto tempo.

À minha mãe, por sempre acreditar, e aos meus irmãos, aos quais tento ser um bom exemplo.

Ao professor Armando Albertazzi, pelo voto de confiança e por proporcionar diversas lições ao longo desses anos.

A todos aqueles que tornaram minha estadia na Alemanha agradável e valiosa, em especial aqueles que mantive um contato mais próximo.

Aos meus amigos, em especial aqueles que estiveram mais próximos durante esses anos de doutorado, Artur Seibert, Bernardo Cassimiro e Lucas Iervolino.
Aos colegas do Labmetro, em especial Tiago Júnior de Bortoli, que me auxiliou na parte técnica.

À Rosana e Marieta, da secretaria do Labmetro e da Pós-graduação em Engenharia mecânica respectivamente, pelo apoio, competência e paciência.

Aos colegas e amigos do Instituto Senai de Inovação, em especial Alexandre Reeberg e Flávio Barbosa, por todo apoio e ensinamentos.

Às agências de fomento CAPES e CNPQ por proporcionarem as condições para realização da pesquisa.

RESUMO

Materiais compósitos são ostensivamente utilizados em campo para reparos em tubulações na indústria do petróleo e gás e a avaliação da qualidade dessa aplicação é feita por ensaios não destrutivos como a shearografia. A alta subjetividade presente na detecção e avaliação de imagens de shearografia dificulta a automação desse método em campo, algo necessário na indústria do petróleo e gás principalmente porque os locais de inspeção são de difícil acesso ou perigosos. A inteligência artificial pode auxiliar nessa tarefa, mais específico na forma de detectores de objetos baseados em aprendizado profundo, que por sua vez é uma subárea de aprendizado de máquina. Esses detectores são capazes de identificar a partir de observações as características necessárias para detectar a localização do defeito na imagem e classificá-lo quanto ao seu formato. O conjunto de dados de imagens de shearografia para essa tarefa foi criado a partir de defeitos artificiais que emularam falhas de adesão entre placas metálicas e a primeira camada de compósitos de fibra de vidro por meio de almofadas de politetrafluoretileno de diferentes formatos preenchidas com ar. Um modelo de detecção de objetos base foi treinado utilizando o detector RetinaNet com uma *Residual Neural Network - 50* (ResNet-50) como *backbone*. Esse modelo alcançou uma média de precisões médias de 0,772, onde a melhor precisão média esteve relacionada a detecção de defeitos quadrados, com 0,901, e a pior relacionada a defeitos semicirculares, com 0,681. De forma suplementar, foi desenvolvido um método não-supervisionado para classificar a imagem de uma franja em útil ou não e adicioná-la ao conjunto de dados como exemplo positivo ou negativo, respectivamente. Esse método, chamado de *Automatic Useful Fringe Pattern Classificatio*n (AUTO-UFP), busca poupar o trabalho oneroso que é feito comumente de forma manual. Ao comparar o conjunto de dados criado automaticamente e aquele criado de forma manual, o F-Score obtido foi de 0,901, o que mostra alta similaridade entre essas duas abordagens. O modelo treinado com esse conjunto de dados obtido automaticamente obteve 0,06% de perda na média das precisões médias em relação ao resultado do modelo-base. Ainda, ocorreu a convergência muito mais rápida quando o detector foi treinado com o dataset automático, o que aponta que o AUTO-UFP é também uma forma de preprocessamento de imagens que mantém apenas padrões de franja com características mais úteis ao detector. Esses resultados confirmam a capacidade do uso de aprendizado profundo em conjunto com a shearografia e são um primeiro passo em direção a automação desse ensaio não destrutivo e abrindo espaço para novos desenvolvimentos.

Palavras-chave: Materiais compósitos. Detecção de defeitos. Shearografia. Aprendizado profundo. Detectores de objetos.

RESUMO ESTENDIDO

Introdução

Materiais compósitos são reconhecidos por terem comportamento mecânico comparável ao de ligas metálicas tanto em desempenho quanto em eficiência estrutural, mas com apenas uma fração do peso dessas ligas e sendo praticamente imunes à corrosão. Essas características têm levado ao emprego desses compósitos em diversos campos, inclusive na indústria do petróleo e gás, onde peças inteiras são produzidas ou reparos de danos de corrosão ou seções danificadas são realizados. O reparo é realizado por camadas, em aplicações sucessivas de fibra e matriz, e falhas e imperfeições podem surgir durante esse processo como é o caso de faltas de adesão entre o compósito e o substrato no qual ele é aplicado. Usualmente, a qualidade e estado desses reparos são investigados por meio de ensaios não-destrutivos, já que reconhecidamente garantem a integridade das peças em análise. A shearografia é um método não-destrutivo, sem contato, e sensível à variações que ocorrem na superfície do material devido a um comportamento do defeito que difere de sua vizinhança após a aplicação de uma excitação, também denominada carregamento. A alta subjetividade presente na detecção e avaliação de imagens de shearografia dificulta a automação desse método em campo, algo necessário na indústria do petróleo e gás principalmente porque os locais de inspeção são de difícil acesso ou perigosos. A inteligência artificial pode auxiliar nessa tarefa, mais específico na forma de detectores de objetos baseados em aprendizado profundo, que por sua vez é uma subárea de aprendizado de máquina. Esses detectores são capazes de identificar a partir de observações previamente adquiridas as características necessárias para detectar a localização de um ou mais defeitos na imagem de shearografia e simultaneamente classificá-los quando ao formato que originou o padrão de franjas conforme a necessidade. Contudo, para que redes para esse fim sejam realmente efetivas, é necessário que as observações apresentadas para o treinamento sejam pouco ruidosas, com alto contraste e possuam parâmetros que caracterizem o defeito em questão. Além disso, quanto maior o número de classes e a complexidade das imagens, maior o número de observações necessária para treinamento. Quando esse número de observações é insuficiente, é comum o aumento artificial dessa quantidade, seja por métodos de visão computacional, seja por meio de mais aquisições pelo sistema de medição, realizada geralmente por um projeto de experimentos que cubra as possibilidades de incidência dos defeitos.

Objetivos

Existe um campo de pesquisa a ser explorado no que se refere ao uso de shearografia para inspeção de compósitos, especialmente na pesquisa pela sua automação com o intuito de permitir medições em locais perigosos. Em específico, há avanços possíveis na geração de conjunto de dados e aplicação de métodos de aprendizado profundo para detecção e quantificação de defeitos. Dada a realidade do uso da shearografia na indústria do petróleo e gás, foi decidido dar um passo em direção à redução da subjetividade associada a interpretação dos resultados de imagens de shearografia, na forma da hipótese de que o uso de detectores de objeto baseados em aprendizado profundo, até então subutilizados para esse fim, são de grande valia mesmo com imagens

ruidosas e complexas. Assim, o objetivo principal deste trabalho é analisar a viabilidade dessa abordagem para o fim específico da localização de defeitos de falta de adesão entre a primeira camada de laminado e um substrato metálico juntamente com a classificação de seu formato. Para a realização desse objetivo principal é necessário elencar outros dois objetivos secundários: o uso de shearografia para a criação de um conjunto de dados específico para esse fim, contando com a manufatura de placas e defeitos emulados; e a automatização da classificação de qualidade de um padrão de franja para facilitar a manipulação desse conjunto de dados.

Metodologia

Almofadas de politetrafluoroetileno preenchidas com ar e com diversos tamanhos e formatos foram criadas e colocadas sobre placas metálicas para serem laminadas com fibra de vidro a fim de emular faltas de adesão e possibilitar um estudo sobre a classificação de seu formato. Shearografia foi utilizada para geração de mapas de fase, que posteriormente tiveram as áreas dos defeitos devidamente delimitadas e relacionadas aos formatos quadrado, semicírculo e triângulo. Para isso, a placa era aquecida por lâmpadas por um certo período, e após o desligamento dessa fonte de calor, duzentos e cinquenta *frames* eram adquiridos conforme a temperatura ia caindo. Um detector do tipo RetinaNet com uma *Residual Neural Network - 50* (ResNet-50) como backbone foi treinado, validado e testado em exemplos originários de diferentes placas como forma de evitar vazamento de dados. Até este momento, a delimitação dos defeitos para o treinamento foi feita de forma manual assim como a identificação da utilidade dos padrões de franja, já que nem todos os *frames* adquiridos durante o ensaio eram úteis, ou seja, podiam ser considerados exemplos positivos para o treinamento. Foi desenvolvido então um método não supervisionado baseado em operações básicas de visão computacional para realizar esta última etapa. Chamado de Automatic *Useful Fringe Pattern Classification* (AUTO-UFP), este método testa um padrão de franjas quanto ao nível de ruido, contraste e números de salto de fase e fornece a utilidade da imagem em relação a parâmetros úteis experimentalmente definidos.

Resultados

O detector de objetos treinados com a utilidade das franjas classificada manualmente para construção do conjunto de dados, também chamado de modelo base, alcançou uma média de precisões médias (tradução livre de *mean average precision*) de 0,772, onde a melhor precisão média esteve relacionada a detecção de defeitos quadrados, com 0,901, e a pior esteve relacionada a defeitos semicirculares, com 0,681. A comparação entre a classificação de utilidades de padrões de franja feita de forma manual e aquela feita pelo AUTO-UFP resultou em um F-Score de 0,901, o que mostrou uma alta similaridade entre conjunto de dados. O detector de objetos treinado com esse conjunto de dados obtido automaticamente obteve 0,06% de perda na média das precisões médias em relação ao resultado do modelo-base. Esses resultados confirmam não apenas a viabilidade do método de classificação automática de utilidade de padrões de franja, mas também a capacidade do uso de aprendizado profundo em conjunto com a shearografia, sendo um primeiro passo em direção a automação desse ensaio

não destrutivo e abrindo espaço para novos desenvolvimentos.

Considerações finais

Neste trabalho foi apresentado um método para emular falhas de adesão em laminações de fibra de vidro a fim de criar um conjunto de dados para treinamento de um detector de objetos baseado em aprendizado profundo. O detector de objetos treinado com imagens de shearografia obtidas a partir desses defeitos emulados mostrou-se viável, inclusive para classificação de formatos de defeitos. Ainda, o método que analisa a utilidade do padrão de franja para inclusão ou não de uma imagem no conjunto de dados teve bom desempenho, e a perda de desempenho do modelo treinado com esse conjunto de dados é aceitável frente ao tempo economizado na análise de cada imagem. Outro benefício é o uso de imagens de shearografia com salto de fase, e não imagens com os saltos de fase removidos, o que simplifica o treinamento e uso do método ao dispensar um passo oneroso de processamento e suscetível a erros.

Palavras-chave: Materiais compósitos. Detecção de defeitos. Shearografia. Aprendizado profundo. Detectores de objetos.

ABSTRACT

Composite materials are extensively used for pipeline repairs in the oil and gas industry, and their quality assessment is done by nondestructive testing such as shearography. The high subjectivity present in the detection and evaluation of shearography images makes it difficult to automate this method in the field, which is highly desirable in such industry mainly because several inspection sites are dangerous or difficult to access. Artificial intelligence in the form of object detectors based on deep learning can help in this task since they can learn the features needed to detect the location of the defect or defects in the image and classify them regarding their shape at the same time. The dataset of shearography images for this task was created from artificial defects that emulated a lack of adhesion with different shapes and sizes. These defects were positioned between metal plates and the first layer of glass fiber composites employing polytetrafluoroethylene cushions filled with air. A base object detection model was trained using the RetinaNet detector with a Residual Neural Network-50 (Resnet-50) as the backbone. This baseline model achieved a mean average precision of 0.772. The best average precision is related to square defects detection, with 0.901, and the worst is related to semi-circular defects, with 0.681. Additionally, an unsupervised method was developed to classify the usefulness of fringe patterns and add the image to the dataset. This method aims to save the expensive work of labeling images, which is often done manually. When comparing the automatically created dataset and the manually created one, the F-Score obtained was 0.901, which shows a high similarity between these approaches. The model trained with this automatically created dataset had a 0.06% loss in mean average precision regarding the base model result. Also, much faster convergence occurred when the detector was trained with the automatic dataset, which points out that AUTO-UFP is also a form of image preprocessing that maintains only fringe patterns with features that are more useful to the detector. These results confirm the capability of using deep learning in conjunction with shearography and are a first step towards the automation of this nondestructive test while opening space for further developments.

Keywords: Composite materials. Defect detection. Shearography. Deep learning. Object detection.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Network
AP	Average Precision
AUTO-UFP	Automatic-Useful Fringe Pattern
CFRP	Carbon Fiber Reinforced Polymer
CNN	Convolutional Neural Network
DL	Deep Learning
FN	False Negative
FP	False Positive
FPN	Feature Pyramid Network
GFRP	Glass Fiber Reinforced Polymer
GPU	Graphics Processing Unit
IoU	Intersection Over Union
mAP	Mean Average Precision
ML	Machine Learning
MLP	Multilayer Perceptron
NDT	Non-Destructive Testing
OD	Object Detector
PTFE	Polytetrafluoroethylene
Q	Square
ReLu	Rectified Layer Unit
SC	Semi-circle
T	Triangle
TN	True Negative
TP	True Positive
UFP	Useful Fringe Pattern

CONTENTS

1 INTRODUCTION

1.1 BACKGROUND

Composite materials are recognized for having mechanical behavior comparable to metallic alloys in performance and structural efficiency, but with a small fraction of their weight and being practically immune to corrosion. For this reason, their use has increased in recent years, especially in fields such as shipbuilding, aeronautics, biomedicine, automotive, and the oil and gas industry (MARINUCCI, 2011; EGBO, 2020; AHMAD et al., 2020).

Although components made entirely of composite materials are used in the oil and gas industry, they are also used on another essential application: the repair of corroded or damaged sections of steel pipelines, which is also the mains focus of this work. Defects such as delamination of composite layers and loss of adhesion between the first composite layer and the metallic tube can occur during the application of the composite or its useful life. The evaluation of these defects is of great importance to avoid severest problems (SEICA et al., 2006; ROHEM et al., 2016; MARTINEZ et al., 2021; MACEDO et al., 2018; SILVA, 2017).

Usually, the integrity of composites is analyzed through Non-Destructive Testing (NDT), i.e., methods that do not cause any harm to the working piece during inspection (BĚLSKÝ; KADLEC, 2019; WANG et al., 2020). Shearography is an example of such a test, being a non-contact optical NDT method sensitive to changes in the material's surface due to delaminations, detachments, damage, or matrix cracking. Shearography measurement reveals surface deformations in the form of fringe patterns that appears after using a loading method that can be vibration, heat, and pressure, among others (LIU, Z. et al., 2011). Shearography is explained in detail in section 2.2.

However, the significant subjectivity associated with interpreting results from shearography images, namely shearograms, is worrying. This subjectivity causes ambiguous characterization and makes it challenging to apply automatic methods. Even more challenging is quantifying the defect severity and the risk of a future collapse that a component is subject to. Given the aforementioned increasing use of composites, several works have emerged and still emerge, often due to the need for industrial application methods within this scenario. A search in the literature with the terms *shearography*, *defects* and *composites* revealed a large number of papers that use this method in a qualitative way to detect defects in composites, but few that quantify the defect concerning its cause, intensity or shape, as it is seen below.

One of the most common techniques to emulate delamination defects is the flat bottom holes, which are the absence of material on the rear of the object to be inspected. It is expected that the lack of material will behave like the air layer present in delamination defects. The work of De Angelis et al. (2012) makes use of this method

to produce artificial defects in Carbon Fiber Reinforced Polymer (CFRP) and then uses an iterative approach based on the similarity between experimental and theoretical resonances to estimate the depth and diameter of the defect. The authors pointed out the accuracy of 90% while defining the contours and errors up to 9% in depth evaluation. Limitations are associated with the need to adjust parameters, such as those referring to the mechanical properties of the material to define the resonance frequencies for the vibration loading. Only circular defects shape was analyzed.

Another way to assess the physical dimensions of defects presented in shearograms is through pixel-millimeter calibration, as described by Benedet et al. (2017, 2020), who performed this study on twenty-eight pipelines repaired with GFRP. In this method, there is the construction of a relationship between the defect and the resulting fringe. Several sources of errors and difficulties impaired the results, such as duct curvatures, lens distortions, fringe formation limits, defect depth, and lack of information from manufacturers about defect construction. Artificial defects tried to emulate delamination and adhesion faults, but no details of their structure, such as material and thickness, were given. The authors also point out standardization and automation to reduce the problems related to fringe formation. It was not clear if a batch was dedicated only for calibration and a separate set for validation.

The mechanical behavior of a defect can be computationally simulated via finite elements. In this way, it is possible to create defects of the most varied shapes and thickness present in different materials with arbitrary repairs and analyze them with shearography. Naturally, the mechanical properties and the loading must be adjusted. One work that uses this technique is that of Bortoli et al. (2019), which emulates holes caused by corrosion in metal pipelines repaired with GFRP. The authors' work aims to create a lookup table with the relationship between the hole and the fringe obtained during the finite element simulation, which depends on the loading procedure, which is by pressure in this case, and the repair thickness. The application occurs by an iterative process, where the fringe obtained in an actual acquisition is compared exhaustively until the best match on the table is obtained. The method reached an average error of 13%, and the authors pointed out future work using holes with different shapes.

Still in this line is the work of Felipe et al. (2020), which also estimates the shape and size of holes but in plates with GFRP repairs and with two hole shapes: circular and ellipsoid. The difference with the work of Bortoli et al. (2019) lies in the loading, which was done by vibration and pulsed heat. Factors that impaired the results were related to resin on the edges of the holes, which decharacterizes them; lack of resolution in the finite element mesh; and problems during loading. The authors reiterate the need for good quality fringes, i.e. high contrast and low noise, and the strong dependence that finite element simulation has on the mechanical properties of the materials in use.

Wavelets transformation was the method chosen by Revel et al. (2017) to es-

timate defect size in composites. They also proposed to reduce the uncertainty in shearography tests by correcting the lateral displacement as a function of the defect under analysis. Wavelets were also used by Allevi et al. (2019) to improved the defect size and morphological estimation in shearography inspection. These last authors point out limitations in performing quantitative defect measurements from the phase profile boundaries, representing boundaries. In these two works, no classification is made, but rather the quantification of a defect area, and only for a single defect on the image.

In addition to quantifying the defect, some works seek to improve the steps involved in shearography testing or facilitate its automation. Fantin et al. (2013) have developed a tool for highlighting fringes to facilitate the interpretation of shearography results. This method is based on the calculation of the image divergent, which tolerates parallel fringes but highlights those irregular ones as defects. The work of J.-F. Vandenrijt and Georges (2014) can also be cited as an auxiliary method since it localizes one or mores fringes in the shearography image. This is done by calculating the gradient on the X-axis and Y-axis of the shearogram, binarizing this gradient, and post-processing the result, yielding a mask that highlights only the defect location.

These works point out the effort to improve and speed up acquisitions, making the method more appropriate for industrial environments that are often dangerous. Tools currently widely used in several industry areas and can further help these improvements are based on Artificial Intelligence (AI), specifically machine learning in the form of artificial neural networks. This type of algorithm represents a multidisciplinary approach involving many fields by merging logic, rational deduction, expert systems, and symbolic learning to take reasonable actions, seeking the best decision according to the situation to which it was presented (RUSSELL; NORVIG, 2009; ENGELBRECHT, 2007).

Results are provided in the master's thesis work of Di Martino (2015), where convolutional neural networks were used to detect anomalies in composites after their manufacturing. An interesting suggestion given by the author for future work was the possibility of combining two convolutional neural networks, one of them selecting a region of interest in the image and the other performing the learning of defect patterns.

The common characteristic of the works developed before is the insertion of numerical parameters in the networks and not directly from images (DI MARTINO, 2015). This is expected since neural methods for object detection and recognition developed more prominently after the 2000s. The introduction of convolutional networks, for example, was by LeCun et al. (1998) and since then, algorithms have been created, and processing power has increased mainly due to the use of Graphics Processing Unit (GPU). This kind of processing unit was initially used for rendering, but its usage came to light in the field of machine learning because they are capable of parallel processing (LECUN et al., 2010, 2015; VISIN et al., 2015; KRIZHEVSKY et al., 2012; RUSSAKOVSKY et al., 2015; SIMONYAN; ZISSERMAN, 2015).

The literature reveals few papers that relate shearography to machine learning and even fewer that link these two terms to quantitative defect analysis. There are only three papers in the Scopus database that match these keywords and their variations. In the work of Fröhlich et al. (2018), there is a first attempt to identify the presence of one or more defects in shearography images obtained from pipelines repaired with GFRP. This binary classification is performed by a proprietary architecture convolutional neural network developed for a low number of observations. With an accuracy of 79% in the validation set, the authors pointed out the low number of examples as a limiting factor (279 images for training and validation). Fröhlich et al. (2018) did not use an Object Detector (OD) thus there was not a defect localization.

The work by Chang et al. (2020) used Faster Region-CNN (Faster R-CNN) type detectors to detect blisters in tire manufacturing, both in sidewall and treads. It is evidenced and also mentioned by the authors the superiority of object detectors over classical methods such as support vector machines, random forest, and Haar classifiers.

In Ye et al. (2019) there is a development closer to the subject of this work. The authors employed a faster R-CNN type network to detect defects in pipeline thermal protection layers. The input images come from rotating and stitching images across the pipeline section. Again, there was no multiclassification, but just a warning of the defect presence and its location in the image. The size of the dataset was also problematic, with 200 images to be divided into training and validation. This work and the work of Fröhlich et al. (2018) are detailed in section 2.4.

What feeds machine learning methods are the observations that make up the dataset. This is one of the possible factors that hinder applying AI in combination with shearography, since obtaining authentic images to create data sets is expensive, either because the defects are not naturally present or because it is challenging to quantify defects properties. An alternative is the creation of artificial defects that emulate those present in real applications. Besides creating flat bottom holes, delamination defects can also be made by pulling out material from the side of the working piece or inserting interlayer materials, such as Teflon Polytetrafluoroethylene (PTFE) or Kapton tape (polyimide).

Creating artificial defects produces the same response in ultrasound testing, blocking the waves as air would, but the same does not always apply for the other NDT. Sometimes thermography and shearography produce different responses even to each other. Teflon, for example, shows more substantial deformation compared to the other two methods (BLAIN et al., 2017). This is also pointed out in the work of J-F. Vandenrijt et al. (2019), who insightfully investigated these techniques in CFRP. They state the difficulty of detecting delamination defects in general by shearography methods, at least when the defects are at specific depths.

1.2 OBJECTIVES

The cited works, both those that deal with the quantification of defects and those involving artificial neural networks, show possibilities of development in using shearography for composite inspection, especially in the research for its automation. Advances are possible in the area of dataset generation by creating artificial defects and using artificial intelligence tools to locate and classify the most varied types of defects. Given the reality of the author's laboratory where this work was developed, and the use of shearography in the oil and gas industry, it was decided to take a step towards reducing the subjectivity associated with the interpretation of shearograms, as it is hypothesized that object detectors, under-utilized so far in this area of interest, are of great value, even using shearography images that are noisy and highly complex.

1.2.1 Main objective

The main objective of this work is to analyze the feasibility of using deep learning-based object detectors to locate and classify the shapes of lack of adhesion defects in metallic plates repaired with GFRP. Some steps are necessary to achieve this goal, and they are presented in the next section.

1.2.2 Specific objectives

The **(i) creation of a data set is the first specific objective** since there is no consolidated and correctly cataloged data set of composite failures, specifically lack of adhesion, in the literature. Such a development relies on the design of plates with emulated defects.

This data set is the starting point for the development of intelligent solutions, such as **(ii) location of the defect in a shearography image using deep learning-based object detectors** and **(iii) classification of the shape of the defect that originated the fringe pattern**.

The **(iv) cataloging of defects images in the form of useful or non-useful fringe patterns** has been automated, as it is a costly process, highly subjective, and prone to errors when done manually. This is a parallel development and is an alternative to facilitate the achievement of the overall goal.

1.3 WORK STRUCTURE

The remainder of this work is structured as follows:

- Chapter 2 presents the theoretical background necessary to understand the methods that were used in this work, such as those that permeate the application of

composites in plates, the acquisition of images by shearography, the use of artificial neural networks for detecting objects in images, and the metrics to evaluate the results in this context.

- Chapter 3 presents the contributions proposed by this work, that is, the design and construction of artificial adhesion defects, the acquisition parameters by shearography, the creation of the dataset manually, and the method of automatic dataset creation.
- Chapter 4 has the results of the shearography tests, the quantification of the datasets created, and the comparison of object detectors created from the manually built dataset with the automated dataset.
- Chapter 5 finalizes this work with the conclusion, a discussion about the results, limitations of the proposed methods, and suggestions for future work.

2 THEORETICAL FUNDAMENTALS

This chapter provides the theoretical background necessary to understand what was developed in this work. It starts from the basics about composite materials in section 2.1, covering the production of specimens and some applications. In section 2.2 are the fundamentals of shearography, the non-destructive test used in this work. Section 2.3 explains machine learning fundamentals, extending to convolutional neural networks, object detection, and evaluation metrics. This chapter ends with section 2.4, where two works that apply machine learning with shearography images are detailed.

2.1 COMPOSITE MATERIALS

Composite materials result from the manufacture of two or more distinct chemical or physical phases properly arranged with an interface separating them. In the end, this union possesses characteristics that are not present in each component alone (KRISHAN, 2012). The intent is to obtain lightweight materials with mechanical properties comparable to many metallic alloys. Such properties depend on manufacturing attributes such as matrix orientation, the volume fraction of the reinforcement, and chemical and the mechanical composition of the phases (MALLICK, 1997). Figure 1 shows how the fiber, or reinforcement, is surrounded by matrix and some weave styles that the fiber fabric can take. In figure 2 there are real examples of actual fiber woven.

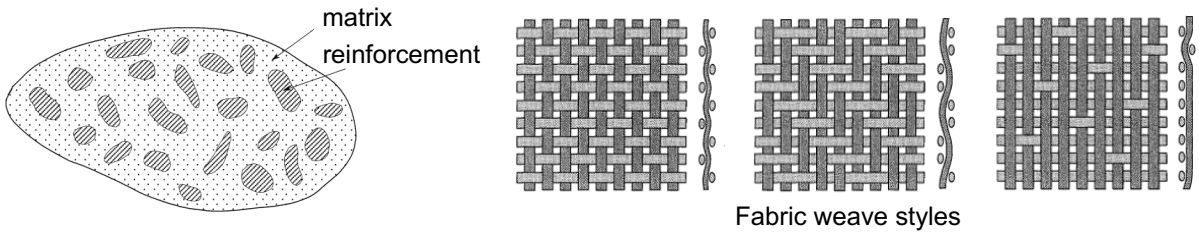


Figure 1 – Composite material structure and some fiber fabric weave styles. Adapted from Berthelot (1999).

(a) Carbon fiber woven (BENCHMARK, 2021).

(b) Glass fiber (C&CN, 2021).

Figure 2 – Example of fiber woven.

2.1.1 Composite materials on industry

Composite materials are commonly used in applications where there is a need for lightweight and high mechanical strength, as in aeronautics, automotive, marine, and, more recently, the oil and gas industry.

In aeronautics, composite materials constitute tails, wings and fuselages, displacing conventional materials such as aluminum and titanium alloys as primary structures. Still, composite materials are not only used to build parts but also to repair these structures, as seen in figure 3, where a helicopter stabilizer is repaired with composite material patches (IRVING; SOUTIS, 2020; PAVLOPOULOU et al., 2015).

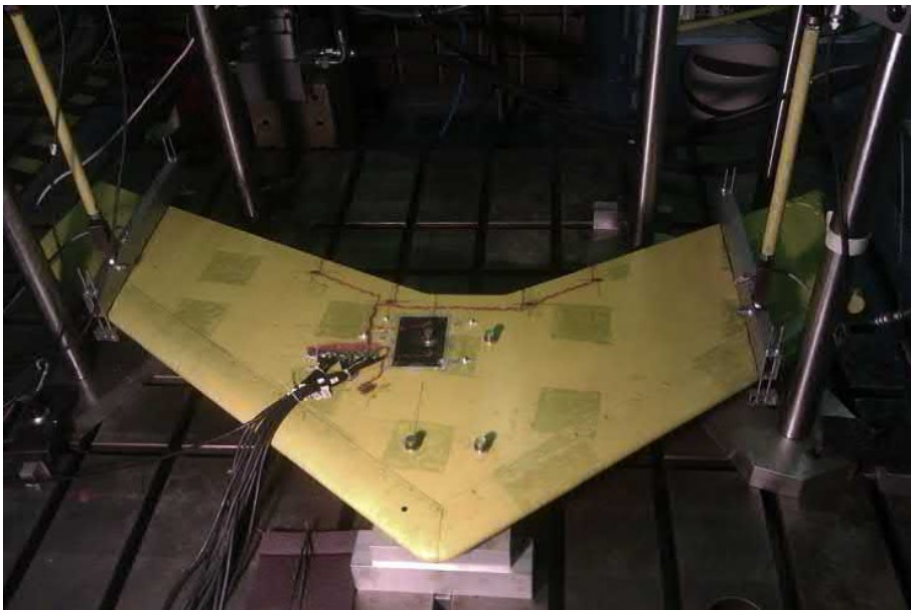


Figure 3 – Composite bonded patch repair of a helicopter stabilizer (PAVLOPOULOU et al., 2015).

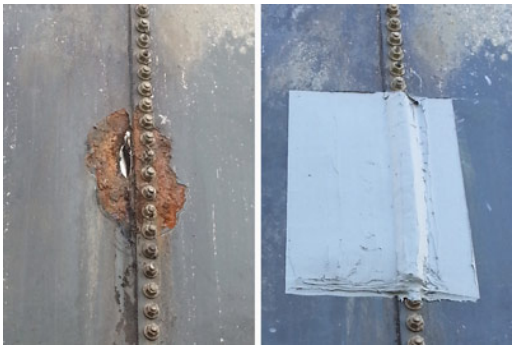
In the automotive industry, just like in aeronautics, composite materials decrease the total weight of vehicles. This accomplishment reduces fuel consumption, being used mainly to construct structural elements, such as sat structures, front panels, wheels, door cars roof, spoilers, etc (RAVISHANKAR et al., 2019; FARUK et al., 2017).

Materials for applications in the marine industry demand more attention than in the above sectors since there is also the influence of wind, tides, waves, and saltwater. These environmental challenges are present in the use of watercraft, submersibles, offshore structures, and pipelines, for example. Also, many of these structures are difficult to access, making maintenance difficult. Therefore, composites in this situation must still offer corrosion resistance and be reliable over long periods (RUBINO et al., 2020).

There are also critical applications of composite materials in the oil and gas industry. Metal pipelines carrying these substances are susceptible to corrosion and consequent failures, even more in offshore scenarios. Also, due to the fluid exerts'

internal pressure, minor defects grow much faster over time (RUBINO et al., 2020). In this sense, repairs are made in patches or gloves, which wrap or fully cover the defective or defect-prone area. Figure 4 shows a pipeline repair in the form of a glove and a tank repair in the form of a patch. The process of applying the composite is called lamination, and it will be discussed in section 2.1.3.

(a) Pipe repair (BUDHE et al., 2018).



(b) Tank repair (BIG, 2021).

Figure 4 – Repair with GFRP.

The lamination process takes place using both CFRP and GFRP. The difference between these materials lies in the density and strength, as presented in table 1. Despite the strength of carbon fiber composites, the cost of this material exceeds those of GFRP (RUBINO et al., 2020).

Table 1 – Properties of reinforced fibers (CHALMERS, 1991).

Material	**Density (g/cm^3)**	**Fiber Diameter (μm)**	**Young's Modulus (GPa)**	**Tensile Strength (GPa)**
Carbon HM	1.8	7-10	400	20-2.8
Carbon A	1.9	7-10	220	3.2
Boron	2.6	130	400	3.4
E-Glass	2.5	10	70	1.5-2.0
S-Glass	2.6	10	90	4.6
KEVLAR 49	1.45	12	130	3.6
polyethylene	0.97	12	117	2.6

2.1.2 Defects on composite materials during oil & gas applications

Corrosion is the major problem in pipelines and occurs both internally and externally. Two important applications concern repairing pipelines to avoid propagating these failures and the reinforcement of tanks or ships with already weakened metal walls. Figure 5 shows some types of defects caused by corrosion. Internal corrosion takes the form of pitting, and it is uniform, but it can also be induced by fluid flow. External corrosion is the result of oxidation and can lead to cracking due to stress (KARAMI, 2012). Real examples of these defects are seen in figure 6.

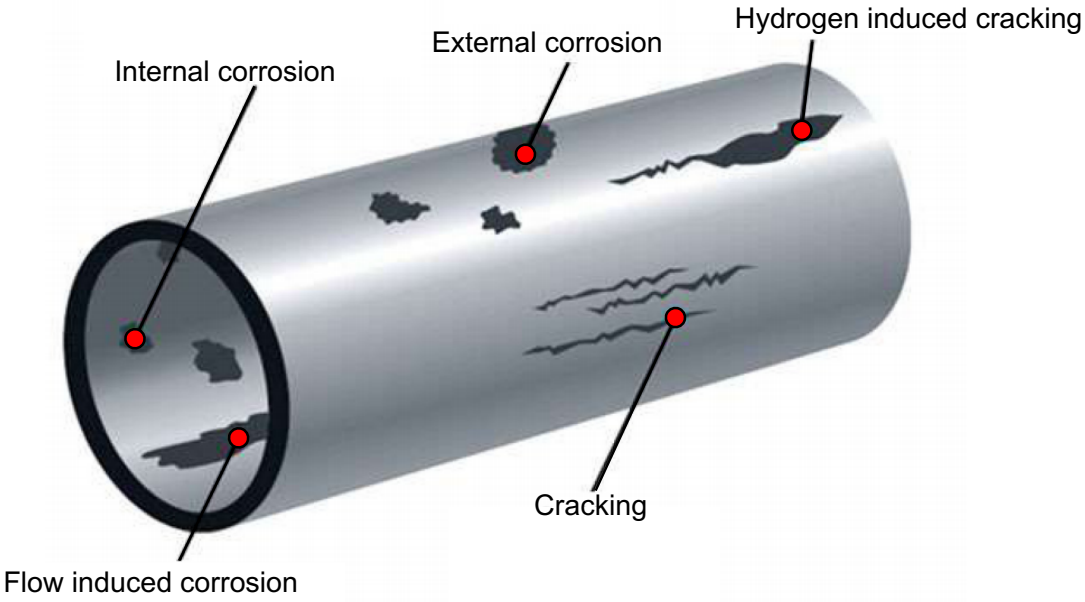


Figure 5 – Types of defects caused by corrosion on a pipeline. Adapted from Karami (2012).

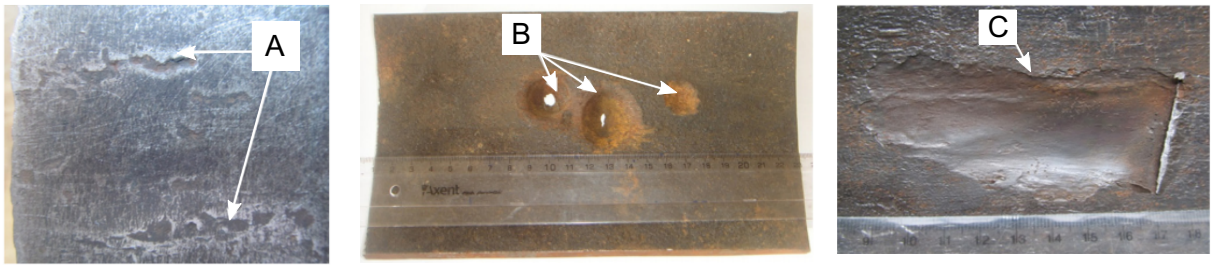


Figure 6 – In-service defects of oil and gas pipelines: pitting signalized by A, corrosion pits by B, and film peeling by C. Adapted from Maruschak et al. (2016).

Defects can also occur during the manufacturing and placement of composite materials or their usage. During manufacturing, the most common defects are porosity and small voids in the matrix, both caused by curing times, temperature, pressure, or vacuum in the resin flow (MALLICK, 1997; SMITH, 2002; GHOBADI, 2017). Heslehurst (2014) points out the existence of fifty-two possible defect types on composites, ranging from microscopic flaws to significant impact damages. They are divided according to when they arise during service life, relative size, and location or origin.

A common defect is delamination, which is the lack of contact between two layers of laminated fiber. This defect can also occur between the laminate layer and its substrate, which can be made of another composite material and originates during lamination or use. There are a few ways to emulate a delamination defect, as shown in figure 7. The intention is that the inspection system identifies the lack of material in the region of interest and changes the measurement result relative to the defect

neighborhood. Delamination is achieved by removing some material directly from the rear of the piece, also called flat bottom holes; by inserting materials between layers, such as PTFE or Kapton tape; or by systematically removing intermediate layers, a method called pull-out, which is simplified if this removal occurs at the sides of the part of interest (BLAIN et al., 2017; VANDENRIJT, J-F. et al., 2019).

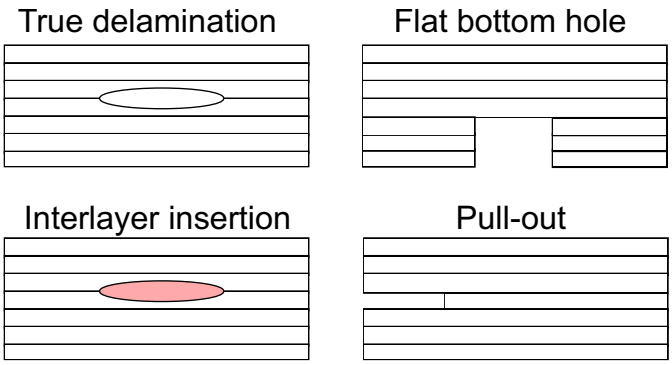


Figure 7 – True delamination and possibilities to emulate such a defect. Adapted from Blain et al. (2017).

Figure 8 shows artificial delaminations of the flat-bottom-hole type created on metal plates coated with GFRP. Naturally, the inspection is performed on the face with the composite. In this work, the working pieces are also metal plates coated with GFRP. However, the delamination defect will be emulated by air-filled PTFE cushions, a detailed process in the 3 section. For the sake of clarity, delamination defects between the metal substrate and the first layer of laminated fiber are referred to as lack of adhesion hereafter.

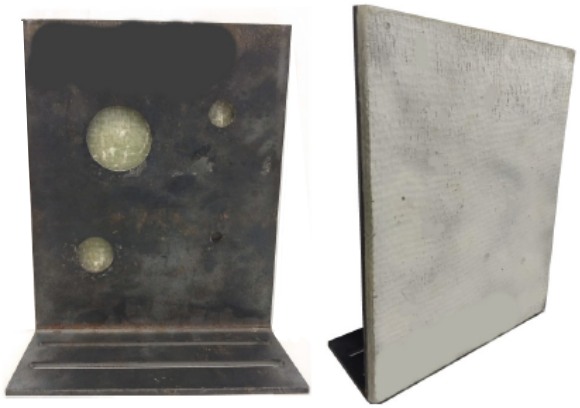
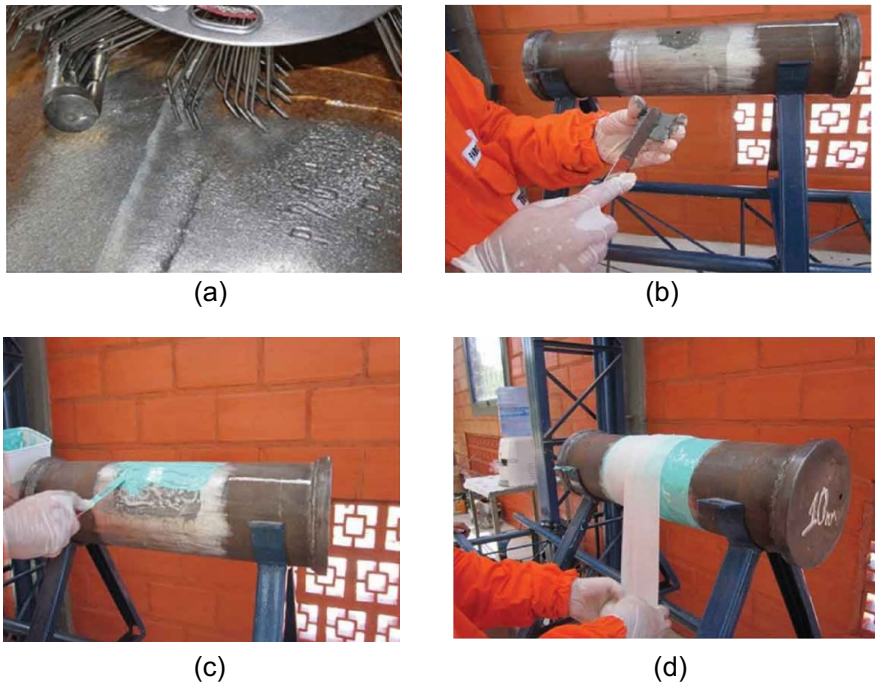


Figure 8 – Example of flat bottom holes for emulating delamination on a metal plate with a GFRP lamination. Adapted from Felipe et al. (2020).

2.1.3 Composite lamination

Lamination is the process of preparing and applying one or more layers of composite on another material and bond these layers with an adhesive material, i.e., the

matrix. Standards define the lamination steps in the oil and gas industry, including even the thickness of the repair. The first step of a field lamination procedure is to clean the metal substrate of solid residues like paint, rust, grease, and liquids (oils and solvents). Next is to prepare the surface by filling in the undercuts with resin. The next step is to spread the resin in and around the defect so that the fiberglass wrapping covers the entire area of interest. After that, one has to wait for the composite to harden, induced by vacuum, for example. These steps are shown in figure 9, in a process that follows the DNVGL-RP-C301 (2015) standard.



(a) (b)

(c) (d)

Figure 9 – Procedure for repairing a steel tube using glass fiber/epoxy composite in an offshore unit: (a) surface cleaning; (b) surface preparation; (c) epoxy resin application; and (d) glass fiber wrapping. Adapted from Barros et al. (2017).

The process mentioned above is performed in the field when there is no possibility of removing the compromised pipeline segment. It is possible to perform this lamination automatically when in controlled environments, both for correction and for strengthening of the pipelines (or other structures, such as wind turbines) before use. If performed this way, lamination is also called fiber winding. Figure 10 shows a schematic of such a procedure. Fiber strands are combined before passing through a resin bath and are then wound onto the part of interest. The number of yarns, the type of resin, the yarn tension, and the rotation speed is some of the adjustment parameters that consider the final application of the part.

After repairing these pipelines, it is still necessary to follow the progression of defects to know when a new intervention is required — NDT such as shearography help in monitoring this evolution. NDT are often used to ensure the integrity of components and coatings of composite materials.

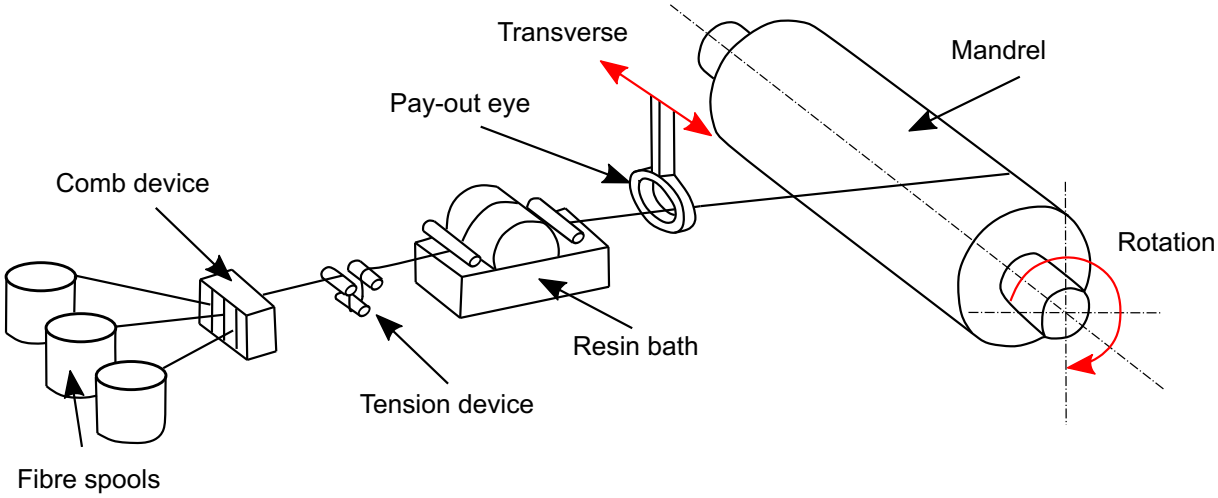


Figure 10 – Schematic diagram of filament winding technique. Adapted from Ma et al. (2019).

2.2 SHEAROGRAPHY

Speckle pattern shearing interferometry, or just shearography, is a speckle interferometric method that measures the displacement difference between overlapped points of laterally-displaced images. With such an information, it is possible to calculate the derivative of the strain field on a surface. In the original article, its authors indicate shearography for localization of disbounding and separations in structures of composite materials (HUNG, 1982). In shearography, the part of interest is illuminated by coherent light, an inherent necessity for creating the speckle effect. A source of coherent light, like a laser, is composed of waves whose behavior is predictable at various points in space and time.

Interference is the phenomenon that underlies shearography measurement. In it, two or more waves overlap and form a resulting wave. This superposition can be constructive, destructive, or something in between. Waves in phase cause the constructive effect, where their intensity increases. Waves with opposite phases cause the destructive effect, where the intensity decreases and may cancel each other out (HECHT, 2017).

A complex granular pattern, name speckle, emerges from interference if this source of coherent light hits on a rough surface if the roughness is appropriate. This speckle is called objective speckle when a camera acquires this image. The size of this granular aspect is a function of the characteristics of the acquisition system, such as the F-number (KARBHARI, 2013; HUNG, 1982). Figure 11 shows an example of a speckle pattern and gives an idea of its development.

The interference speckle pattern is created when two speckle patterns are laterally displaced while maintaining an interference region, as shown in figure 12. This displacement is called shear, and its magnitude must be analyzed on a case-by-case

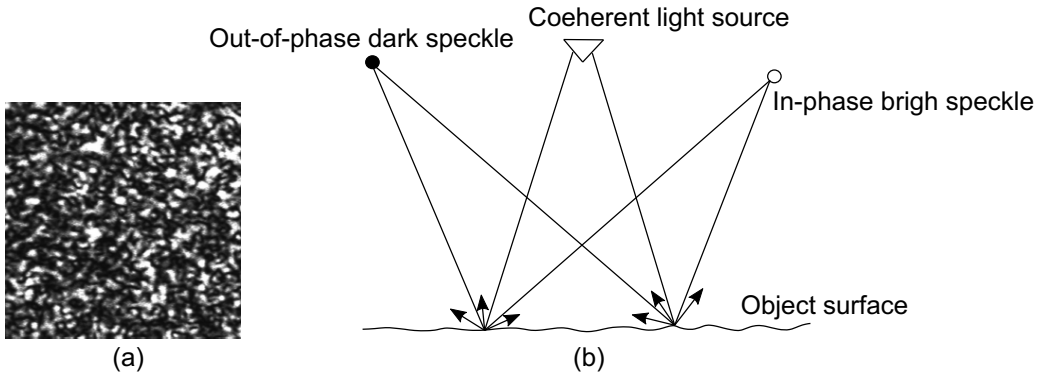


Figure 11 – (a) A typical laser speckle pattern and (b) the formation of objective speckle on an object surface. Adapted from Yang and Xie (2016).

basis because of possible loss of correlation between these two images.

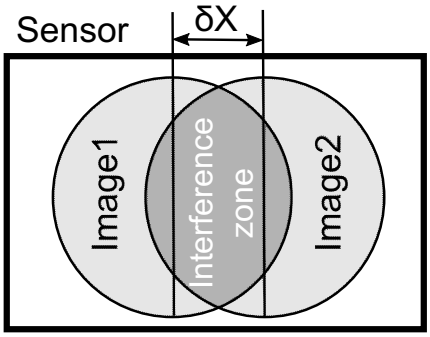


Figure 12 – Lateral displacement of known size δX. Adapted from Barrera (2018).

In this work we used a shearography configuration called one-shot, that simultaneously obtains the phase maps in three different shear directions from a single image Barrera et al. (2018). Figure 13 shows this configuration. Three wedge prisms (W) were combined in front of a triple aperture disk (L) to generate the spatial carrier fringe that contains multiple shear possibilities.

Figure 14 displays the importance of shear direction on detecting four types of groove-like defects in a plate, an influence that is directed related to the amount of movement of the lateral displacement mirror. Detection with horizontal shear, image B, is unable to access the bottom groove. The same is true for the groove-like defect in the upper left corner for the vertical shear. Such a fact highlights the correct choice requirement regarding shear direction and magnitude.

The output of the one-shot method is a speckle pattern with the carrier fringes. After a Fast Fourier Transform it is possible to define lobes (filtering) and capture only components of each shear. The figure 15 shows the result for each of these filterings.

To detect the existence of defects it is necessary to excite, or to load, the part. The phase map obtained after this loading is subtracted from the one obtained before, which is called the reference phase map. The loading is responsible for causing changes on the test material's surface, which highlights the regions that may have defects. The

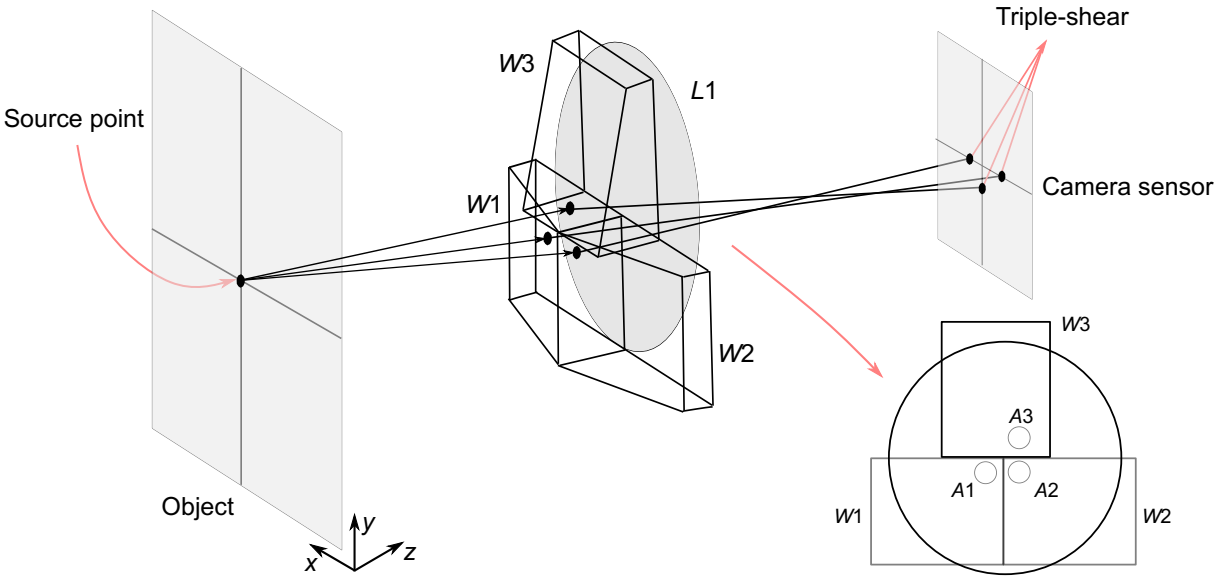


Figure 13 – One shot shearography configuration (BARRERA et al., 2018).

(a) (b) (c)

Figure 14 – Digital shearograms with different shearing directions: (a) a shape with four groove-like defects; shearogram with horizontal shearing; (c) the shearogram with vertical shearing. (YANG; XIE, 2016).

form and intensity of this loading have a major impact on the detection capability of the shearography system and are defined experimentally in most cases. Figure 16 shows four loading methods: thermal, mechanical, vibration and vacuum. Yang and Xie (2016) points out that loadings based on pressurization, vacuum, and thermal type have the best results for detecting delaminations and disbounds in composite materials when the flaws are not deep. The problem with heat loading is the lack of uniformity, as it is common for the heat to reach certain places on the plate first because it is typically done using a lamp. Two or more lamps mitigate this problem.

The next step for defect quantification is to unwrap the wrapped phase image. The wrapped phase map originally has values between π and $-\pi$. Phase unwrapping aims at removing 2π jumps discontinuities resulting from the modulus 2π range limitation of the arctan function from the phase map, which results in continuous and

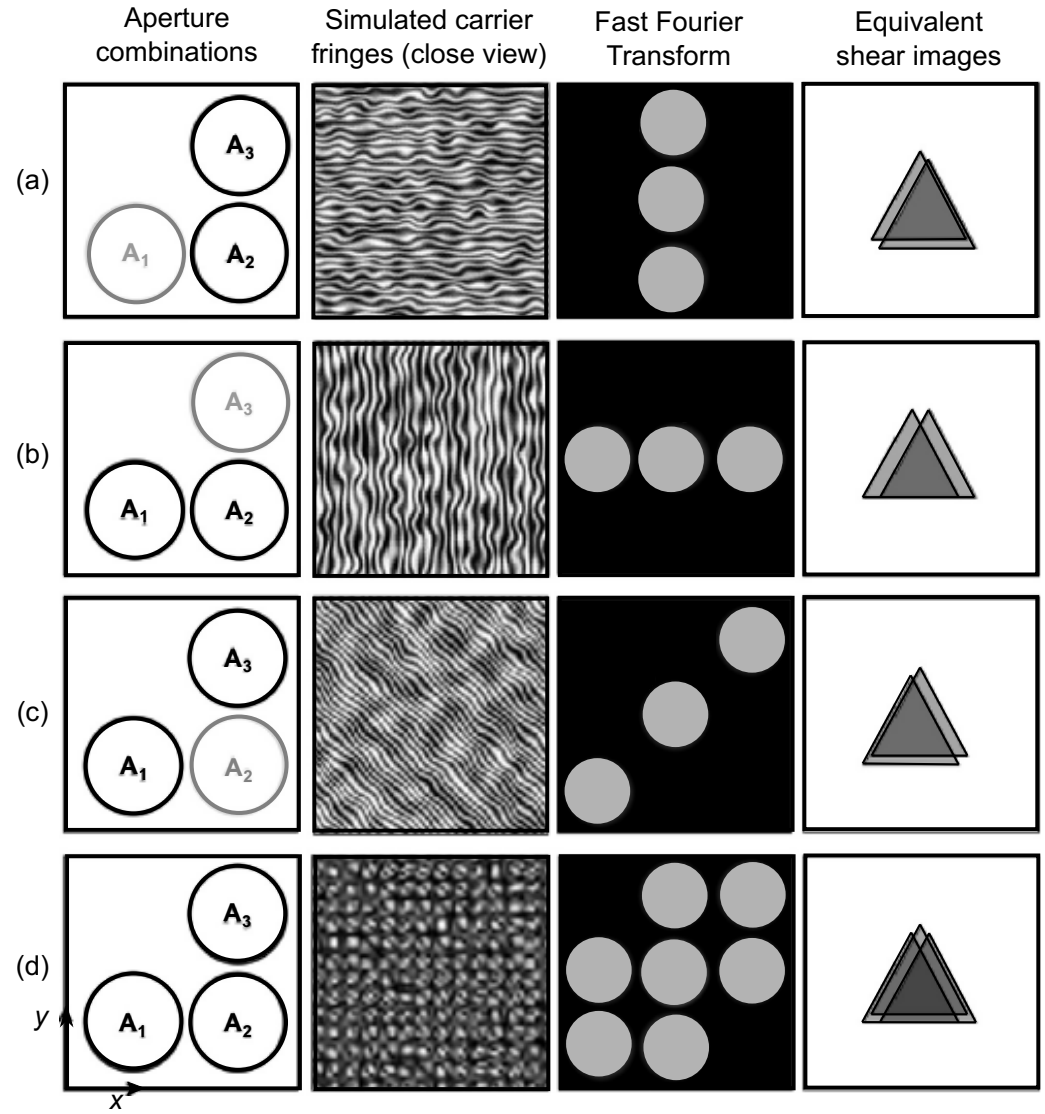


Figure 15 – Diagram of the different aperture combination, the simulated speckle carrier fringe, the equivalent spectra by Fast Fourier Transform and the equivalent shear images: (a), (b) and (c) shows the three combination apertures, (d) show the result combination of the three apertures. Adapted from Barrera et al. (2018).

quantitative measurement of the phase, recovering expected physical quantities (KARBHARI, 2013). One method to prepare the wrapped image to phase unwrapping is by sin-cos smoothing as shown in figure 17. Successive sine and cosine applications achieve this smoothing, followed by *arctan* operation. These iterations can reduce high-frequency elements and improve the contrast of the input image. This filter is indicated in this situation because traditional kernel-based filters would de-characterize the phase images(KARBHARI, 2013; AEBISCHER; WALDNER, 1999).

The phase unwrapping process involves adding multiples of 2π to the phase values of the pixels so that the absolute value of the phase difference between adjacent pixels does not exceed π (KARBHARI, 2013). Elaborate algorithms, such as the one

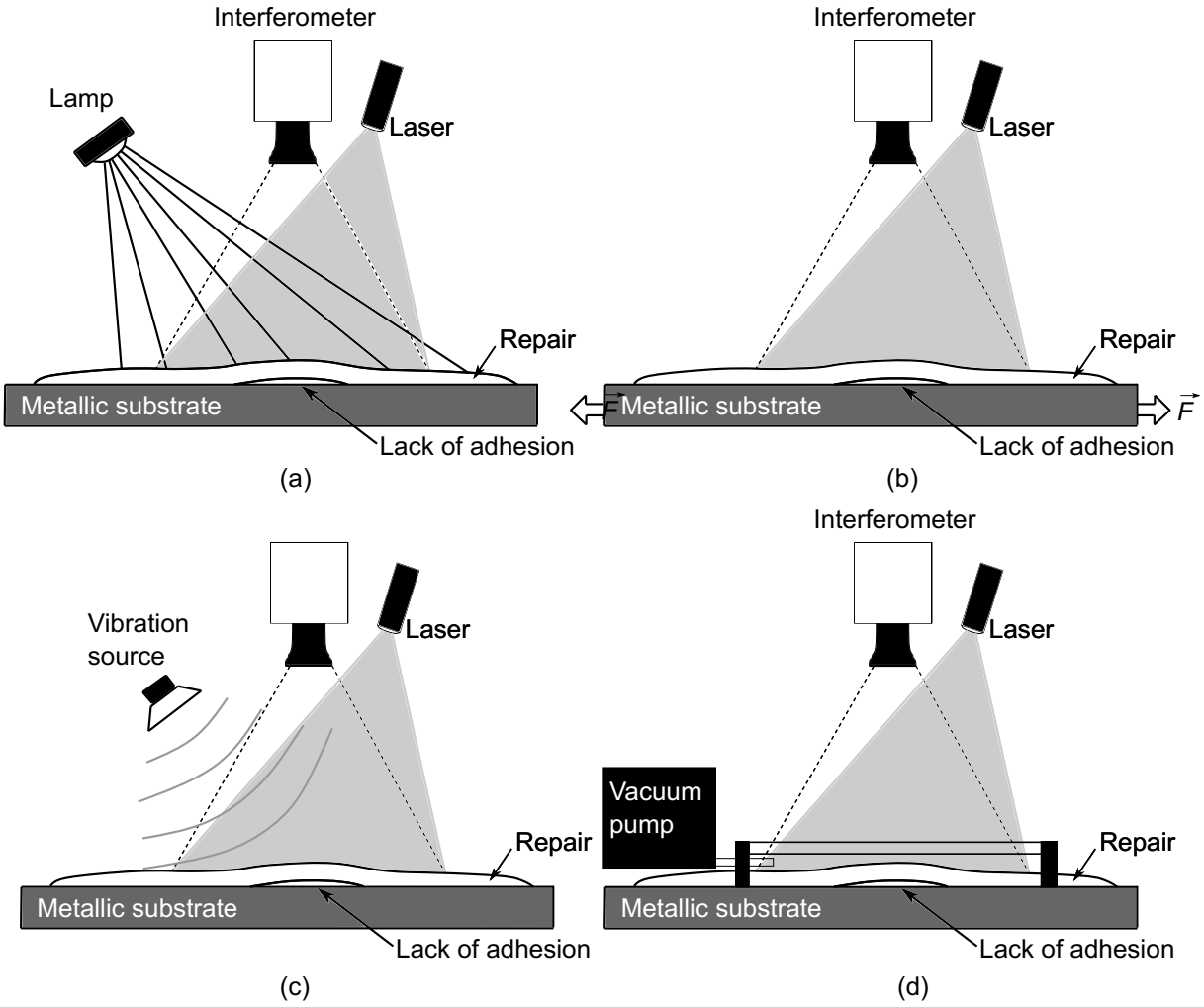


Figure 16 – Loading methods: (a) thermal; (b) mechanical loading; (c) vibration; and (d) vacuum . Adapted from Silva (2017).

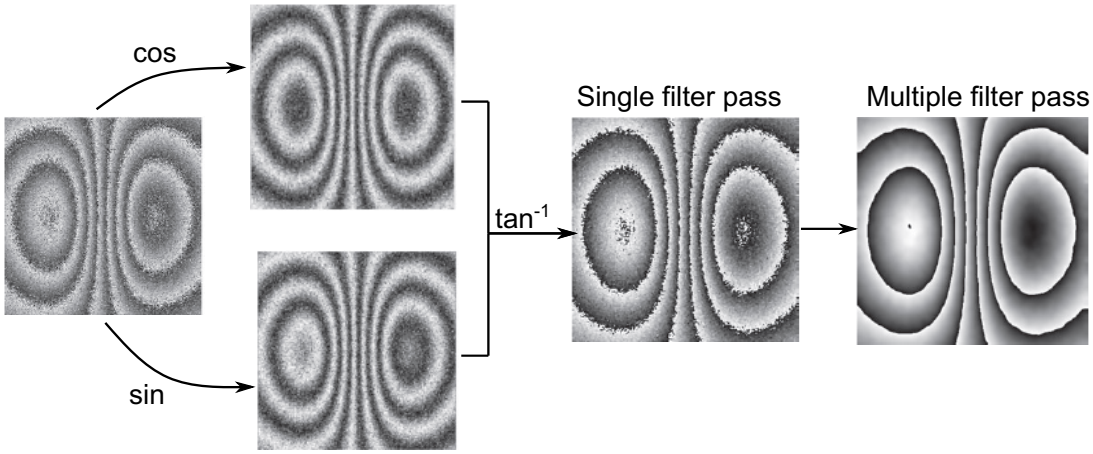


Figure 17 – (KARBHARI, 2013)

presented by Herráez et al. (2002) and used in this work, can overcome noise and discontinuities in filtered phase maps. This algorithm is of the quality-guided path type and relies on unwrapping the highest quality pixels with the highest reliability values first. Those pixels with the lowest quality and the lowest reliability are processed later

to prevent error propagation (HERRÁEZ et al., 2002; YANG; XIE, 2016). A result of an unwrapping procedure is demonstrated in figure 18, where is possible to see the discontinuities removal. After this procedure is it common to discretize the image to 8-bit resolution with values ranging from 0 to 255.

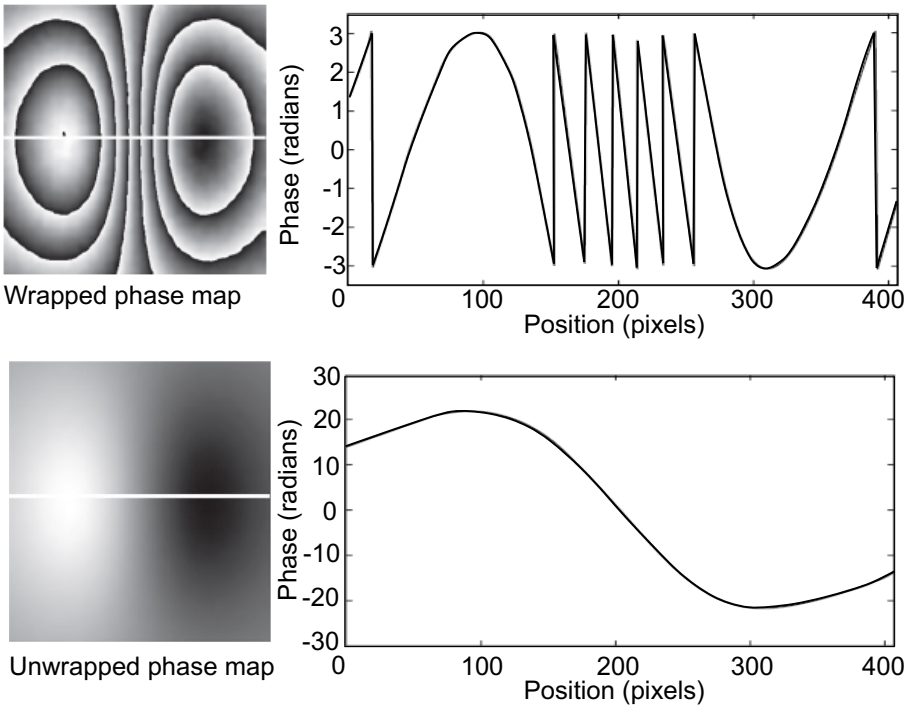


Figure 18 – Example of wrapped phase map with its discontinuities and the corresponding unwrapped phase map with discontinuities removed (KARBHARI, 2013).

2.3 MACHINE LEARNING FUNDAMENTALS

The main objective of this work is to detect adhesion failures in shearography images and classify the defect shape that originated the defect. OD are the most suitable methods for this task, as they locate the fringe in the image and simultaneously classify it. Current object detectors are Deep Learning (DL) algorithms, a branch of Machine Learning (ML), which in turn belongs to the field of AI.

The field of AI is multidisciplinary, involving philosophy, mathematics, neuroscience, economics, computer engineering, psychology, control theory, cybernetics, and linguistics. Each of these areas has contributed to the development of the field to its current level. As for its definition, AI is understood as taking the best possible action by a rational agent in a given situation. Thus, the construction of these agents is one of the purposes of AI (RUSSELL; NORVIG, 2009).

Machine learning, on the other hand, has a more narrow definition. Géron (2017) defines it as "the science (and art) of programming computers so that they can learn

from data". Mitchell (1997) states that "A computer program is said to learn from experience *E* with respect to some task *T* and some performance measure *P*, if its performance on *T*, as measured by *P*, improves with experience *E*". Experience *E*, which can be presented in the form of a dataset, is the fuel of machine learning. One would think that the more of this fuel, the better performance on task *T* the agent would have when performing *T*. However, this is not always the case, either due to the agent's limited capacity, quality of the data, or task difficulty.

This section introduces the basic concepts underlying the training of Artificial Neural Network (ANN)s. These concepts are essential when training Convolutional Neural Network (CNN). Such information about training and CNN is valid for explaining OD, as is the last subsection, which gives the most important evaluation metrics to this work development.

2.3.1 Basics of neural network training

The ML method, and hence DL, consists of four components: the data from which information is extracted, the model, the cost function that evaluates the model, and an algorithm that adjusts the trainable parameters. Deep learning methods also provides other learning elements from the input data and use techniques that enhance learning if compared to shallow methods.

Zhang et al. (2020) explains that these networks are called deep because there are many computational layers capable of addressing low-level features, which is why such a method benefits from a large volume of data, going deeper in complexity than other shallow ML methods, as complexity here is related to the numbers of trainable parameters. The elimination of the feature engineering step is the second advantage mentioned by the authors, which is the search for parameters that express the variables of interest that represent inputs related to the desired output. A deep network can learn which features are relevant while connecting them to the correct result. Still, this kind of solution is only possible because there is a lot of technological development behind it, like more powerful GPU, large volumes of data collected or likely to be collected, storage capacity, and efficient optimization functions (ZHANG et al., 2020).

Thus, DL is recommended for problems, like the one in this work, that involves searching for patterns in complex images with barely apparent relationships between the input (a fringe) and the output (the shape of the defect that originated it). Relating fringes to this output is a supervised learning problem, defined as the learning where the agent receives inputs with values associated with a single label and can map the path between this input and the output. In unsupervised learning, there is no such input-output relationship, and the agent is responsible for learning the connections between the observations (RUSSELL; NORVIG, 2009).

The simplest architecture of a DL method is the MLP and it is presented in figure

19. They are composed of several layers of basic units called neurons, hence also the nomenclature neural network.

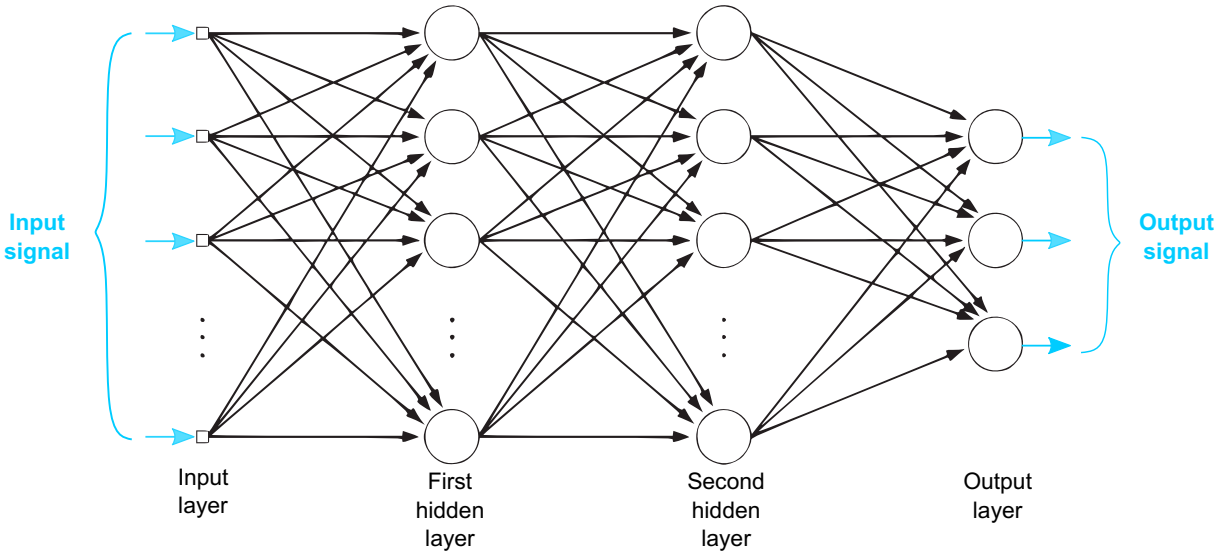


Figure 19 – Architectural graph of a MLP with two hidden layers (HAYKIN, 2009).

A neuron is the primary learning unit of an MLP, as presented in figure 20. A linear combiner takes multiplication weights w from an input signal x and sums them with a bias b, which is a constant added to help the weights to fit the data. This summation v passes through an activation function φ that acts to normalize the output, which is the input y to the next layer (HAYKIN, 2009).

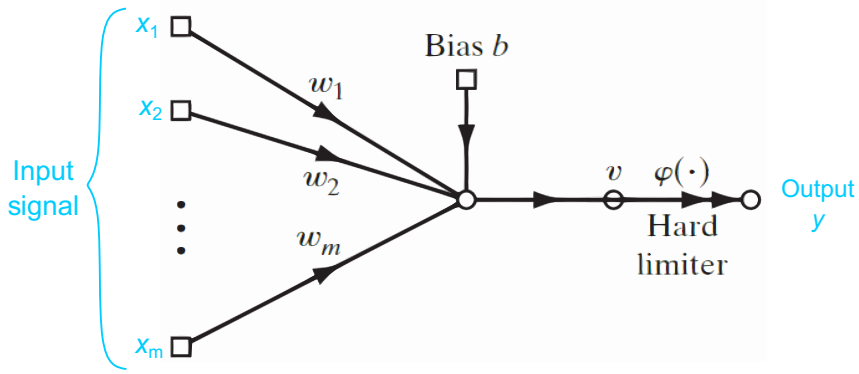


Figure 20 – Signal flow graph of the perceptron (HAYKIN, 2009).

The activation functions decide if and how the output of one neuron is propagated to the next layer while also introducing non-linearity to the neural network. The ReLu function, as shown in figure 21, and its variants are the most popular choice because of their simplicity of implementation (ZHANG et al., 2020). This function allows the propagation of only positive values, zeroing out the others, and follows equation $ReLu(x) = max(x, 0)$. As we will see later, in the error backpropagation method, it is necessary to use the derivative of the activation function. The derivative of a ReLu

function is well behaved, as seen in the right image of figure 21, except for input values equal to zero. Although rare, it is considered in these cases for practical purposes that the value of the derivative is zero.

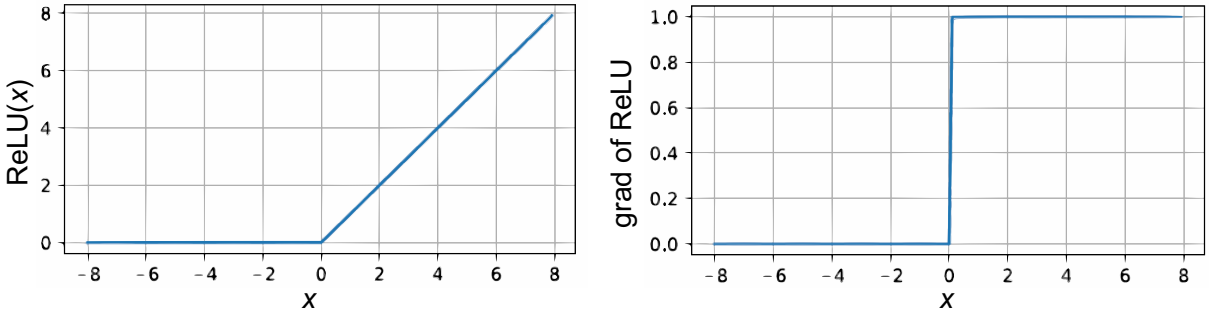


Figure 21 – Left: ReLu activation function and (right) its derivative (ZHANG et al., 2020).

There are two possibilities for the output of a classifier network, noting an interest in a categorical output that represents the best choice among all classes. The first of these is the sigmoid function. Figure 22 shows this function and its derivative. Following the equation $sigmoid(x) = 1/(1 + exp(-x))$, where x is any real number, this function propagates only positive values between 0 and 1, smoothing inputs smaller than minus four or larger than four. If the input is close to zero, the function behaves approximately linearly. When using a sigmoid, each class prediction has values between 0 and 1, and for multiclassification, the summation can be greater than one. For binary classifiers, it is necessary to apply a cut-off threshold to categorize the output. The advantage of this approach is that it is possible to perform multi-class classifications, that is, more than one answer for the same input.

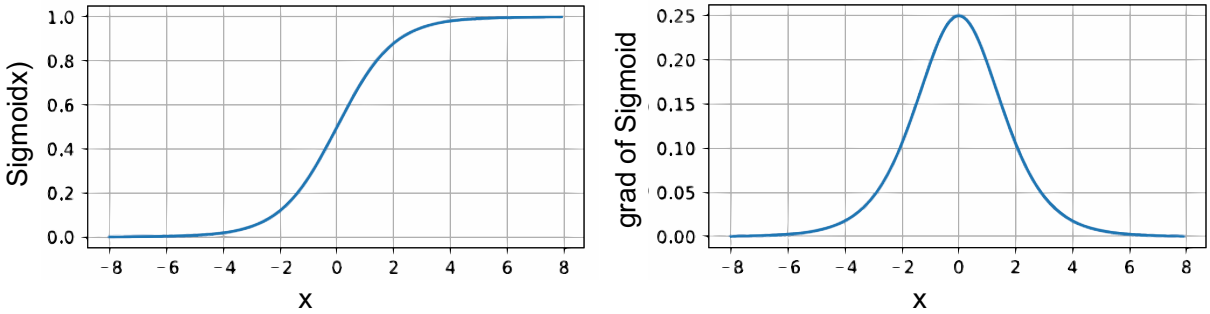


Figure 22 – Sigmoid activation function and its derivative (ZHANG et al., 2020).

The third activation function of interest is called Softmax, defined by equation 1, where $\mathbf{x}_n$ is a vector of size equals to the number of classes to be classified (STEVENS et al., 2020). For any inputs, the output is always a non-negative number. This function has a feature that the sigmoid function does not: Softmax can output each class's probabilities $\hat{p}_k$ between 0 and 1, and those probabilities sum to one. So, the result of

classification is that class with the highest probability. Equation 1 presents the formula for softmax calculaton.

$$\text{softmax}(x_1, ..., x_n) = (\hat{p}_1, ..., \hat{p}_1) = \left(\frac{e^{x_1}}{e^{x_1} + ... + e^{x_n}}, ..., \frac{e^{x_n}}{e^{x_1} + ... + e^{x_n}} \right) \tag{1}$$

With the output set, the model still needs to know how far the answer is from correctly classifying an arbitrary number of examples. It is the role of cost functions to express this value. The cross-entropy cost function, given by equation 2, is the most widely used for multi-classification tasks and has synergy with the Softmax function because it with probability-type inputs. This method is used to measure how well a set of estimated class probabilities matches the target classes (GÉRON, 2017). The lower the class probability calculated by the neural network against target one, the higher the penalty, as seen in figure 23. In cross-entropy equation, $y_k^{(i)}$ is the probability of the target class k to which instance i^{th} belongs, and can be one if the target belongs to the class in question, or zero if not; Θ is the parameter vector of the model; J is the function (objective) cost, which must be minimized; and $\hat{p}_k$ is the output of Softmax function, where

$$J(\mathbf{\Theta}) = -\frac{1}{m} \sum_{k=1}^{m} \sum_{k=1}^{K} y_k^{(i)} log\left(\hat{p}_k^{(i)}\right) \tag{2}$$

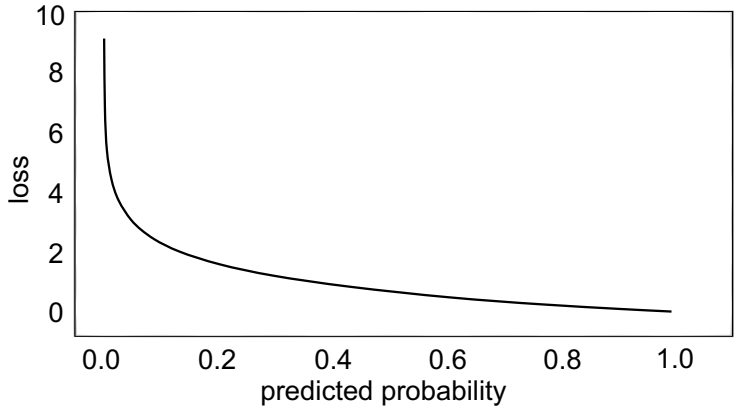


Figure 23 – Cross entropy loss when true label equals 1.

The derivative of the cost function, given by equation 3, is needed for the calculation performed by the optimizers during error backpropagation.

$$\nabla_{\theta^k} J(\mathbf{\Theta}) = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{p}_k^{(i)} - y_k^{(i)}\right) \boldsymbol{x^{(i)}} \tag{3}$$

With the result of the cost function, it is necessary to update the trainable elements (weights and biases) of the net accordingly. Initially, the synaptic weights and biases are initialized randomly with arbitrary distributions before training begins. Updating the weights and biases of a neural net happens in two steps: propagation and error backpropagation (RUMELHART et al., 1988; HAYKIN, 2009):

Propagation the vector with one or more inputs is presented to the input nodes. The input values' effect is propagated throughout the neural network, layer by layer, causing an output vector. During this step, all synaptic weights are fixed.

Error Retropropagation (backpropagation) all synaptic weights (trainable elements) are adjusted according to the value of the cost function. This signal is backpropagated by the net in the direction of the weights and detectors, which, using partial derivatives and chain rule, can identify the error portion of each previous connection.

The search of the minimum error values in a multidimensional error space occurs during error backpropagation. There are several solutions in this convex space, but only one is optimal. The optimizing algorithm is expected to ignore local minima and find the global minimum. There is a vast literature on optimizers and several assumptions for them to work. Learning rate is an essential hyperparameter of an optimizer and defines the step size or how much the weights are adjusted concerning the loss gradient. A hyperparameter is a parameter that requires a manual set and is often prone to adjust. More information about an optimizer theory, its developments, and implementations can be found in Nocedal and Wright (2006) and Zhang et al. (2020).

Adaptive Moment Estimation (ADAM) optimizer was the choice for this work (KINGMA; BA, 2015). According to the authors, this method is easy to implement, computationally efficient, requires little memory, is invariant to the rescaling of the gradients, is suitable for problems with large amounts of data and parameters, is appropriate for nonstationary objectives and issues with sparse and noisy gradients, and has easily interpreted hyperparameters with little need for adjustment.

This update of the weights takes place during a period called step. During a step, the neural network is presented with one observation or a batch of an arbitrary number of observations, namely batch size. The unique input to a machine learning method is called an observation. This observation consists of various features from which the model performs training and prediction. In images, the number of features is the product of the width, height, and channels. An image will not always have the same width and height, i.e., its dimensionality may change. This situation occurs in this work, where each defect has its own dimension. It is common for classifiers and detectors to tailor their inputs to a single size to deal with these different dimensions.

The neural network is only updated after all observations of a batch have been presented to it. Commonly, the number of steps is the ratio of the total number of observations and the batch size. For example, a dataset with one hundred observations and batch size equals ten would take ten iterations to be fully presented to the neural network. The number of steps it takes for the neurak network to receive all the examples is called an epoch. Also, the batch size is a training hyperparameter and helps to control generalization, a concept that will be explained shortly. However, large batch sizes demand more processing power since more samples need to be stored in memory and

propagated through the neural network.

It is expected that the dataset will consist of many observations and that these observations will be representative of the distribution of each class. However, it is not trivial to define a priori the number of observations needed for a network to perform satisfactorily. The amount of observations depends, for example, on the task difficulty, the method, the input complexity, and the cost of obtaining new samples. Task difficulty is associated with the problem itself. For example, it would take fewer examples to classify whether an object is a flower than to which species it belongs. The methods are related to the number of trainable parameters of the chosen model. Deep models have more trainable parameters, so they need more examples to output a coherent result. As for the observations, they can come with characteristics such as occlusion, making it necessary to have a larger number of them to express the totality of the object of interest. Finally, the cost of obtaining new observations can be impeding, either because of the price of doing new tests, the time, or the rarity of the events that create these examples.

These are the basic steps in training any neural network model. However, its architecture, represented by the number of neurons and hidden layers, is arbitrary for unsolved problems or fixed in models proven effective. Its architecture defines the complexity of the model, which is the starting point for its choice. The model's capability can be estimated before and confirmed after training, as it is also related to the dataset. A complex model for a given problem has a low training error and a high variance. This scenario means that the model has memorized the examples, reducing its generalization. For instance, in a polynomial curve fitting, what dictates the model's complexity is the polynomial order, which is directly related to the number of free parameters.

It is possible that performance does not reach the expected level even with a sufficient amount of data and a model capable of processing it. It is first necessary to define a metric to determine when a method is to be considered valuable. This metric will be accuracy at this moment, but section 2.3.4 will present all metrics used in this work. The best answer to a problem comes from an oracle that knows the actual probability distribution that originated the observations and makes predictions from it. This prediction error, 1 – *Accuracy*, is called the Bayes Error and is irreducible (GOODFELLOW et al., 2016). The second smallest error is called human-level performance. This error comes from the accuracy achieved in the performance of a task by one or more experts. While the Bayes Error is challenging to quantify, the human-level performance comes from feasible experiments. However, it is known that deep neural networks can outperform human-level performance in several tasks, such as recommender systems, logistics, speech recognition, and even some image classification.

The generalization concept is useful to check if a trained model can improve and reach, or even surpass, human-level performance. Generalization is understood

as the ability of an intelligent method to provide consistent results when given inputs never seen during training. Two situations impair the ability to generalize: overfitting and underfitting. These are two concepts that are related to the bias-variance trade-off. This trade-off is related to simultaneous reduce of the variance and the bias although they are inversely proportional. Figure 24 shows three linear regressions with different fitting characteristics. In this case, random data were generated by a quadratic distribution. The first image in the figure shows the fitting of a linear function for this distribution, which results in underfitting. The appropriate capacity is, as expected, found by a quadratic function; overfitting is obtained by fitting a polynomial of degree nine (GOODFELLOW et al., 2016).

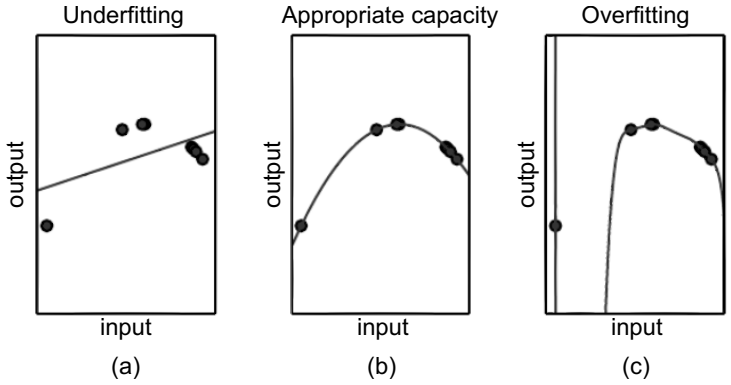


Figure 24 – Fittings coming from three different degrees of polynomial (GOODFELLOW et al., 2016).

It is common to split the data set into three parts to estimate and even avoid lack of generalization. One of them is entirely dedicated to training, the second is for validation, and the third is for testing. The difference between the resulting model error after training and the human-level performance is called avoidable bias. This implies that it is possible to improve the model with some strategies, such as increasing the model's capacity, training for longer, changing the optimization algorithm, or adjusting the neural network parameters. The validation slit is the part that approximates the actual use of the model and should follow the same distribution as the test split. Applying the trained model to the validation set yields the validation error, and the difference between this error and the training error is called the variance. A high variance implies more data to be acquired, regularization techniques, or changing neural network parameters.

Allowing the model to train more than necessary, i.e., let its complexity increases indefinitely, is another overfitting source. Early stopping is a strategy to force it to stop at the best possible point of generalization. In this strategy, at each epoch, the trained model is evaluated on the validation set and a counter is started if it has not improved. This counter waits for an arbitrary number of epochs and it stops the training and returns the best model obtained so far if the validation does not improve.

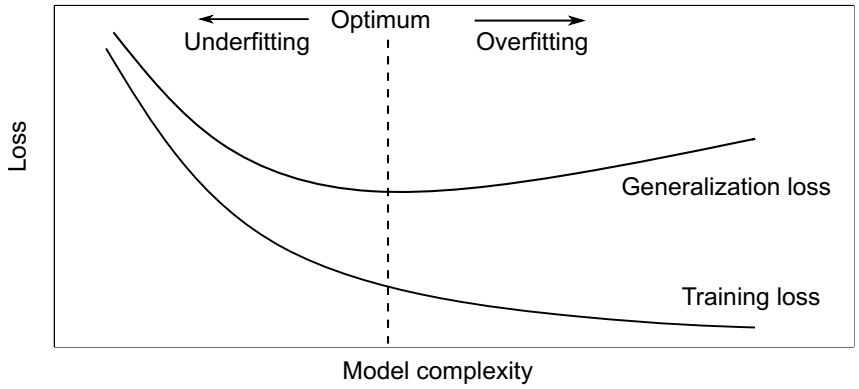


Figure 25 – Influences of model complexity on underfitting and overfitting. Adapted from Zhang et al. (2020).

2.3.2 Convolutional neural networks

Girshick et al. (2014) have already demonstrated the superiority of CNN over methods based on parameter extraction. Feature extraction from images in a CNN occurs by convolution blocks composed of filters. The difference from traditional filters is that they are trainable, i.e., each filter learns the appropriate kernel values during training. A convolution operation is a dot product (element-wise multiplication) between the filter and the input image patch, as shown in figure 26, that can benefits from GPU processing. The result is summed to result in a single value. After doing this, we displace the filter laterally over some pixel quantity position and do it again. This amount of displacement is called stride, and it is a parameter prone to adjusting. All convolution processes are performed independently on each channel of the image, thus each channel can provide different features for the learning task.

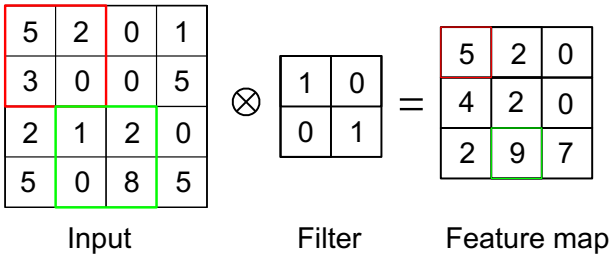


Figure 26 – Example of convolution of stride equals one.

Equation 4 represents a convolution operation in a simplified way, where the input data is convolved with a set k of learnable kernels W plus biases b, generating a new feature map X_k. This new feature map is the input to an activation function σ (JIANG et al., 2019).

$$\mathbf{X_k^{l+1}} = \sigma(W_k^l * \mathbf{X}^l + b_k^l) \tag{4}$$

Another important operation performed on a CNN is pooling. Pooling is similar to convolution, but there is no trainable filter, just a kernel responsible for traversing

the feature map obtained by convolution and summarize the grid *nxn* to a single value. This value can be the maximum value (Max pooling), average pooling, minimum (min. pooling), among others. In general, this operation makes the representation invariant to small translations, progressively reduces the size of intermediate interpretations, the number of parameters, and the computational effort for a CNN (GOODFELLOW et al., 2016; JIANG et al., 2019). In specific, Max pooling has the advantage of privileging maximal activations over sparse information (CHOLLET, 2017). Figure 27 shows an example of max pooling calculation with stride equals one.

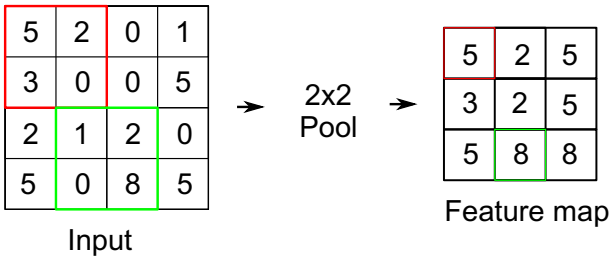


Figure 27 – Example of max pooling with stride equals one.

Convolutional layers not only serve to extract features but also change the dimension of the images. As features are extracted and their dimension changed, the features become more complex or high-level. Thus deeper networks extract more complex information from an input. After the last convolution layer, there is the flattening process, which is nothing more than the vectorization of the resulting feature maps to be fed to an MLP. This MLP performs the classification, as seen earlier. During CNN training, the error signal is propagated through the convolutional layers, and the updating of the filters occurs just like a neuron. Figure 28 shows one of the first CNN architectures developed (LECUN et al., 1998).

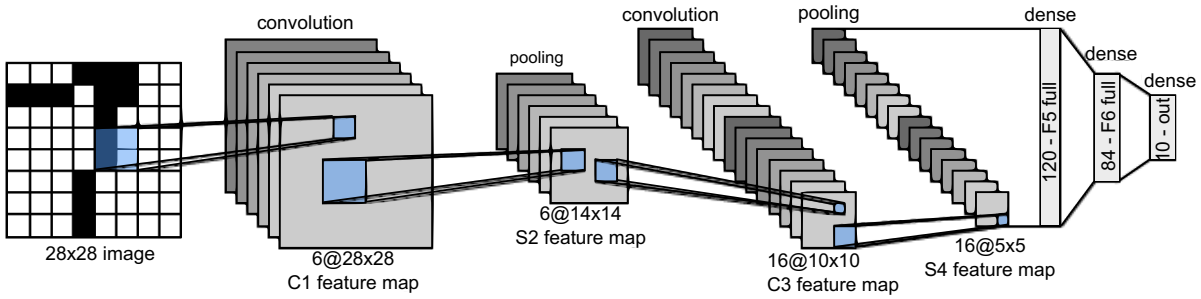


Figure 28 – Data flow in LeNet (LECUN et al., 1998). The input is a handwritten digit and the output is a probability over ten possible outcomes (ZHANG et al., 2020).

It is expected that the more convolutional layers are added to the neural network, the more complex features are extracted. However, there is a depth limit, just as hidden layers in an MLP. From this limit, the performance starts to get worse because of a

phenomenon called vanishing gradient, which limited the performance improvement of models like AlexNet (KRIZHEVSKY et al., 2012), Visual Geometry Group - VGG (SIMONYAN; ZISSERMAN, 2015), and GoogleLeNet (Inception) (SZEGEDY et al., 2014) that relied on the depth and a large number of parameters. In this phenomenon, the back-propagated signal during training gradually and exponentially disappears as it travels from the end to the beginning of the neural network, causing the initial layers to have little or no adjustment in weights (SUSSILLO; ABBOTT, 2015; UNTERSUCHUNGEN. . . , n.d.; BENGIO et al., 1994). Goodfellow et al. (2016) further mention that "Vanishing gradients make it difficult to know which direction the parameters should move to improve the cost function." Unlike overfitting, the vanishing gradient is already visualized in the model's training performance, where high errors are noticed.

ResNet introduced the residual block concept to CNNs to minimize the vanishing gradient phenomenon and enable deeper networks. For such a task, the authors proposed that each convolutional block should have its own identity as an element, with the hypothesis that it would be easier to optimize the residual map than to optimize the original map. This residual block, shown in figure 29, is characterized by a connection that feeds the following layer's input with its own identity. In this block, batch norm stands for batch normalization, conv is convolution, and ReLu is Rectified Layer Unit.

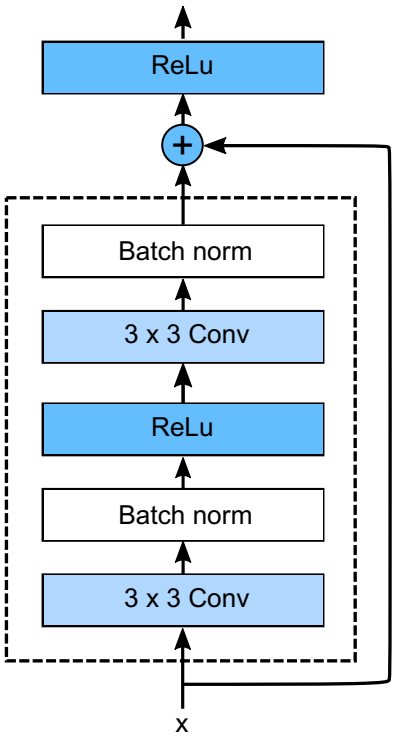


Figure 29 – ResNet skip block. Adapted from Zhang et al. (2020).

Thus, the network must learn only the difference between the identity and the block's output, the residuals, hence the name Residual Neural Network. The residual block has two convolutional layers that preserve the same dimension as the input layer since, at the end of this residual block, the sum of the identity occurs. However, this

dimension is reduced at some points, and the same should happen with the residual link. This block also occurs the batch normalization, which aims to accelerate the convergence of deep networks (IOFFE; SZEGEDY, 2015). First, the inputs (in this batch) are normalized by subtracting their mean and dividing by their standard deviation, obtained statistically from this mini-batch. Then the scale coefficient and the scale offset are applied with updated values as the training occurs with the whole dataset (ZHANG et al., 2020).

A ResNet starts with a 7x7 convolution layer, followed by batch normalization and a 3x3 max pooling. After this max pooling starts the residual blocks, and after numerous blocks there is a global average pooling ending in a fully connected layer. Here, the global average pooling replaces the flattening because it averages each channel's feature map to build the vector that feeds the fully connected network. The number of layers with trainable parameters defines the capacity of the network. A Resnet50, for example, has 50 trainable layers: the first convolutional layer, fourty-eight convolutional layers, and finally the fully connected layer.There are other possibilities for ResNets with other layers, such as 18, 34, 101. These different numbers of trainable layers are directly related to the network's feature extraction capability. Shallow networks like ResNet-18 cannot extract more complex patterns (HE et al., 2016).

2.3.3 Object recognition

Classification networks such as ResNet are trained to classify a single class instance per input. There are situations where there is interest in multiple classes in the same image and their positions. Object recognition (or object detection) is a set of computer vision and image processing techniques that can detect objects in an image and classify them (RUSSAKOVSKY et al., 2015; JIAO et al., 2019).

An OD is composed of a backbone, neck, and head. The backbone is responsible for extracting features from the input image. State-of-the-art detectors make extensive use and benefit from deep learning algorithms, such as ResNet, for the backbone function. In the neck are additional auxiliary methods and aggregator blocks, such as the Feature Pyramid Network (FPN) (LIN, T. et al., 2017), which will be commented on again later. The head is of two types according to the pipeline they use: one-stage or two-stage. Single-stage detectors simultaneously extract and classify the objects of interest. Examples of single-stage detectors are Single-Shot Detectors (LIU, W. et al., 2015), YOLO (REDMON; FARHADI, 2018; BOCHKOVSKIY et al., 2020; REDMON et al., 2016), and RetinaNet (LIN, T. et al., 2020). Two-stage detectors first generate large numbers of candidate detections and then classify them into the categories of interest. As examples of two-stage detectors, we have Faster Region-Convolutional Neural Networks (Faster R-CNN) (REN et al., 2015), and Region Based Fully Convolutional Networks (R-FCN) (DAI et al., 2016). Two-stage detectors are slower for prediction

tasks than those with one stage and until recently had state-of-the-art accuracies. The accuracy of one-stage methods was lower but still considered comparable, making these methods more balanced (JIANG et al., 2019; BOCHKOVSKIY et al., 2020). However, a recently developed detector called EfficientDet had superior results, as seen in Figure 30 where the *Y*-axis is the average precision on the Common Objects in Context (COCO) dataset, and the *X*-axis is the number of operations (floating operations per second, or FLOPS), in this case referring to the number of operations required to run a neural network model. Usually a larger model requires a higher number of operations. More on evaluation metric on section 2.3.4.

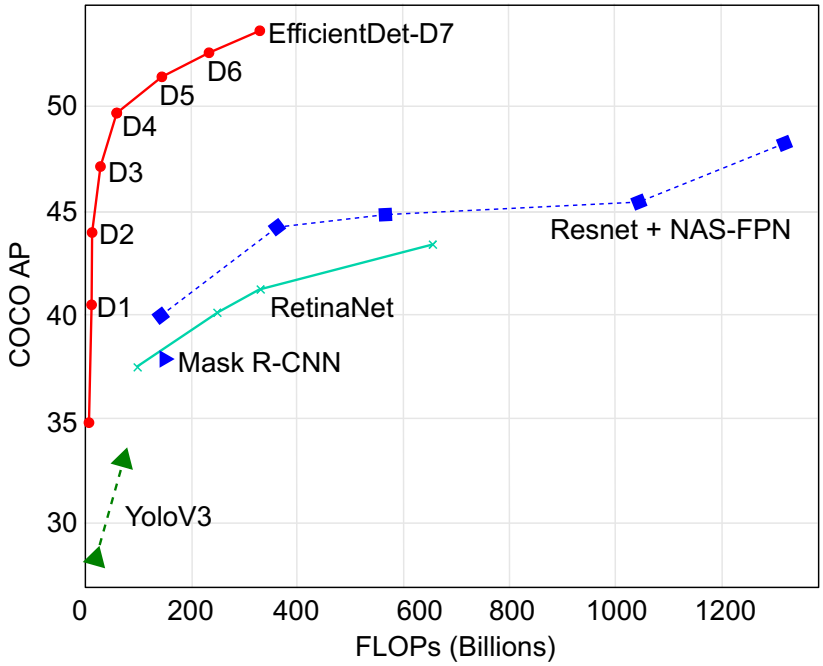


Figure 30 – Comparison of different detectors average precision of one and two stages on COCO dataset. Adapted from Tan and Le (2020)

.

At a certain point, it was necessary to define which detector would be used in this work. Since the field of detection algorithms is in constant update, a state-of-the-art detector was hardly a possible choice. At the time of this choice, the established models were SSD, YoloV3, RetinaNet for single-stage detectors, and FasterRCNN for two-stage detectors, as shown in 30. Nowadays, other possibilities have brought more accuracy (EfficientDet) and speed in inference (YoloV4). Given the exploratory proposal of this work, so the detector chosen was RetinaNet using a ResNet50 as the backbone, which has an average performance regarding AP and FLOPs. In preliminary tests, more complex backbones brought problems of adjustment and even divergence during training.

Retinanet is a one-stage detector with focal loss as the classification loss function. RetinaNet inherits previous one-stage detectors' speed while greatly overcomes the disadvantage of one-stage detectors' difficulty to train unbalanced positive and neg-

ative examples. Its architecture is demonstrated in simplified form in figure 31. In this architecture suggested by the authors, the backbone is a ResNet, the neck is an FPN, and the head is a dense prediction based on anchors. The goal of FPN is to perform integration of parameters extracted by the convolutional network, combining weak and strong semantics at different levels. Feature Pyramid Network produces new cumulative feature maps that benefit from the extraction hierarchy present in CNN, where each level of the pyramid can be used to detect objects at a different scale. Each of these feature maps from the pyramid levels is input to other subnets responsible for classification and detections.

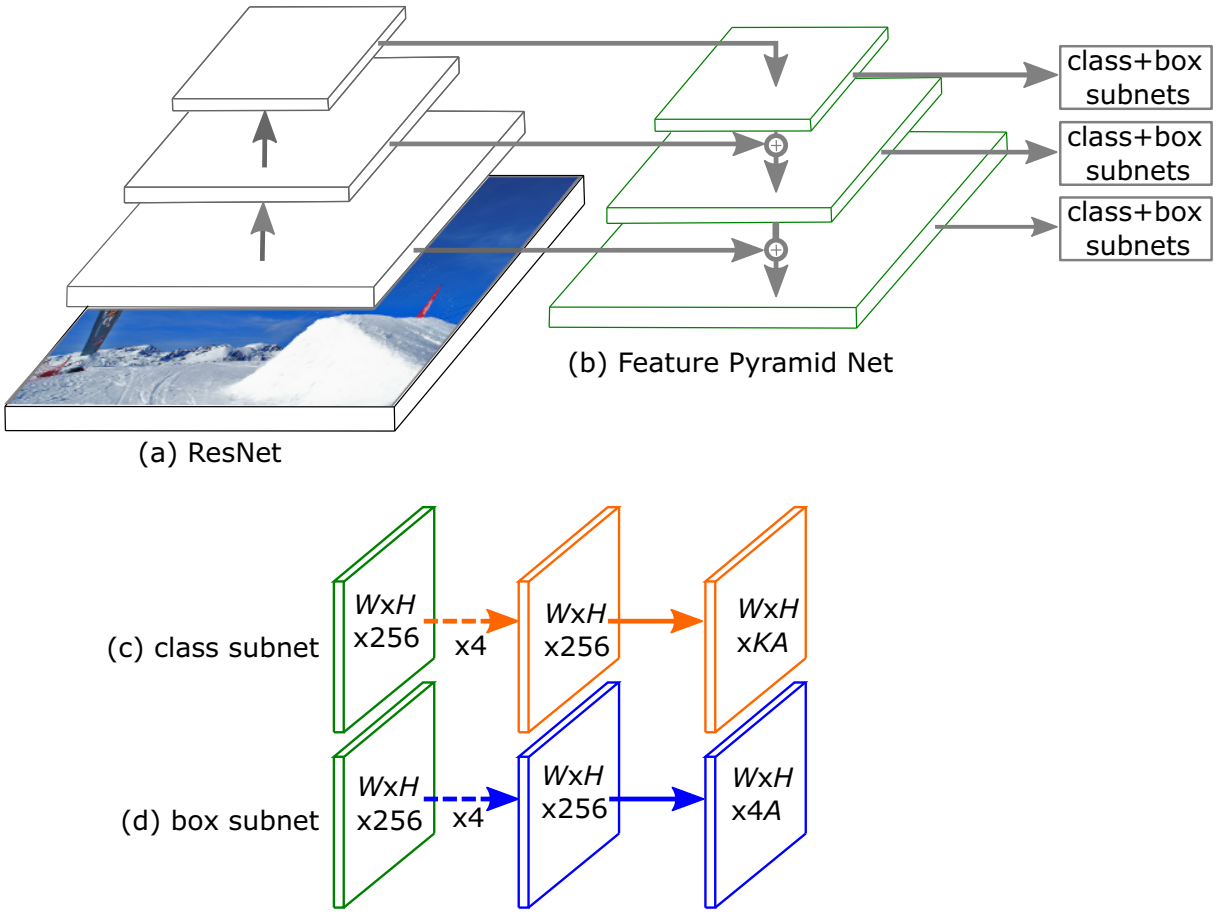


Figure 31 – The one stage RetinaNet architecture with (a) ResNet as backbone and (b) FPN as neck. Subnets (c) and (d) are responsible for classifying the classes and the anchors respectively (LIN, T.i et al., 2017).

Each subnet is composed of four layers of convolutions. The output of the classification subnet uses the focal loss cost function instead of cross-entropy, figure 32, which is where the Retinanet contribution lies. According to the T. Lin et al. (2020), the output layer is a sigmoid rather than a softmax, which leads to greater numerical stability. Cross entropy, the blue curve, punishes incorrect examples more severely, causing class unbalancing when too many easy examples are provided to the network. The focal loss adds a modulation component $(1 - p_t)^{\gamma}$, which modulates the curvature

of the cross-entropy function to give less focus to the harder weights by pushing the curve downward. In this equation p_t is the input probability and the constant γ is called focusing parameters and serves to change the shape of the Focal Loss curve, and has the best value suggested by the authors as two (LIN, T.i et al., 2017). The threshold that tells if there is a class or not in a given detection (through the sigmoid function) is also called confidence score, an important term discussed in section 2.3.4.

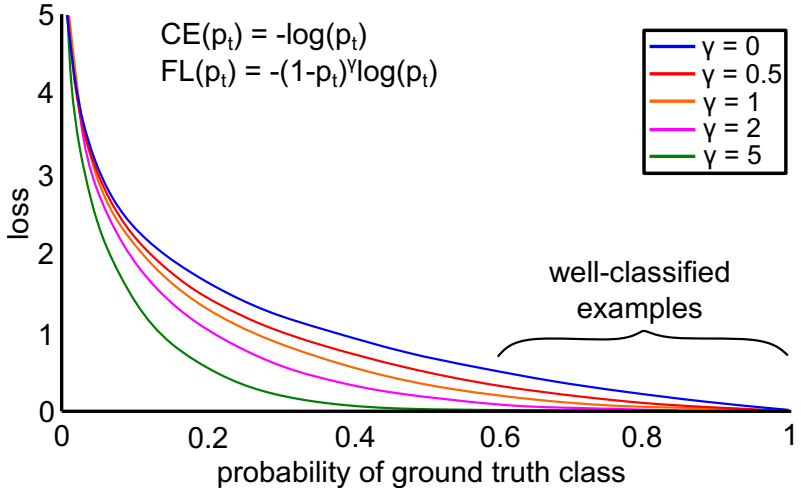


Figure 32 – Comparison of focal loss (FL) with different focusing parameters γ and cross entropy (CE) (LIN, T.i et al., 2017)

.

The subnetwork related to box regression relies on anchor and the concept of Intersection Over Union (IoU). Anchor boxes were introduced in (REN et al., 2015) and have become a standard element even in single-stage detectors. They are randomly created geometric elements with arbitrary sizes. The smallest anchor box created is 32x32, which means that objects up to this dimension can be detected. For smaller objects, it is necessary to adjust the anchor boxes and retrain the model. Its Intersection over Union must be at least 0.5 for an anchor to be associated with ground truth. The IoU, or Jaccard Index, quantifies overlaps between two areas (JACCARD, 1901). Figure 33 brings the visual representation of IoU, which is an overlap of the images divided by their union.

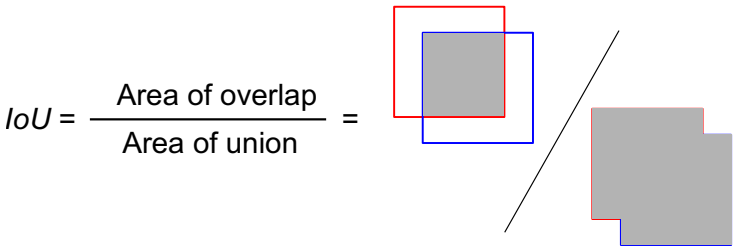


Figure 33 – Intersection over Union visual representation.

Box regression is responsible for calculating the ground truth's degree of belongingness to the best anchor. This subnetwork is independent of the subclass that

performs the classification, which is another advantage over the two-stage methods. The weights of this network are updated based on the result of the cost function *smooth L1 loss* (GIRSHICK, 2015).

There can be more than one bounding box related to the same ground truth. One way to get around this problem is through Non-Maximum Suppression (NEUBECK; VAN GOOL, 2006; ZHANG et al., 2020). This method iteratively chooses the anchor box with the highest confidence and removes any other overlapping anchors similar to it, identified by a certain *IoU* threshold.

The difference between object detectors with other ML methods lies in the way the examples are presented to the net. An input to a detector is an image combined with a file containing coordinates called bounding boxes, which tell the detector where the class in question is (ZHANG et al., 2020). Creating such a dataset is a challenging task that is usually performed manually, given its complexity and subjective nature.

As for training, a detector follows the principles mentioned for the other machine learning methods. For example, it is necessary to define the number of epochs, batch size (the number of steps is calculated) automatically, and whether it is required to use data augmentation. Data augmentation is a way to increase the training set's variability by performing operations like shear, zoom, and vertex flip on the input images (GOODFELLOW et al., 2016). However, augmenting the data is not always a good idea. For example, applying shear and zoom to fringes could inhibit rather than help training by mischaracterizing their features. This is why data augmentation methods were not used on this work dataset development.

2.3.4 Evaluation metrics

Machine learning methods make silent mistakes. It is not trivial to identify the cause of any poor performance, but it is possible to infer it from evaluation metrics. It is also by the metrics that the comparison of different methods and different models is made.

The basic metrics for classification problems are True Positive (TP) when the positive inference of a method is correct; True Negative (TN), when the negative inference of a method is correct; False Positive (FP) when the inference is positive, but the true result is negative; and False Negative (FN), when the inference is negative, but the true result is positive. From these base metrics come other metrics, such as Precision and Recall. Precision, defined by $Precision = TP/(TP + FP)$, represents the proportion of positive classifications that were actually correct. The Recall is defined by $Recall = TP/(TP + FN)$ and defines the proportion of positives that the model correctly identified. A practical example would be rare flaw machine detections. The precision would be $Precision = TP/Total\ positive\ results$ and the $Recal == TP/Total\ flaw\ cases$. Note that the TN is not inserted into precision and recall because there are often too many TN in a

classification and detection object, which would cause an unbalance. This is also what happens when using accuracy, defined by $Accuracy = (TP + TN)/(TP + TN + FP + FN)$ (PADILLA et al., 2020).

There is a balance between precision and recall. A binary classifier might predict that all examples are of a certain class, and it would certainly hit all those that were indeed of that class. This inference would generate a Recall of 100% but a low Precision value. It is necessary to analyze the problem and define the metric of interest. However, through the F-score, it is possible to join these two metrics in the form of a harmonic mean, providing a complete way to analyze a classifier. The F-Score score is defined by $F_1 = 2 * (Precision * Recall)/(Precision + Recall)$. Figure 34 presents two hypothetical vectors and all the steps until the resulting F-score.

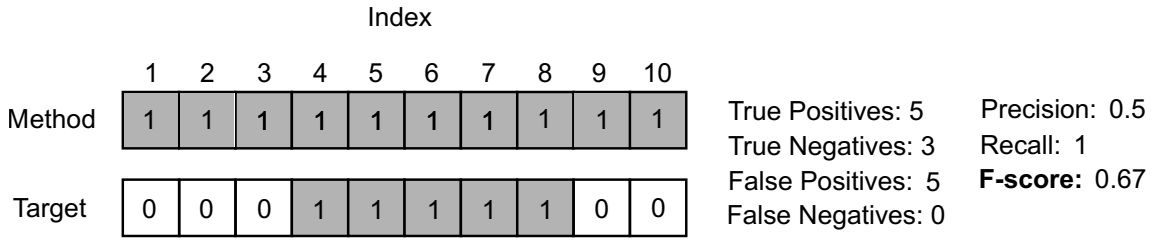


Figure 34 – F-score calculation from two hypothetical vectors.

Competitions were partly responsible for the improvement of detection algorithms. Each competition used a different metric to evaluate the results and select the best model. RetinaNet uses the Precision x Recall curve and AP, according to the Pascal VOC Challenge rule (LIN, T. et al., 2020; EVERINGHAM et al., n.d.). The Precision versus Recall curve is constructed with different confidence scores and acquires a serrated shape. There is an interpolation before averaging the Precision over the Recall to simplify the curve. Initially, eleven points were used for this interpolation, but nowadays, all points on the curve are used to calculate the area under the curve, according to equations 5 and 6. In both equations, $\rho(r)$ is the measured precision at recall r. It is not the Precision of each point used during AP calculation, but rather an interpolation between points, providing the maximum Precision whose recall values are greater than r (PADILLA et al., 2020).

$$AP = \sum_{n=0} (r_{n+1} - r_n) \rho_{interp}(r_{n+1}) \tag{5}$$

$$\rho_{interp} = \max_{\bar{r}:\bar{r} \geq r_{n+1}} \rho(\bar{r}) \tag{6}$$

Figure 35 shows a hypothetical graph of AP (blue lines), the interpolations (dotted red lines), and the resulting area of interpolation (filled in red). The resulting area defines the AP and comes from equation 5. The closer to one, the better.

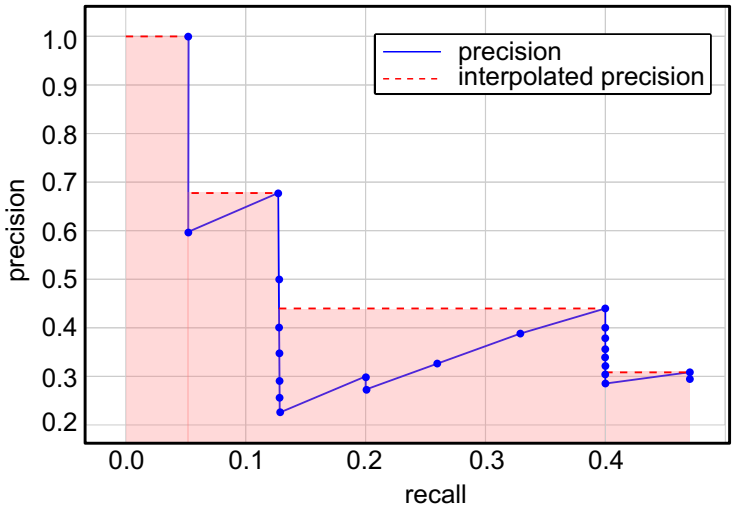


Figure 35 – Hypothetical AP curve. In blue the precision curve, in doted red the interpolated precision, and in red the area under the curve. Adapted from Padilla et al. (2020).

Each class has its own AP. The mAP is used to unify those values. This metric is the arithmetic mean of the average precision values for an information retrieval system over a set of query topics (BEITZEL et al., 2009). It can be expressed as equation 7. So, the average of AP defines the overall model performance mAP of the detector, where

$$\text{mAP} = \frac{1}{n} \sum_{n} AP_n \tag{7}$$

2.4 APPLICATION OF MACHINE LEARNING TO SHEAROGRAPHY IMAGE ANALYSIS

As mentioned during the introduction, there are few works that seek to improve shearography tests through the use of artificial intelligence, and two of the mentioned works are also related to images coming from pipelines. Fröhlich et al. (2018) brought a first attempt of the use of CNN in the binary classification of defect presence in pipelines repaired with GFRP, which can be seen in figure 36.

Figure 36 – A pipeline sample with its three composite repairs (FRÖHLICH et al., 2018).

The biggest challenge encountered during this work was the size of the dataset, which was composed of only 296 images separated arbitrarily between 168 images with major defects and 128 images with minor or no defects. Figure 37 shows examples of these defects and pipelines without any defects.

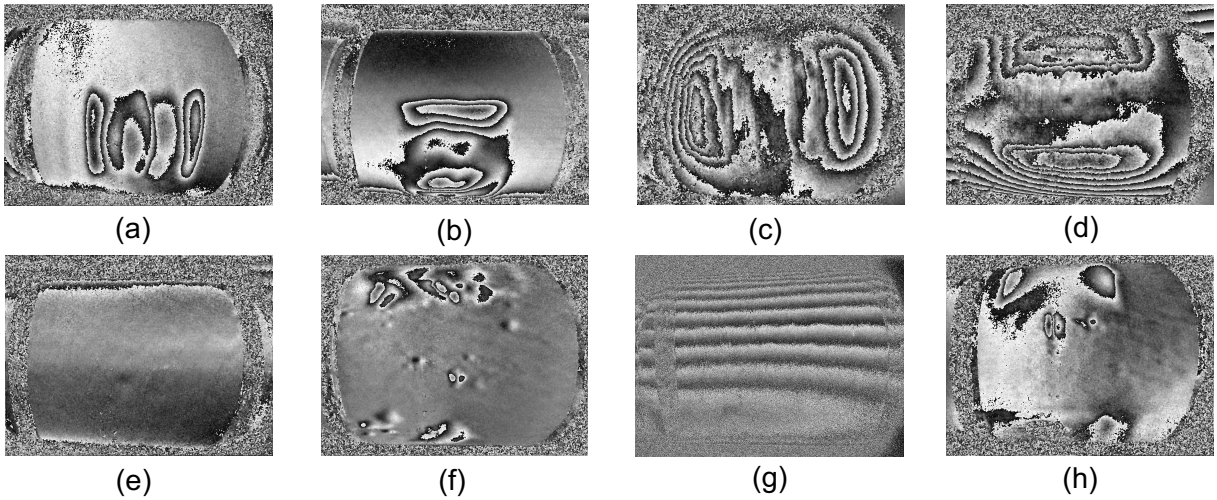


Figure 37 – Shearography difference phase maps with the following situations: a) lack of adhesion detected with horizontal shear; b) lack of adhesion detected with vertical shear; c) a wall thickness loss area detected with horizontal shear; d) a wall thickness loss area detected with vertical shear; e-h) no defect or minor defect detected. Adapted from Fröhlich et al. (2018).

The best performing CNN in this work reached 79% in the validation set, which consisted of 20% of the total images. The overfitting caused by the low complexity of the neural network was solved with data augmentation and dropout layers. On the other hand, dropout is responsible for randomly neglect some neurons during the training step to avoid too much co-adapting (SRIVASTAVA et al., 2014). Fröhlich et al. (2018) conclude by suggesting the possibility of multiclassification and parameter engineering of fringe patterns to ensure the best learning from machine learning methods.

The automation of shearography testing, specifically its use in pipeline inspection, has taken a step forward with the work of Ye et al. (2019). These authors set up a measurement system composed of a rolling mechanism, the shearography assembly mounted on a translation mechanism, and a control unit. This mechanism is represented in figure 38. The part under analysis is rotated automatically. In each of these partial rotations, a segment of the piece is presented to the shearography system for inspection.

Measurement system presented by Ye et al. (2019) aims to find adhesion faults between thermal protection coats and steel pipe by using OD, more specifically a Faster R-CNN. The measurement results are presented after stitching the images, where each segment can be traced and related to its location in the pipe. Figure 39 shows the result of a measurement, with the defects marked.

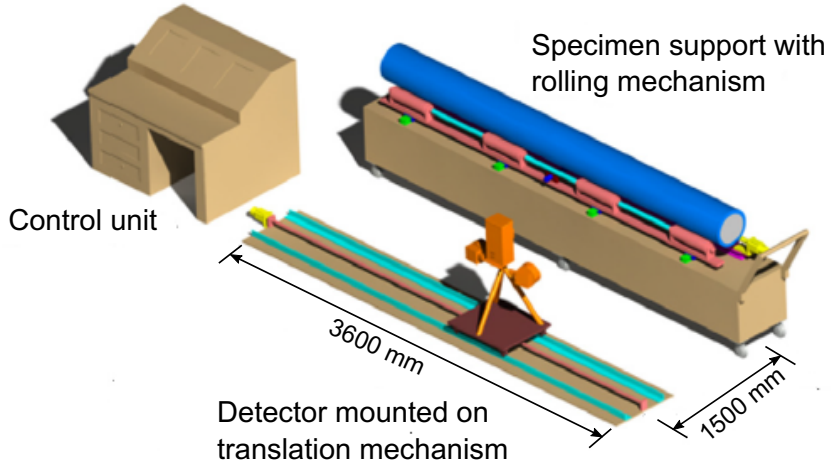


Figure 38 – Scheme of the test system (YE et al., 2019).

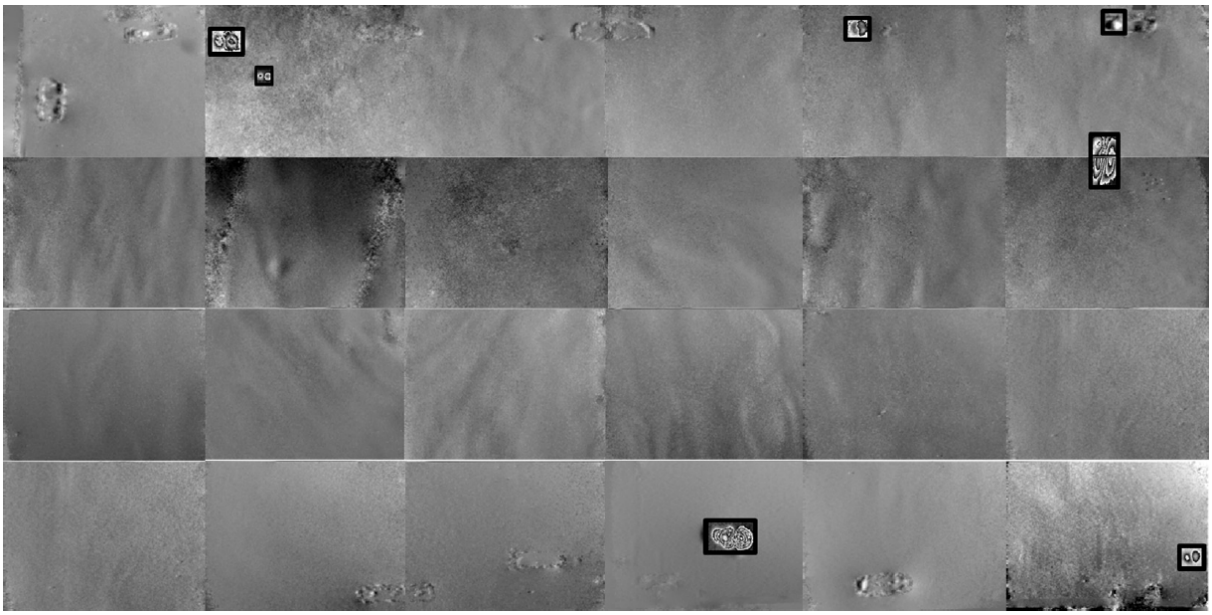


Figure 39 – The defect distribution over the entire cylindrical surface detected by the system (YE et al., 2019).

Finally, the authors pointed out the low hit rate for minor defects, indicating a problem that occurs due to the lack of small examples in the dataset. Also, minor defects are challenging to detect by the shearography system itself.

3 DEVELOPMENT

This chapter presents the development carried out in this work. The first step to accomplish the fringe pattern classification objective was producing the artificial defects that emulate lack of adhesion with different shapes and dimensions. Metallic plates were manufactured, and then they were laminated with GFRP with artificial defects of different shapes and areas. The data set was generated from the acquisition of shearography images from these plates. Annotations for the object detector were initially manually done, i.e., the fringe patterns representing the defects were delimited by manually selecting the corners of a rectangle around them. In this stage, there was a subjective criterion to consider if the fringe pattern coming from a particular defect was useful or not. The concept of the usefulness of a fringe pattern is also explained in this chapter. Because of this subjectivity and because it is a laborious job, a method was developed to help identify this usefulness, namely AUTO-UFP. Section 3.1 brings the details of the dataset development, from the design of cushions that emulate the adhesion defect to the acquisition of images of the laminated plates through shearography images. Section 3.2 demonstrates how AUTO-UFP yields a decision regarding fringe pattern quality.

3.1 DEVELOPMENT OF ARTIFICIAL DEFECT AND IMAGE ACQUISITION

3.1.1 Cushion production and arrangement

As presented in section 2, lack of adhesion, just adhesion defects are related to a lack of composite‘s first layer adhesion onto another substrate. A mix of interlayer insertion and pull-out strategy was considered here, resulting in a significant difficulty to keep the air in the desired shape and position on the plate. The adequate approach was to artificially emulate this defect with two PTFE layers attached at their edges by an adhesive tape. Air was manually inserted into this cushion using a syringe. Polytetrafluoroethylene was the cushion material mainly because it has mechanical properties comparable to the composite matrix in use.

Several issues had permeated the making of the air cushions. The two layers of PTFE have to be pressed on the edges so that no air escaped through the PTFE-tape interface, but there was a risk in this procedure of losing the defect correct shape and area. A simple inspection prevented such a problem: it was possible to check if the inner edges had kept the desired shape by positioning the specimen in front of a light source. The next step was to slowly insert the needle through the rubberized adhesive tape region between the PTFE. This insertion was the most problematic step. Sometimes the tip of the needle ruptured a PTFE layer, resulting in the material being unusable. Moreover, it was impossible to keep track of the air volume inside cushions because of the air insertion method, but their thicknesses were around 2.5 mm. These steps are

summarized in figure 40.

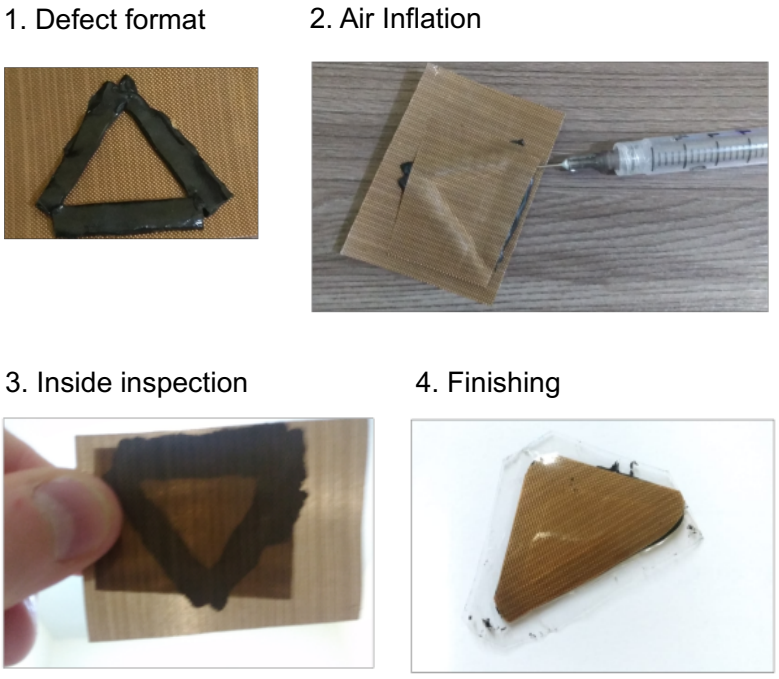


Figure 40 – Production steps for a Teflon air cushion production.

Defects were designed to have three shapes and three surface areas: Q, T, and SC. Square defects served as the basis in terms of area for the development of the other shapes. Figure 41 shows these defects areas and their relation among shapes.

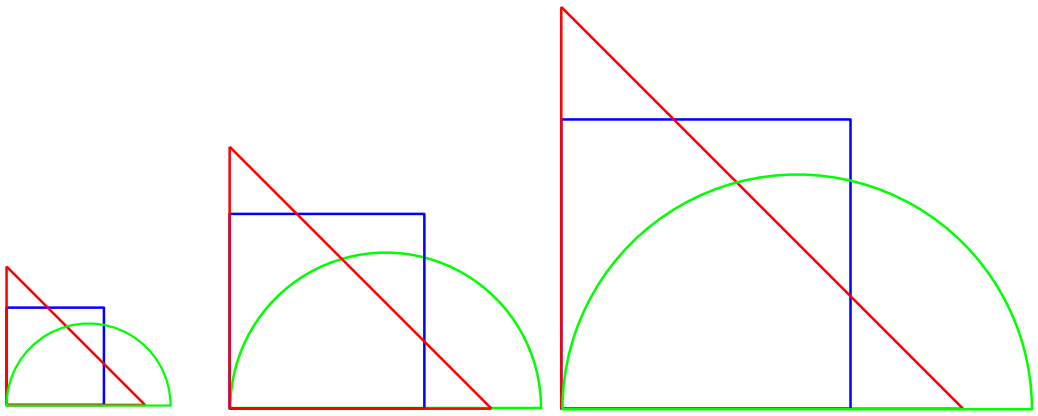


Figure 41 – Defect areas.

The metal plate onto which the defects were added were 3 mm thick AISI 1020 carbon steel plate with a work-space area of 200 mm by 200 mm and an L-shaped base with slits for fixation during testing. The cushions were placed on this plate work-space and, on top of them, the fiberglass composite laminates were packed layer by layer a total of eight times. This composite was a GFRP (2:1, 0/90), with a density equals 380 g/cm3.

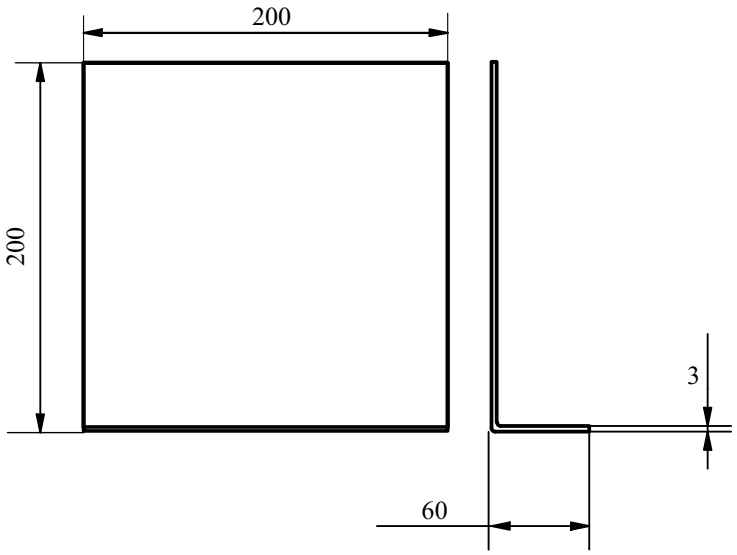


Figure 42 – Metal plate draws onto which defects were added. Dimensions are in millimeters.

Figure 43 shows the localization of each cushion on the plate. This experimentally defined arrangement intends to minimize the effects of fringe overlapping during shearography acquisition. Plates with square-shaped defects are referred to as Q-plates and the other one as SCT-plates. One can notice that Q-plates have more and larger defects because they were laminated first, acting as a proof of concept. It turned out that the larger squared defect, top-left cushion (with an area of 2500 mm^2), leaked on every plate during lamination. It also happened to one of the three top-right squared defect cushions, with an area of 1600 mm^2. That is why there are only three defects of each type on SCT-plates.

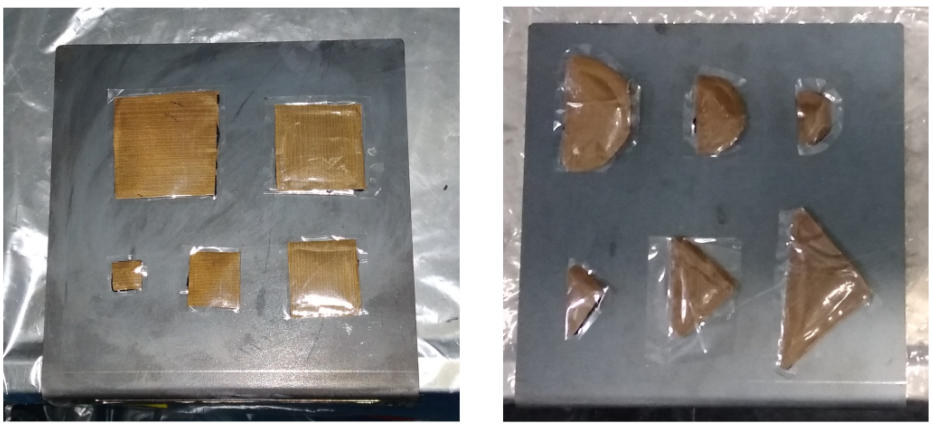


Figure 43 – Left image: the arrangement of squared shaped defect cushions on Q-plate. Right image: the arrangement of semi-circle shape defect cushions (upper part of the metal plate) and triangle shape defect cushions (lower part of the metal plate) on SCT-plate.

The epoxy matrix was a mixture of resin and a catalyst in the appropriate quantities and proportions. Lamination consisted of the intercalated deposition of this epoxy

matrix with the fiber layer up to the desired thickness. A peel ply tissue was applied over the last fiber layer to allow the excess resin to run off and separate the composite from the finishing layers. This tissue was covered with a layer of felt, which retains the excess resin that runs off from the lamination step. This set was packed in a hermetically sealed bag, where vacuum was induced for more or less eight hours. This procedure was finished by removing barbs and painting the surface to aid in the acquisition of images.

Cushions were arranged with a small portion of tape under them to prevent the resin during lamination from entering the small gap in between the PTFE and the metal plate. It also contained the cushion from drifting during lamination stresses. Lamination occurred after this fixation, taking care to cover with resin the whole surface of the metal plate, including the top face of the cushions. Figure 44 shows the appearance of the metal plate after this procedure. In the end, the total composite layer height was around 6 mm.

Figure 44 – Left: SCT-plate with semi-circular after lamination. Right: after it has been housed in a sealed bag and vacuum has been induced.

3.1.2 Shearography image acquisition

Figure 45 shows the layout of the test bench. The plate was positioned perpendicular to the camera's optical axis, and two halogen lamps with 350 W were responsible for the thermal loading. Laser and camera properties are summarized in table 2.

Heat emitters were turned on for 1 s, 3 s, or 5 s to heat the plate in different tests. heating was turned off immediately after a period and the image acquisition began.

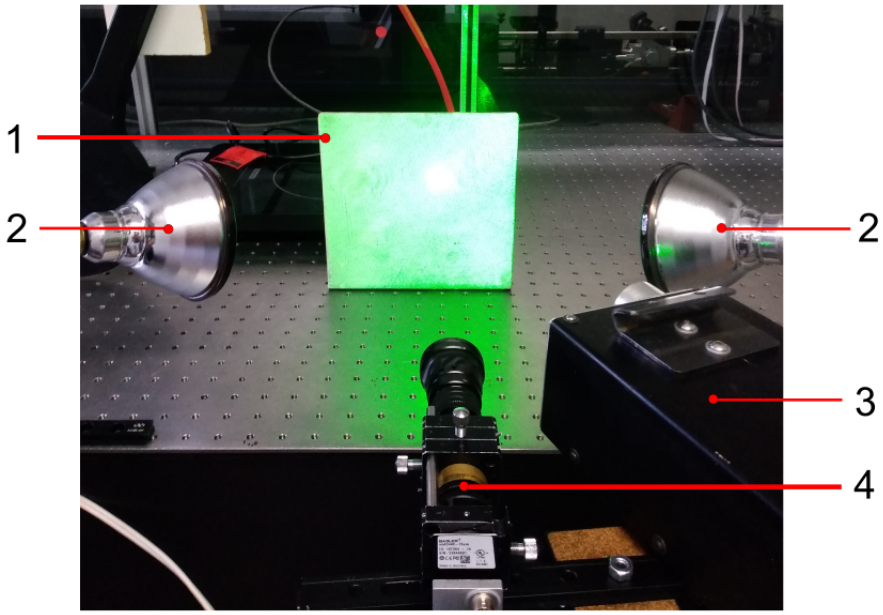


Figure 45 – Shearography specimen and equipment during an acquisition. Legend: 1-plate with defects; 2-lamps for thermal loading; 3-laser emitter; 4-one-shot shearography system.

Table 2 – Summary of laser and camera properties for shearography acquisition. The laser emitter was a Torus 532 MPC3000.

Property	Value
Laser power	500 mW
Laser wavelength	532 nm
Laser coherence	$\leq$100 m
Camera resolution	2448 x 2048 pixels
Image acquisition rate	75 Hz
Lens focal distance	12 mm
Field of view	20 x 15 degrees

Two-hundred-fifty frames were acquired during plate cooling and used for producing the phase maps, as described in section 2. The output of the shearography method is an 8-bit gray-level image, which is considered a form of discretization. Changing the image that is initially processed in radians causes a loss of resolution. It is not common to use images with nearly continuous pixel values as input to neural networks, but this can be accomplished. It is a suggestion for future work to do this test and change to 16-bit resolution and compare the responses.

Figure 46 shows the acquisition structure for each plate. These different load duration were previously tested intervals that produced deformations within a range measurable by the shearography device. The measurement procedure was repeated three times for each heating duration, i.e., once the acquisition was finished, the plate temperature was reduced to the initial temperature and the plate was then submitted to loading again for another acquisition. The temperature was measured with a PT100

sensor, and ranged from 23°C with the plate cooled to 29°C of a heated plate.

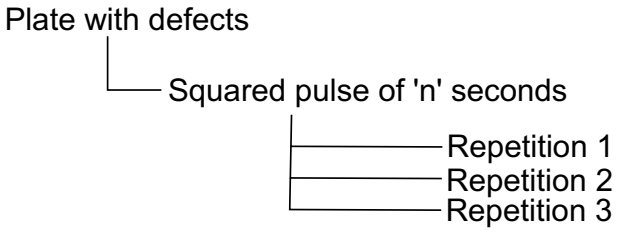


Figure 46 – Structure of the acquisitions for each plate. Three repetitions for each thermal loading duration.

3.2 AUTOMATIC USEFUL FRINGE PATTERN IDENTIFICATION

AUTO-UFP is an algorithm to discard, a priori, images with fringe patterns of insufficient quality to compose the database for deep learning training purposes. A fringe pattern to be considered a UFP should simultaneously have three characteristics: low noise, a minimum contrast difference between sinks and spillways, and a minimum number of well-defined phase jumps. The region of the image from which the fringes appear is called a spillway, and the region where the fringe reduces until it disappears as the fringes progress is called a sink. These features presence mean that it is possible to properly visualize the defect unwrapped phase image and that there are prominent features to train a supervised detector. These findings are made on the wrapped filtered version of the fringe pattern image because the fringe pattern properties are visible and available, and the unwrapping step becomes unnecessary. Considering that a fringe pattern is a UFP is the first step towards fringe pattern detection automation, especially automation with data driven approaches. Automatically analyzing a fringe pattern usefulness minimizes or even eliminates the necessity of an operator for such a subjective task. Besides, many frames coming from a single test make manual annotation for object detectors time-consuming and prone to errors.

Figures 47 and 48 show the appearance of a UFP on its unwrapped leveled version for the three defect shapes. The first figure presents UFP examples, where it is possible to easily visualize the defect on its unwrapped phase leveled image version. The second figure presents non-useful fringe patterns, where the noise and lack of phase jumps inhibit the unwrapping algorithm to provide a defect visualization. It is noticeable in both cases that unwrapped leveled images lack contours, edges, and high contrast. Contours in this context represent the edges of each fringe that builds the fringe pattern of the frame, namely the phase jump regions. It is important to emphasize here that this method receives only one frame at a time containing a single fringe pattern. This limitation makes some prior detection of this fringe pattern necessary, which in this case was done manually.

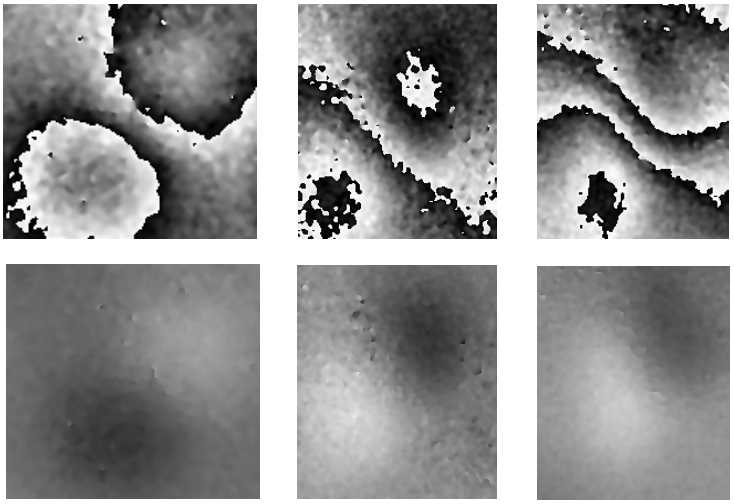


Figure 47 – Filtered wrapped UFP images and their respective unwrapped leveled versions for a (a) square, (b) semi-circle, and (c) triangle-shaped defects.

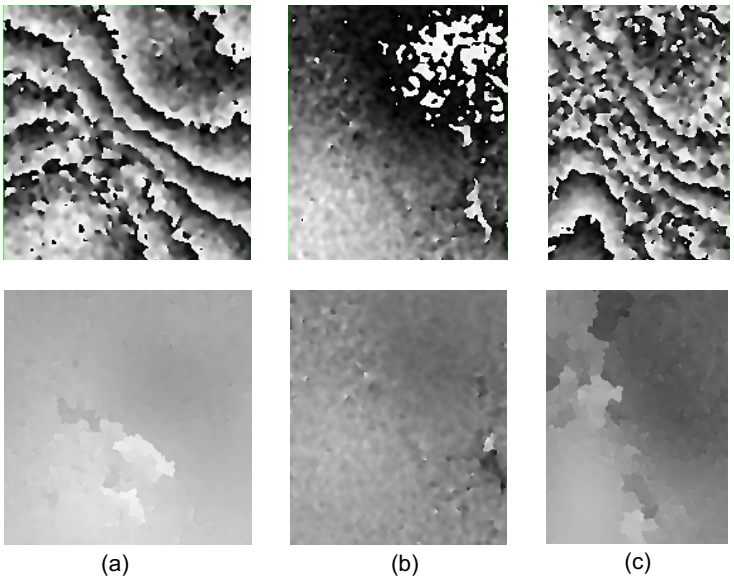


Figure 48 – Filtered wrapped non-UFP patterns and their respective unwrapped leveled versions for (a) square, (b) semi-circle, and (c) triangle shapes of defects.

An unsupervised method was developed that looks precisely for the characteristics representing a UFP in a shearography image. The AUTO-UFP can do that and differentiate the situations presented in figures 47 and 48. Some hyperparameters tuning are necessary, but the dataset for these adjustments is smaller than a dataset for training a specialized CNN with adequate generalization capacity, especially with the noise present in the shearography images and the wide variety of UFP related to each defect format. Moreover, because it is unsupervised, an operator does not need to label each of the fringe patterns, but separate some fringes patterns considered useful and others that visibly present problems for adjustments. As for use, AUTO-UFP is an online method that receives a frame containing a single filtered fringe pattern and returns the

result of 1 for a UFP and 0 for a non-useful fringe pattern. The general flow chart of the UFP method is in figure 49, and the steps are explained in the next sections.

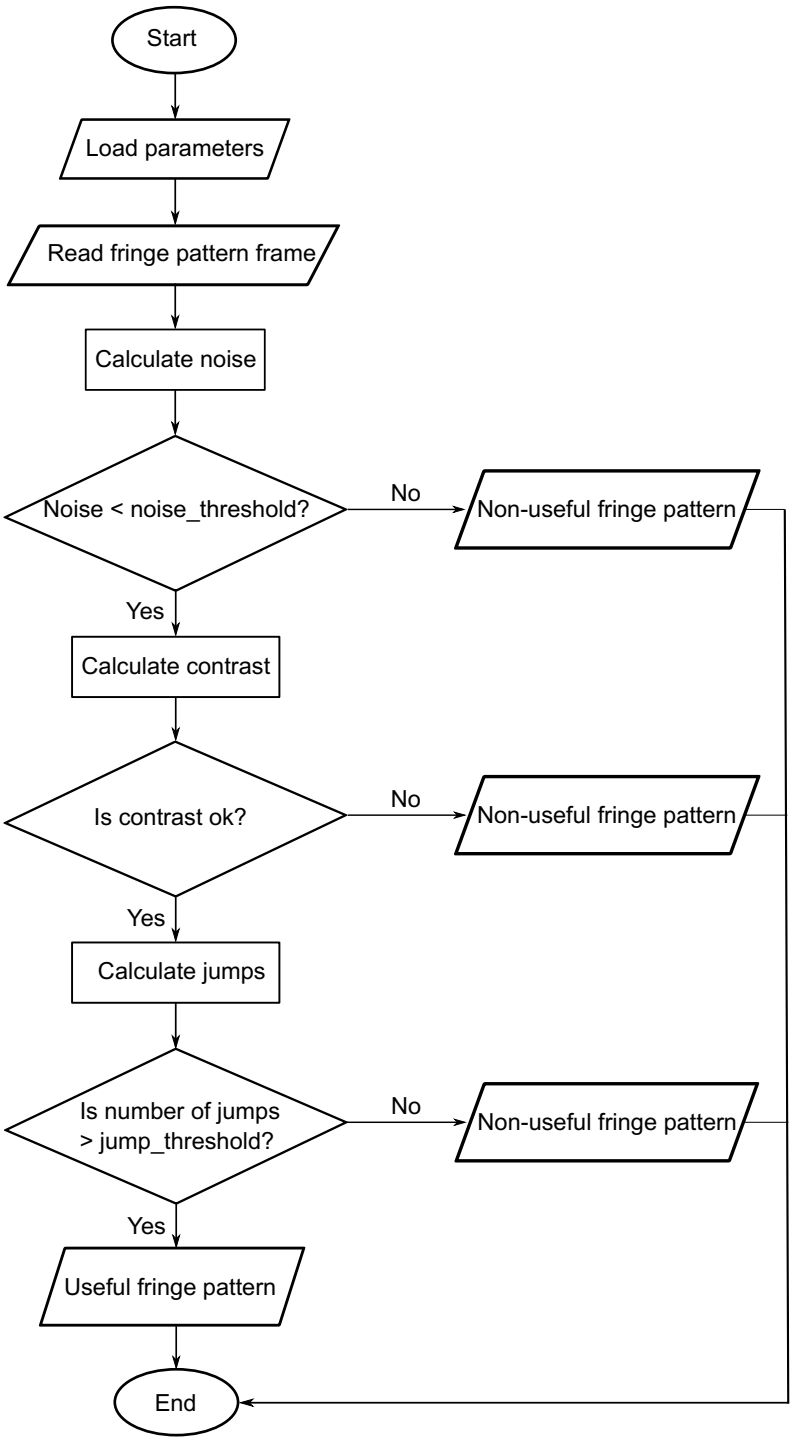
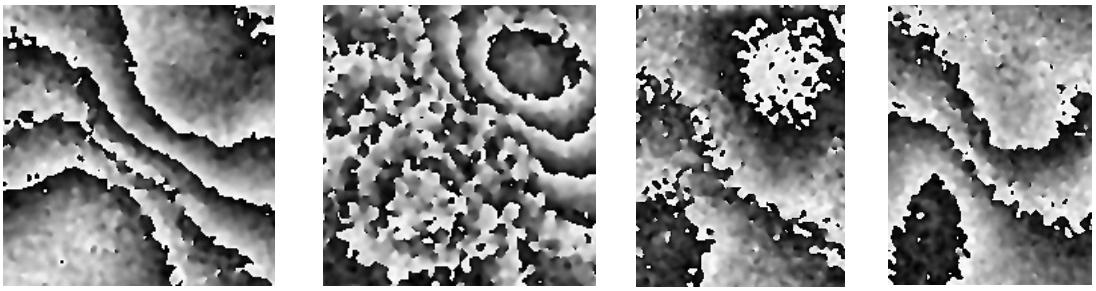


Figure 49 – Fluxogram for AUTO-UFP method.

3.2.1 Noise level check

AUTO-UFP method starts the evaluation with the noise level check on an image containing a single fringe pattern. The method used for noise estimation in the images

was the one presented by Donoho and Johnstone (1994) and implemented by Van der Walt et al. (2014). This algorithm relies on the median absolute deviation of the wavelet detail coefficients. It is assumed that noise necessarily follows a normal distribution. This method traditional application provides a sigma (standard deviation) value for eventual wavelet-based filtering, but the output sigma also works as a noise level indicator. The higher the sigma, the noisier the image is, as shown in figure 50. An acceptance threshold is required to define how significant this sigma value can be. Such a threshold was identified experimentally by applying the method on images with different noise levels belonging to the adjustment image set.

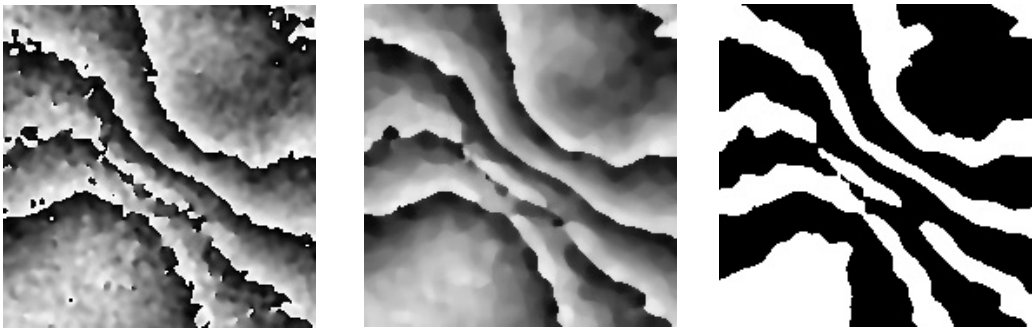


sigma = 1.15 sigma = 3.62 sigma = 1.66 sigma = 1.28

Figure 50 – Different filtered wrapped fringe pattern images and their respective sigma values.

3.2.2 Contrast check

The next steps is the contrast analysis, which depends on a threshold-based binarization of the image performed after a filtering via Gaussian blur. Figure 51 shows an example of this binarization. A direct analysis of the gray values is not enough, because the regions tend to eliminate each other and do not highlight the differences between the image quadrants.



1. Wrapped filtered 2. Blured 3. Binarized

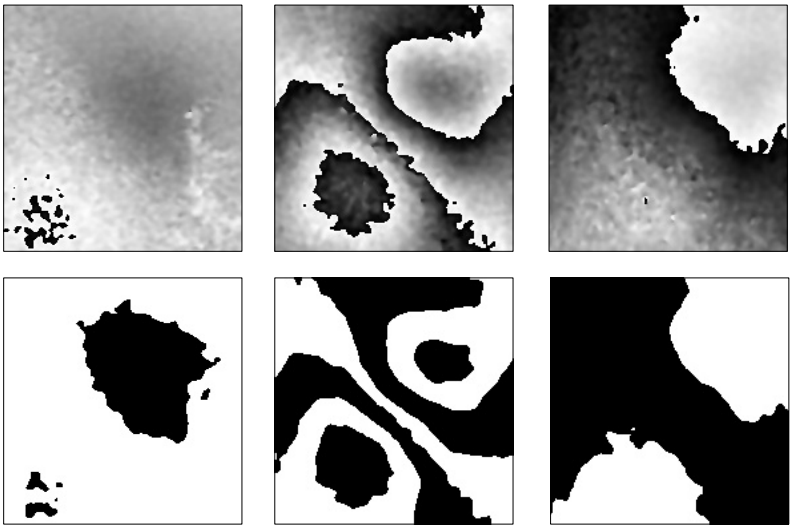
Figure 51 – Steps of binarization of a wrapped fringe pattern image, which is blurred and then binarized. Kernel for blurring equals 9. Threshold for binarization equals 151.

Figure 52 shows results for a global binarization with different binarization threshold values to highlight the importance of this feature. The threshold for binarization is a hyperparameter that must be adjusted.

Blured Threshold 130 Threshold 150 Threshold 170

Figure 52 – Filtered image of the same wrapped fringe patterns of figure 51 and the binarization results for different threshold values.

As the pixels of a binary image are either black or white, an image passes the contrast test if it has a minimum number of white pixels. This black-white fraction is also a hyperparameter and must be adjusted with the adjustment image set. Different proportions of black and white pixels are show in figure 53, which shows three fractions examples with a binarization threshold of 151. The image to the left shows the result of an SC defect, where the fraction of black is smaller. The other two images show square defects with appropriate fractions. Yet, passing the contrast test is no guarantee that the fringe pattern is useful. The most right image of figure 53 would not succeed on the next test because there is only one phase jump.



21% of black pixel 49% 60%

Figure 53 – Examples of filtered fringe patterns with their binary images and respective percentage of black pixels.

3.2.3 Phase jump counting

Phase jump counting is more complicated than contrast analysis because it involves the binarization step, filtering, counting signal peaks, and one extra rule associated with peak position. Initially, a curve of pixel intensities crossing the image diagonally is extracted from the binary image, which comes also from the threshold-based binarization. This curve, namely diagonal line, starts from the upper-right corner and ends on the bottom-left one. Such a direction is chosen because it is already known that the shear is in the *XY* direction. Starting and ending points must change if the fringe pattern is in another shear, as shown in figure 54.

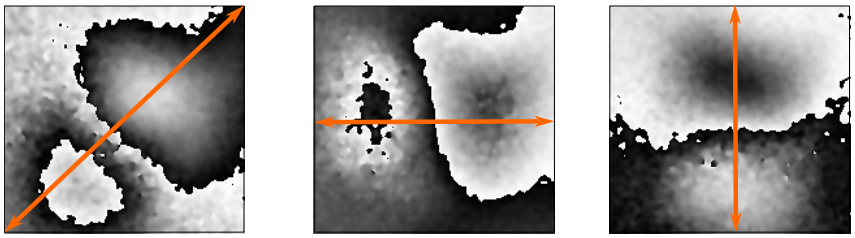


Figure 54 – Different shear directions for a wrapped fringe pattern image of a squared-shaped defect and diagonal lines that characterize the most the phase jumps.

A signal filtering step is necessary because noise level verification takes the whole fringe pattern image, and even in images that pass this verification, the diagonal line can yet be noisy and generate saturated pixel intensities. This filtering process is via Savitzky-Golay filter (SAVITZKY; GOLAY, 1964; VIRTANEN et al., 2020). Savitzky-Golay is a low pass filter based on local least-squares polynomial approximation, i.e., it fits a polynomial of arbitrary order to a certain odd number of input samples in the form of a sliding window with stride equals one. These points are weighted accordingly to a finite response signal, and the evaluation and output come from the middle point of this window. The primary benefit of this filter is the peak shape preservation property (SCHAFER, 2011; SAVITZKY; GOLAY, 1964). The major benefit of the method in this work is plateau removing. Also, window length had five points, and the polynomial order was equals two.

Peaks that represent phase jumps are extracted from this filtered diagonal line. Such peaks come from neighborhood comparison. There is an influence on the height of a point concerning its neighbors. Thus, height is a parameter, and the proximity between peaks is a hyperparameter, reaching the three hyperparameters needed to adjust the method through grid search.

Figure 55 shows an example of the results of each stage of phase jump calculation, from the input image to the peaks counting. There are plateaus of white pixels on image (iii), which are turned into one or two peaks depending on their size via Savitzky-

Golay filtering — only peaks beyond nine indexes of distance are counted. Also, there are peaks in both halves of the filtered diagonal line. The rule of peak location discards curves in which there are phase jumps, or peaks, in only one of the halves. As the minimum number of jumps is two, which was defined experimentally, there should be one jump at each of the filtered diagonal line halves. Image (iv) shows a curve that satisfies this rule.

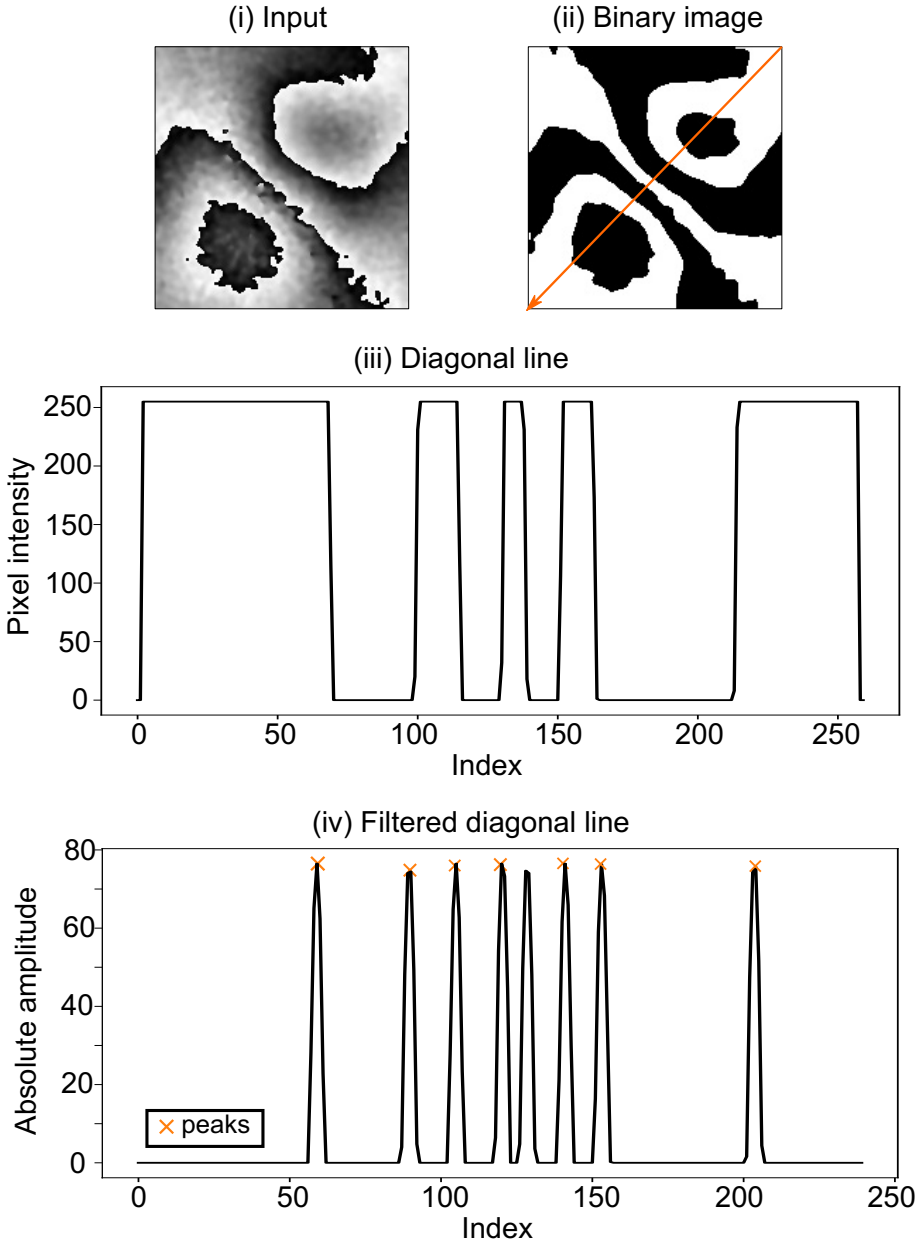


Figure 55 – Stages in the calculation of the number of peaks. (i) Input image. (ii) Binary image and Contrast Line direction. (iii) Contrast Line. (iv) Filtered Contrast Line and peaks found.

Summing up, all parameters that are little influenced by the dataset at first are the minimum number of phase jumps, sigma value for the noise level, the height difference of close peaks during contrast check, and kernel for the Gaussian blur during image

binarization. Moreover, the three hyperparameters that depend on the dataset and objective of the grid search are:

- the threshold for image binarization;
- a fraction between ones and zeros after image binarization;
- and distance between filtered Contrast Line peaks.

A fringe pattern to be considered a UFP should simultaneously have three characteristics: low noise, a minimum contrast difference between sinks and spillways, and a minimum number of well-defined phase jumps. These features mean that it is possible to visualize the defect when the phase wrap is removed (unwrapped leveled image) and that there are prominent features to train a supervised detector. These verification are made on the filtered version of the fringe image because the fringe properties are visible.

4 RESULTS AND DISCUSSION

This chapter starts with the results concerning shearography acquisition, section 4.1. The details of the manually selected data set are in section 4.2. Section 4.2.2 presents the results regarding the Grid search and use of the AUTO-UFP method to generate the dataset automatically. Finally, section 4.3 has the results concerning the detector trained with the manual dataset and the AUTO-UFP dataset, as well as a comparison between these models.

4.1 SHEAROGRAPHY ACQUISITION

Figure 56 shows a fringe pattern frame extracted from the original image of a shearography acquisition and the two subsequent image processing steps of a heat loading of a squared-shaped defect. The original image is noisy and has no well-defined contours, which does not occur in a filtered version. A filtered fringe pattern has the noise significantly reduced and now presents contours and higher contrast. The last processing level suffers from the same lack of contours as the original image, although it expresses the defect area well. These features led to the choice of the wrapped filtered image for all further developments. For each loading condition a temporal sequence of 250 frames was acquired during cooling. Since three heating times were used and three repetitions were performed, a total of 2250 images are available for each specimen.

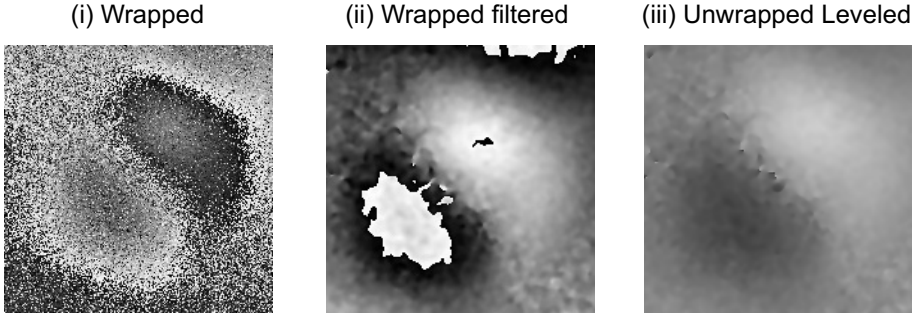


Figure 56 – Example of output of a shearography acquisition.

Figure 57 shows the terminology that will be used from now on to refer to the positions that each defect occupies on a respective plate. Position A of the Q-plate shows a leaked defect. A difference between this leaked defect to the other ones is evident on the unwrapped leveled image. Even more, a visual inspection revealed a recess on the region of a leaked defect. Lamination was made with a few layers of composite, but this was intentional and referred to the exploratory characteristic of this work.

A leaked defect is not useful because it has a different appearance in the filtered version of the wrapped image, as shown in figure 58 in position A of a square plate. Leakage also occurred on one of the plates at position B. Small defects, i.e., those

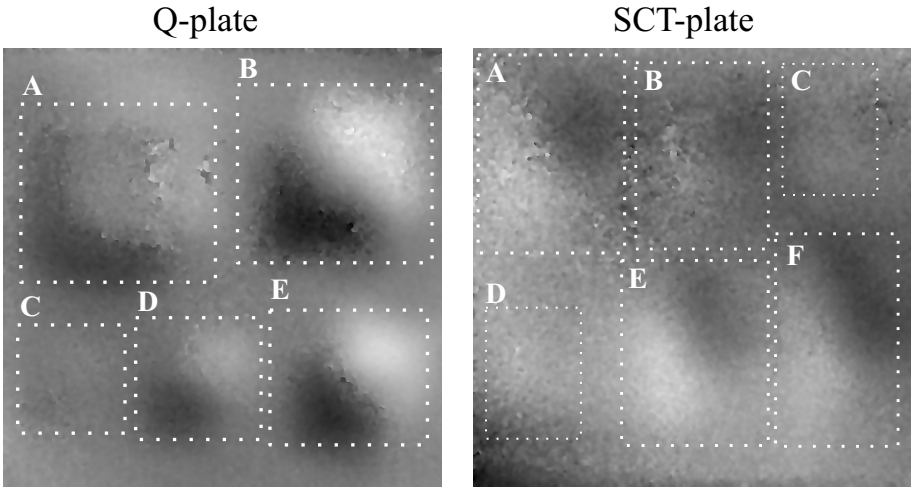


Figure 57 – Position mapping of defects on each plate type.

at position C of the Q-plate and positions C and D of the SCT-plate did not appear in the acquisitions most of the time. So they were not considered for data set building porposes.

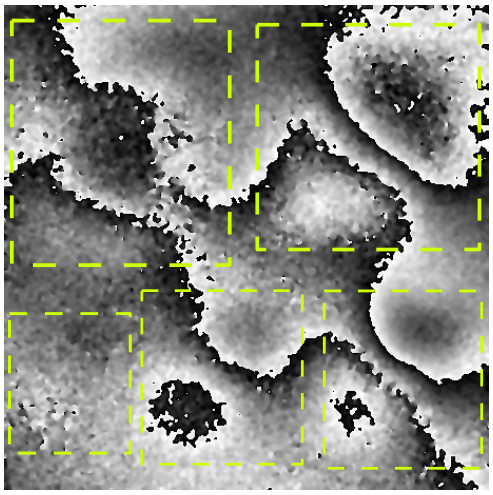


Figure 58 – Acquisition from a Q-plate. Fringe pattern in position A from a leaked cushion.

Given the observations mentioned above, only defects present in positions D and E in the Q-plates and A, B, E, and F positions of the SCT-plates were used for development to standardize the dataset.

4.2 ANNOTATIONS FOR USEFUL FRINGE PATTERN DETECTION

Fringe pattern can present themselves at any location of a shearography image, and the classification task depends on that. An OD can solve these problems simultaneously, provided the dataset used for its training is varied and contains the expected class images. This section presents the manual and AUTO-UFP datasets construction and how it is presented for model training. For all these developments, only wrapped

filtered phase image from *XY* shear directions were used, as they posses more features to be extracted from.

4.2.1 Manual annotation

A batch of shearography acquisitions has 250 frames. Each frame has bounding boxes that were manually annotated to cover the defects in both plates. It was chosen the best frame that matches each defect position in the original image, then this bounding box was replicated for all plates. However, this sequence of 250 frames does not have only bounding boxes with useful fringes within them, e.g., when preprocessed, some do not provide a visible defect. It is necessary to analyze the wrapped filtered image and its unwrapped leveled version to state if a defect is visible.

Figure 59 shows a Q plate frame with its filtered phase map and unwrapped leveled versions. On the left image, the green lines represent the bounding boxes with no overlapping between them. Thus, a bounding box has a useful fringe pattern on the filtered phase map if the defect is evident on the right image. For example, a defect on position *C* is not visible on the figure above, but positions B, D, and E have useful fringes.

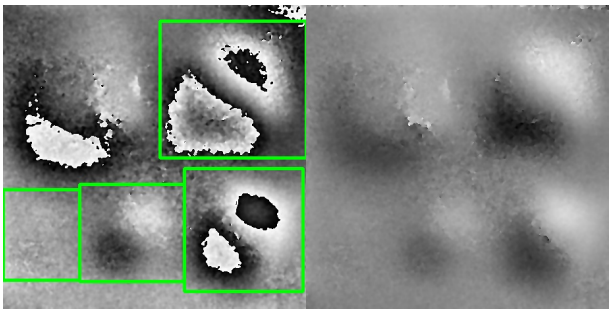


Figure 59 – Bounding boxes of a Q-plate. Upper left defect is neglected since the air leaked during manufacturing.

Each frame was analyzed individually to record the interval of useful fringes per batch of acquisitions. Some frames have useful fringes only for some positions since temperature evolves differently at each region of the plate. Furthermore, some positions have only useful fringes within specific intervals. These intervals change even for the same plates due to acquisition settings.

Table 3 presents an example of useful fringes intervals for a Q-plate acquisition. A range of useful fringes on position D starts at frame 53 and ends at frame 145. Frames within this range are considered positive observations, and all frames outside this range are negative observations. All intervals targets come from the plate and the position from which they were extracted. Plate Q defects always have square type defects. For SCT plates, defects in the plate upper position (positions A and B) have semi-circular shapes, and defects in the lower positions (D and E) have a triangular shape.

Table 3 – Intervals with useful fringes for the D and E positions of a Q1 plate, with 1 second squared pulse duration, from repetition one.

Position	Starting Frame	Closing Frame
D	60	145
E	53	145

One of the three plates with square defects was dedicated to training processes, one to validation, and the third for testing. The amount of observations used for training (fringe pattern examples), validation, and testing is summarized in table 4. It is crucial to notice that doing repetitions of the same plate with the same plate and for the same heat duration can be considered data augmentation. Different heat duration can hardly inhibit the model, but perhaps the three repetitions can reduce model generalization, mainly because it propagates unbalancing. However, this step was necessary to generate the number of observations for the training procedure.

There is an imbalance of the semi-circular defects concerning the other shapes. The implications of this unbalance are discussed in section 4.3. There is also a more considerable amount of observations for the training set than for validation and testing. As a detector, the RetinaNet was chosen for the reasons presented in the theoretical background, section 2.3.3.

Table 4 – Amount of observations used for training, validation, and testing of the fringe pattern detector according to the position and the format of defect that originated the fringe. Q stands for square, SC stands for semi-circle, and T stands for a triangle. Manually obtained.

	Train/Val/Test				
Q		SC		T	
Position	Number of fringes	Position	Number of fringes	Position	Number of fringes
D	824 / 1130 / 932	A	1005 / 318 / 465	E	1349 / 681 / 912
E	826 / 1175 / 1018	B	904 / 333 / 465	F	1349 / 681 / 912
Subtotals	1650 / 2305 / 1950		1909 / 651 / 930		2698 / 1362 / 1824
Total	6257 / 4318 / 4704				

4.2.2 AUTO-UFP results for dataset creation

As presented in section 3.2, AUTO-UFP requires some parameters to be manually adjusted, and the other three hyperparameters have to be found, initially at least, through a grid search. Only training observations (plates Q1 and SCT1) were used for parameter and hyperparameters adjusting. Adjusted parameter values are as follows, and they were obtained experimentally by analyzing nearly ten useful and non-useful fringes:

- Minimum number of phase jumps: 2 (dimensionless);

- Maximum sigma value for noise: 1.5 gray-level;
- Minimum distance of close peaks: 9 pixels;
- Kernel for Gaussian blur prior to binarization: 9x9 pixels.

It was possible to have a first estimate of the hyperparameters values after adjusting these aforementioned parameters, thus defining the grid search intervals. Table 5 shows the intervals of the hyperparameters and their respective averaged F-scores, where a low variability between the combinations can be noticed. These averaged F-scores are calculated from scores obtained from each of the batches of 250 images originated from the training set. The best F-score found of 0.91 is related to a threshold of 150 and tolerances of 20 and 25, among others. A second grid search, more refined than the first, was conducted around these values. This second grid search, presented in table 6, did not improve the F-score of 0.91. Such a constancy demonstrates robustness in the method, thus these hyperparameters could also be considered just parameters and could come from the same experimental analysis performed on few images as the others.

Table 5 – Coarse grid search parameters and their respective averaged F-scores.

Threshold	Factor	Distance	F-score
140	15	5	0.910
140	15	10	0.910
140	15	15	0.907
140	20	5	0.910
140	20	10	0.910
140	20	15	0.907
140	25	5	0.903
140	25	10	0.903
140	25	15	0.901
145	15	5	0.909
145	15	10	0.909
145	15	15	0.907
145	20	5	0.909
145	20	10	0.909
145	20	15	0.907
145	25	5	0.903
145	25	10	0.903
145	25	15	0.901
150	15	5	0.910
150	15	10	0.910
150	15	15	0.909
150	20	5	0.910
150	20	10	0.910
150	20	15	0.909
150	25	5	0.900
150	25	10	0.900
150	25	15	0.899

Table 6 – Refined grid search parameters and their respective averaged F-scores.

Threshold	Factor	Distance	F-score
149	14	9	0.910
149	14	10	0.910
149	14	11	0.910
149	15	9	0.910
149	15	10	0.910
149	15	11	0.910
149	16	9	0.910
149	16	10	0.910
149	16	11	0.910
150	14	9	0.910
150	14	10	0.910
150	14	11	0.909
150	15	9	0.910
150	15	10	0.910
150	15	11	0.909
150	16	9	0.910
150	16	10	0.910
150	16	11	0.909
151	14	9	0.910
151	14	10	0.910
151	14	11	0.910
151	15	9	0.910
151	15	10	0.910
151	15	11	0.910
151	16	9	0.910
151	16	10	0.910
151	16	11	0.910

After this grid search, hyperparameters were set as 151 gray-level for the binarization threshold, 14% for a fraction of black and white pixels tolerated after binarization, and 9 pixels for the distance between peaks during phase jump analysis. These values were used to infer fringe pattern quality over the entire original dataset (training, validation, and testing). Like table 4, table 7 summarizes the number of useful fringe pattern observations found by the method. One can see that the total for the three split does not differ too much from those obtained with manual annotation, except for a larger unbalance on class T. AUTO-UFP granted 6292 fringes for training against 6257 from manual annotation, 3853 for validation against 4318, and 4297 for test against 4704.

Comparing the total number of observations is not enough to confirm this dataset usefulness, neither the F-score showed before. It is necessary to train a detector and compare it to another detector trained with manual annotations to ensure that AUTO-UFP yields a useful dataset.

4.3 FRINGE PATTERN DETECTION AND SHAPE CLASSIFICATION

In this section the object detector training setup is presented, both for the dataset with the manually catalogued positive and negative observations and for the dataset with

Table 7 – Number of observations acquired from AUTO-UFP and used for training, validation, and testing of a fringe pattern detector according to the position and the format of defect that originated the fringe. Q stands for square, SC stands for semi-circle, and T stands for triangle-shaped defects.

	Train/Val/Test				
Q		SC		T	
Position	Number of fringes	Position	Number of fringes	Position	Number of fringes
D	726 / 888 / 808	A	1036 / 280 / 262	E	1433 / 491 / 822
E	831 / 1095 / 936	B	627 / 363 / 395	F	1639 / 736 / 1074
Subtotals	1557 / 1983 / 1744		1663 / 643 / 657		3072 / 1227 / 1896
Total	6292 / 3853 / 4297				

these observations classified using the AUTO-UFP method, as well as the evaluation and comparison of the results of these models.

4.3.1 Detector training

Images from one plate of each type were dedicated to training, one for validation, and the last one for testing. The training for both models occurred with an early stopping of five epochs, where the metric analyzed was the mAP of the validation set. For the training to progress, the minimum delta of this mAP should be at least 0.01 higher than the last best epoch. No data augmentation was used since there is little difference between classes, and changes in shear, zoom, and other operations would alter the fringe pattern format, hindering the learning. Rotations and flips were also not used, so the shear direction was always the same and that there was no defect rotation. However, applying rotations and flips is highly recommended in real applications, where the validation and test set are unknown. Using simulated fringes would be an option, if the problems in the fringe pattern could also be simulated. More on this topic on section 5.1.

The initial learning rate was 10^{-5}, with a reduction of 10% of this value in the case of plateaus. A plateau consists of no reduction greater than 10^{-5} for validation loss. Other parameters were batch size equals one because of GPU memory issues, steps per epoch equal to the number of images in the training set, IoU threshold of 0.5, and a threshold for the classification score of 0.05. The training set consisted of images with positive and negative examples. The validation and test sets had images with no useful fringe pattern on them, making evaluation closer to a real application and, consequently, more difficult.

Figure 60 shows the evolution of losses during RetinaNet detector training with ResNet 50 as the backbone. This backbone has enough complexity and depth for such a small number of observations and a low amount of classes to be detected. A deeper backbone like ResNet 101 would bring overfitting problems. This figure shows

a coherent evolution of both losses. The next figure, figure 61 presents the inter-class evolution of average precision metric and average value, mAP. The best epoch was the tenth, where an mAP of 0.86 was achieved as the five subsequent epochs did not show the necessary improvement.

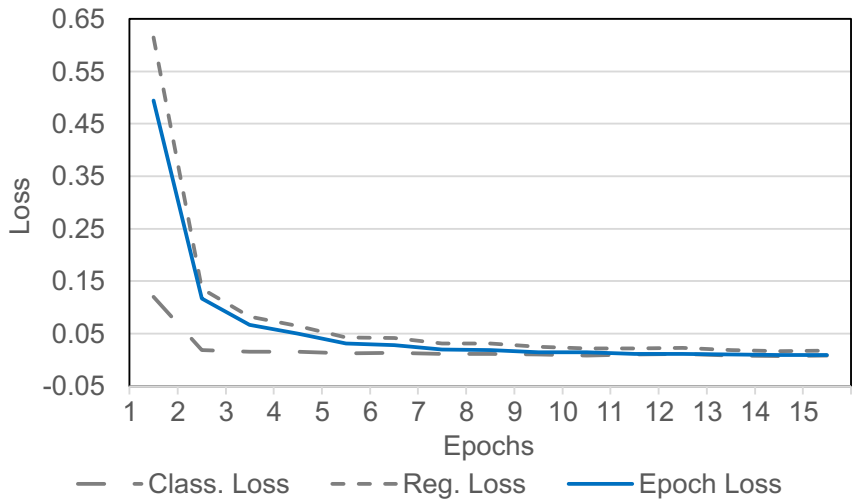


Figure 60 – Progression of losses during the training that used the set created manually.

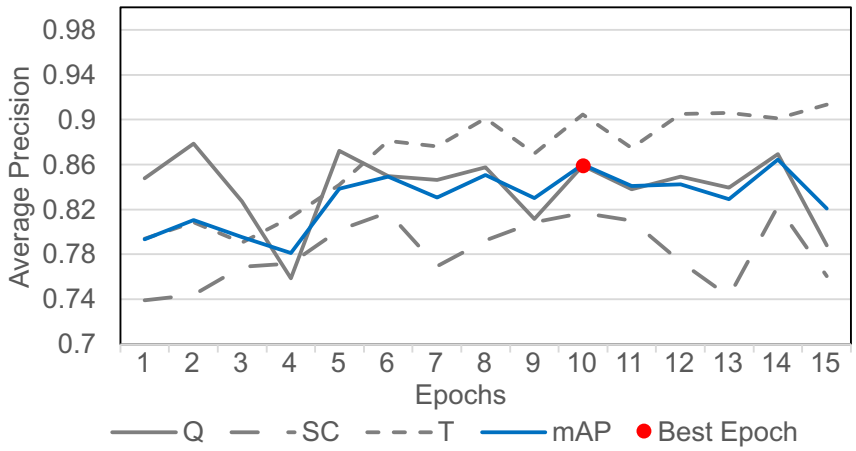


Figure 61 – Progress of the AP and mAP of the validation set during the training that used the manually created set. Best epoch mAP is 0.860, coming from the tenth epoch.

Figures 62 and 63 show the evolution of a detector trained with the dataset created from the AUTO-UFP tool. In this case, training was interrupted in the sixth epoch since the best one was early the first one, with an mAP of 0.865. It is noted in Figure 63 that the performance is decreasing as the training progress, an indication of overfitting. The evolution of losses presented in Figure 62 validates this assumption.

The best interclass result occurred with the square shape defects and the worst for the triangular shape defects.

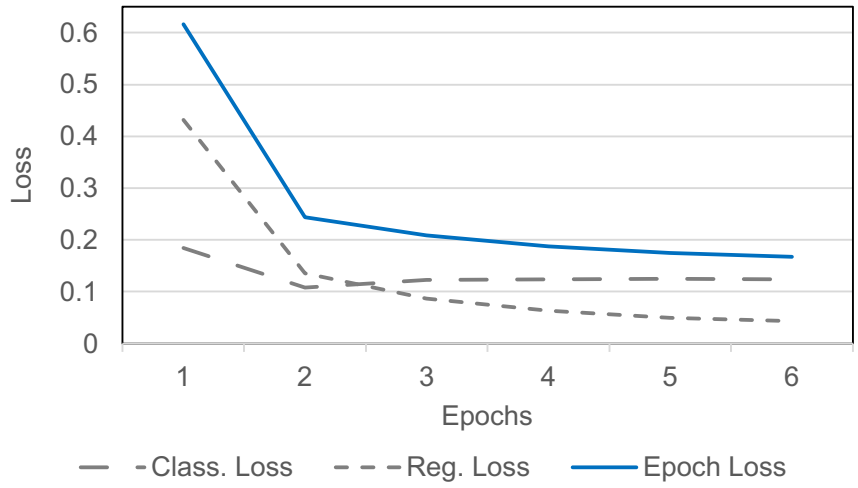


Figure 62 – Progression of losses during the training that used the set created with AUTO-UFP.

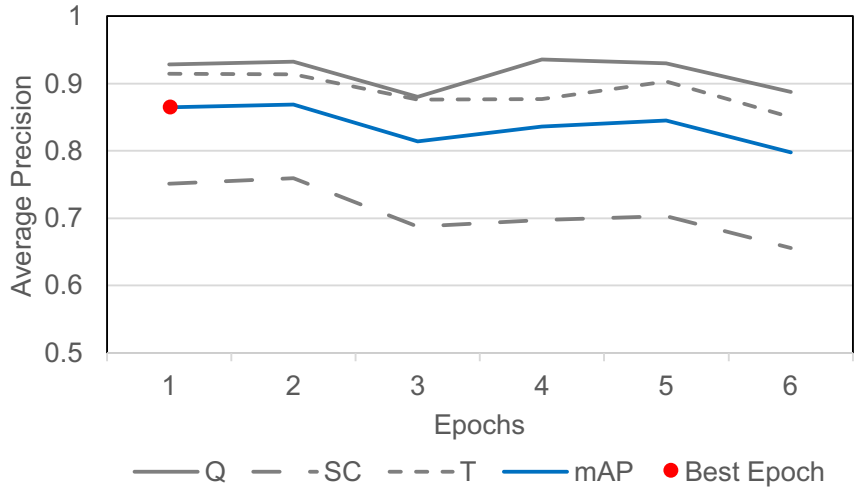
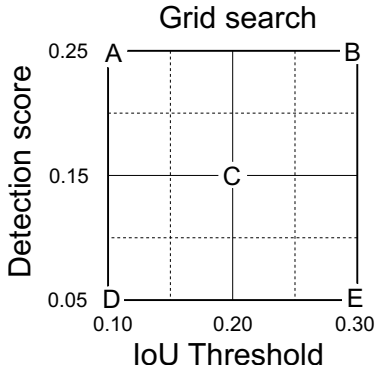


Figure 63 – Progress of the AP and mAP of the validation set during the training that used the set created with AUTO-UFP. Best epoch mAP is 0.865, coming from the first epoch.

By analyzing these progressions, it can be seen that the model trained with the AUTO-UFP converged in the first epoch with the same mAP of the base model. This performance suggests that the AUTO-UFP method not only aids in classifying the observations between positive and negative but also acts like a preprocessing step to the training, helping the model by keeping fringe patterns with valuable features.

Although both models were trained with an IoU score of 0.5 and a classification score of 0.05, another combination of these hyperparameters could produce a better

result. Figure 64 presents the results of a mid-point design of experiments, where an error surface was built to choose the combination that yielded the least error. Such an evaluation came from the application of the trained models on the validation set. For both dataset combination, D provided the best results. The difference between these results and those mentioned in figures 61 and 63 is due to the IoU. Identical results are correct and are the result of number rounding.



models mAP

	Manual	AUTO-UF
A	0.839	0.740
B	0.839	0.740
C	0.842	0.835
D	0.848	0.874
E	0.848	0.872

Figure 64 – Search Grid and respective mAP for different Resnet Models. Combination D yielded the best results for both models.

4.3.2 Test evaluation

There is a possibility of two or more classes sharing close or overlapping regions in multiclass detection problems, like a person riding a bicycle. In this case, both the bicycle and the person must be detected, and their bounding boxes will have plenty of overlapping. Preliminary analyses on fringe pattern detection have shown a strong tendency to overlap, both interclass and intraclass, so the interclass non-max supression was activated during the test to remove such an overlap. Also, a more restrictive IoU was imposed to solve the interclass overlap because of the defects proximity, starting from 10% as shown in table 64. This IoU is also valid for interclass overlaps.

There are many detections with low scores in fully noisy regions due to the nature of our problem, so the maximum number of detections during testing was set to four. One possible solution would be to increase this threshold, but the semi-circular class also has a low score in correct detections. Thus, increasing the classification threshold would eliminate detections in that class.

Tables 8a and 8b present the execution results of the models presented in section 4.3 in their respective test sets. As in the validation procedure, both sets also have negative examples, totaling about 4500 images per set, where the number of positive observations has already been discriminated in tables 4 and 7. The model trained with the manual dataset presented an mAP of 0.772. A manual dataset is built with closed image intervals. A closed interval contains all the fringes between the minimum index and the maximum index. In this way, it is expected that the manual dataset will be

more robust and have less dispersion in the format of the fringes in terms of its quality, which is expressed in the form of contrast and well-defined phase jumps. The cost of this dispersion coming from AUTO-UFP dataset building is 6% on the mAP, resulting in 0.710. Concerning the classes, 8 shows an exchange of AP between classes Q and T. It is reasonable to say that AUTO-UFP has issues to calculate phase jumps for Q-type defects. These type of defects have lots of phase jumps in the fringe pattern, an sometimes these fringes are not continuous or well defined. T-type defect has less phase jumps, but they are continuous and well-defined, thus they are properly perceived by the jump calculation.

Table 8 – (a) Results for the application of the model trained with the manually created dataset and (b) with the dataset created by the AUTO-UFP method.

Manual Test Set		
Class	Instances	AP
Q	1687	0.901
SC	1638	0.729
T	1638	0.681
mAP: 0.772		

(a)

AUTO-UFP Test Set		
Class	Instances	AP
Q	1744	0.667
SC	657	0.551
T	1896	0.917
mAP: 0.712		

(b)

Sc-type defects present presented not good results overall. Four causes can explain such a low mAP related to SC type plates:

1. High number of defects on plate SCT (6 per plate) hinders the creation of well-defined edges, overlapping the fringe pattern evolution.
2. The way manual annotations are made, taking into account the image coming from its unwrapped version. Some fringe boundaries could have been denied this way.
3. The small number of SC type examples about the others, since the same amount of useful fringes has not been formed.
4. Poor heat transfer in the upper region of the plate in terms of acquisition than the lower area, where the square and triangular-shaped defects were.

In this section figures, green boxes represent ground truths, orange boxes are Q detections, blue are SC detections, and dark yellow is T detections. If the score is in black, the detection was incorrect and white if it was correct. Only the first four highest score detections are considered, and their score should be higher than 0.05. Also, the maximum overlap allowed is 10%. The worst scenarios regarding the output of the fringe pattern detection are as follow:

- False positive: it means that the method detected a useful fringe pattern when there was not, implying a decision that should not occur. Refer to section 5.1 to avoid this issue.
- Superposition of fringe patterns: a fringe pattern that comes from a superposition of two or more detected fringes. This happens not only when there are two or more defects in the image but also when there are false positives together with true positives. IoU and classification score need to be correctly adjusted.
- Low detection score: a low score detection is not an issue since it does not affect the bounding box position. However, there is a score threshold, and below that threshold, detections might not occur.

Figure 65 shows examples of successful detections. In the left image, detections and correct classification of Q defects occurred at D and E positions, and a false positive at position C. Fringe pattern detection at position C demonstrates the model capability to handle smaller fringes than those presented during training as long as it has the desired features. Detections like the last one also occurred on Q-type plates with non-leaked defects at position B. The image on the right has the four expected detections — two semicircular defects in the upper region and two triangular defects in the lower region.

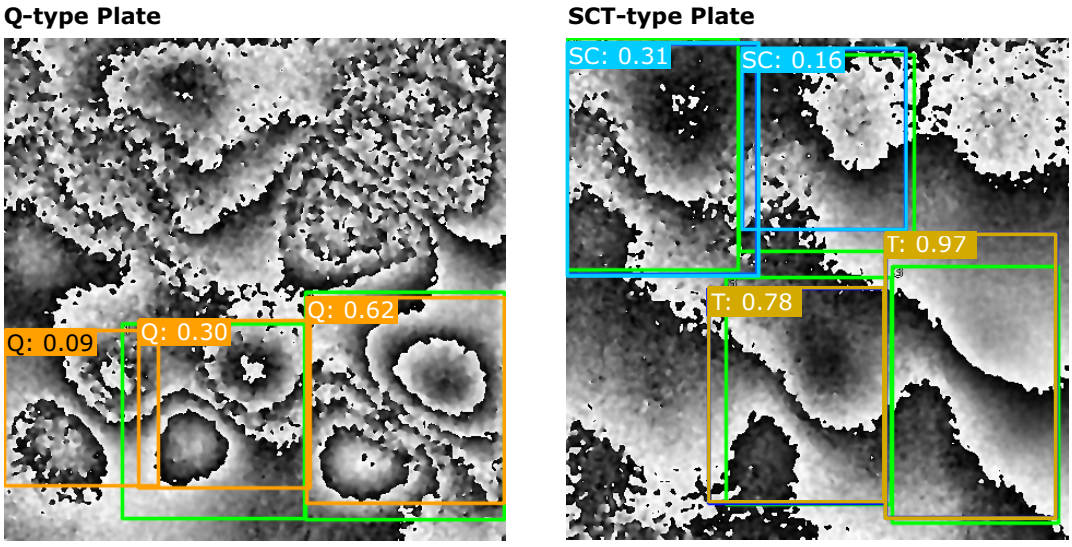


Figure 65 – Examples of acceptable detections on plates of both types.

Due to the network generalization capacity and the subjective criteria of notes made for defining useful fringes, the network can perform accurate detections immediately before or immediately after ground truths. This is the case of the left image of figure 66, where fringes were identified, but there were no manual annotations, thus being considered false positives.

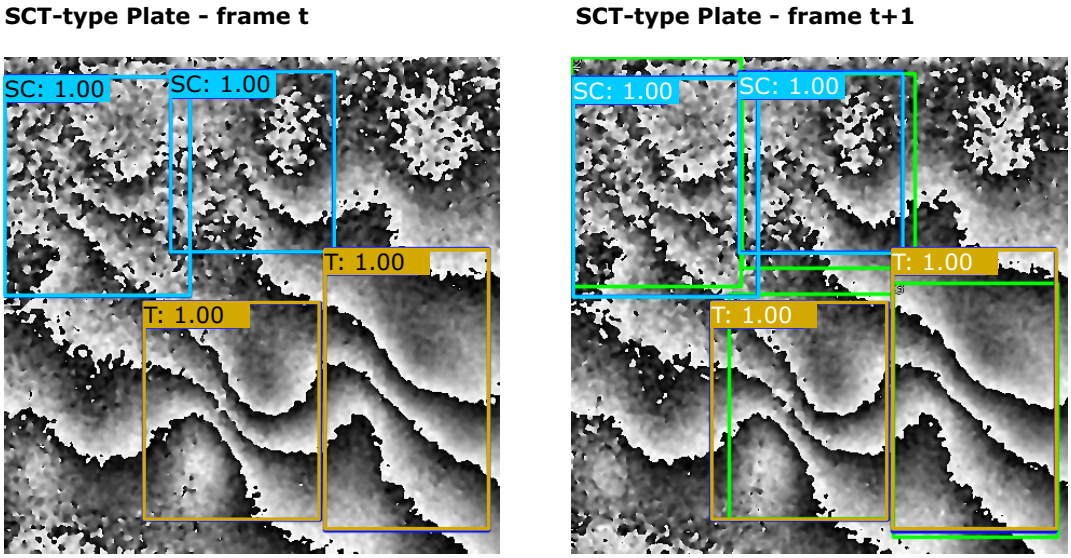


Figure 66 – Examples of the network ability to extrapolate detections in situations where there was no useful fringe pattern annotation due to the subjectivity of manual annotation.

Figure 67 shows several forms of wrong detections. In the left image, a Q-type plate, there is only one correct detection at position D. The image on the right shows detections in noisy places where only a small part of a fringe pattern is formed.

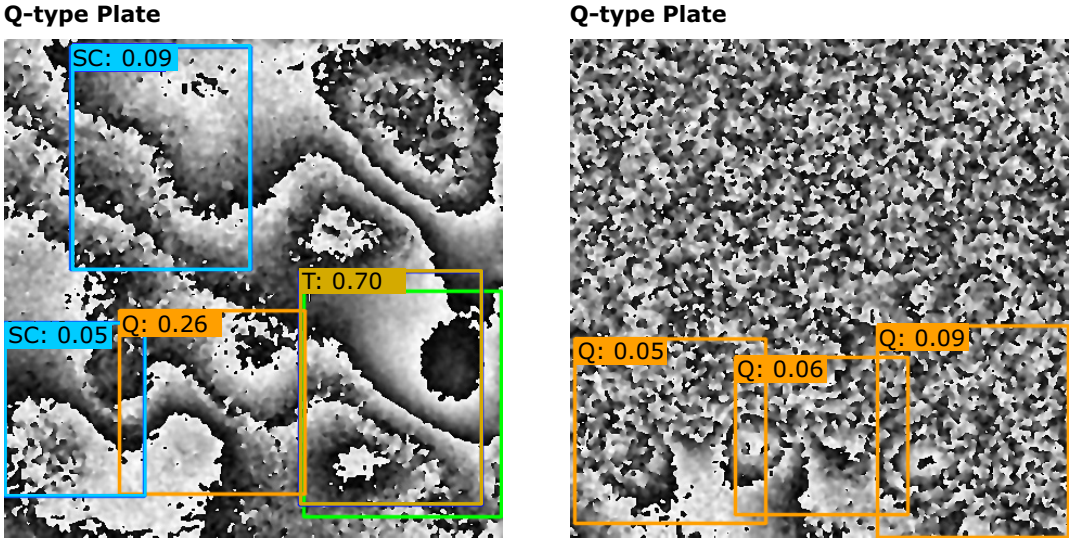


Figure 67 – Examples of incorrect detection for Q-type plates.

Figure 68 presents both correct and wrong detections. In the left image, SC detections have been assigned to fringes formed from a leaked defect. Still, the model could extrapolate and identify the defect on position B. In the right image, there are Q detections in a noisy frame.

Finally, one last situation that can occur is mixing characteristics of two defects in the same detection, as shown in figure 69. There is only one correct detection in this image, on the E position. The other two square defect detections mix characteristics

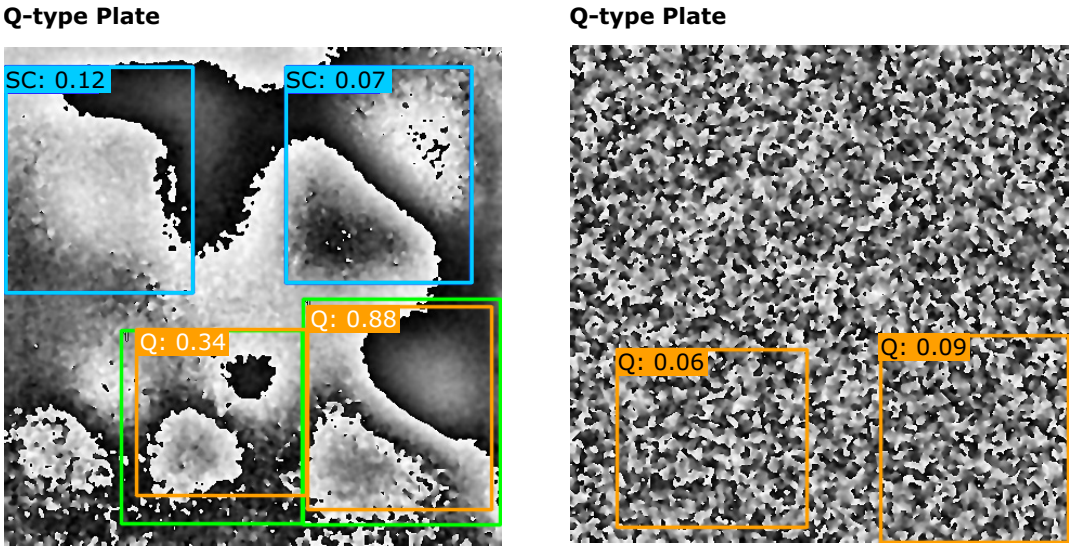


Figure 68 – Examples of correct and incorrect detections for Q-type plates.

of fringes from two distinct positions. The high number of false positives occurred mostly in certain regions of the plates that happened to be noisier than others. It is reasonable to state that the detector is capturing specific unnecessary patterns from these background regions during training and relating them to fringes. Usually, these false positives are punctual and do not all comprise frames in closed intervals.

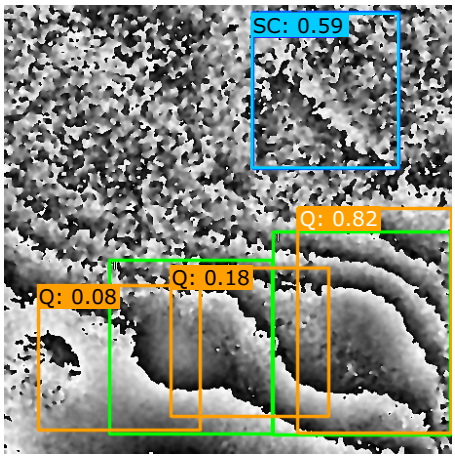


Figure 69 – Example of detections that capture properties of two fringes simultaneously.

5 CONCLUSION

This work proposed an exploratory analysis of the capability of intelligent methods to locate and classify defects of adhesion lack between glass-fiber-reinforced polymer (GFRP) repairs and its substrate from images obtained by shearography.

It is difficult to find publicly available datasets with shearography images, even more with a specific defect type. It is known that delamination defects are more common and more easily reproducible. Artificial delamination defects aim to emulate the existence of air between layers of composites. Adhesion defects are similar in this respect, as they present air between the substrate and the first composite layer applied. Thus the same strategy to emulate the air was used, but with care to ensure that the defect was perceptible. For this purpose, polytetrafluoroethylene (PTFE) pads were created and filled with small volumes of air inserted using a syringe, from which a more realistic behavior was expected than the use of PTFE or Kapton tape alone.

Air cushions of different areas and different shapes (semi-circles, triangles, and squares) were built, placed under metal plates, and then laminated with GFRP in a total of eight layers. The number of layers may be small, and possibly the air volume in the defects is more significant than in a natural situation. It is important to emphasize that the intention was to demonstrate the ability of intelligent tools to capture the shape of these defects. For this, they should be evident.

Three plates for each defect shape with different sizes were developed for a total of nine working pieces. There were also three repeated shearography acquisitions using step heating with different time duration as loading. Each acquisition series provided 250 frames, which with all possible combinations generated 2250 images per plate. However, the inspection method could not identify the more minor defects; therefore, the related images were discarded from the dataset.

Annotations for an object detection dataset are usually done manually by identifying the regions where each defect is located and assigning a class to it. This process was done, but with a limitation since not all frames had useful fringes pattern. The concept of the useful fringe pattern is related to the quality of an image that contains one or more fringes, i.e. if this fringe pattern has information concerning the defect that originated it. So it was necessary to go frame by frame, identifying the location of the fringe, its class, and its usefulness, i.e., if it was a useful fringe pattern or not. This is a subjective analysis made more straightforward by comparing the fringe wrapped with its unwrapped form. The latter demonstrates more easily whether the defect is visible, even though the former has more information about the fringe regarding the number of features.

The localization of defects in shearography images is not trivial. The images usually have a lot of noise, besides the difficulties associated with the fringe itself, which

goes through pre-processing to de-characterize it until a less subjective interpretation is possible. Classical algorithms have problems interpreting these fringes, which motivates artificial intelligence methods, more specifically deep learning, to detect and classify defect shapes.

One plate of each type was dedicated to training, one for validation, and the last one for testing. This strategy was to prevent, or at least reduce, the phenomenon of data leakage. Each defect fringe pattern quality revealed a strong relationship with its position on the plate, possibly because of the incidence and position of the heat coming from the lamps responsible for loading. In this way, the number of fringes considered useful per defect area is not the same. Consequently, the number of fringes by defect shape is not the same: 6257 useful observations for training the model for square defects, 4318 for semi-circular ones, and 4704 for triangular ones. This differentiation may have occurred because square defects had their dedicated plates, while the other defects shared the same working pieces, increasing the density of defects per plate. This unbalance of quality and quantity of fringes in favor of the square defects greatly influenced the detection result. The mean average precision of this model was 0.772, but the average precision for square defects was 0.901, against 0.729 for semi-circular and 0.681 for triangular defects.

These detection results depend on the score and Intersection-Over-Union thresholds settings, both in training and testing. There was low confidence in detecting the triangular and semi-circular defects, which can be mitigated by adjusting the mentioned parameters. However, this change resulted in a large number of false positives, harming the mean average precision.

Each frame had to be individually analyzed to evaluate the usefulness of the fringe and added to the training set, even knowing its position a priori. This classification a laborious task if done manually and with a high error rate when a supervised algorithm is used. High error rates are not acceptable since the result of this automatic evaluation is used to build the detector dataset. An unsupervised pipeline called AUTO-UFP was developed to solve this issue. It takes the wrapped version of the fringe image and calculates the amount of noise, contrast, and phase jumps. The fringe pattern is useful if it passes these three evaluations, considering arbitrary and experimentally obtained thresholds. Faced with supervised methods, the need for only a few images for parameter setting becomes a great advantage. However, these images should be expressive of good and bad cases, requiring knowledge about the defect under study.

The use of AUTO-UFP generated a dataset with 6292 useful observations for square defects, 3853 for semi-circular defects, and 4297 for triangular defects. Again an imbalance in favor of square-shaped defects was present. The quality of the method was evaluated by F-score, where the actual value of a binary classification comes from the manual selection of the fringe. An F-score of 0.91 was obtained after an exhaustive

grid search, which proved to be not very dependent on the hyperparameters of the method. This independence leads us to believe that the technique can be used without significant problems for other types of defects and other formats.

A new detector was trained on the AUTO-UFP dataset and had a mean average precision of 0.712, 0.06% lower than the model trained on the manually constructed dataset. The difference lay in the inter-class results, where the highest average precision was for the triangular defect class with 0.917. Square defects gave an average precision of 0.667, and semi-circular defects an average precision of 0.551. These average precisions are the opposite of those of the base model. Despite the similar number of observations for the classes, AUTO-UFP kept inadequate examples for the square shape, possibly because it could not deal coherently with many phase jumps present in this type of defect. The characteristics held by the method for triangular defects were beneficial for the detector. Another advantage of using this automatic dataset is that the model's training was much faster: the best epoch was at the very first one.

The AUTO-UFP method is indicated for real-time verification of the usefulness of shearography fringes from adhesion defects. It has not been tested for other types of defects. Its most significant benefit is to be unsupervised. The adjustment of the few parameters can be performed on a small number of images.

Thus, the major contributions of this work are summarized as follows:

- A dataset with shearography images of defects arising from adhesion failure consisting of air pockets wrapped with PTFE and positioned on GFRP laminated plates;
- a method that classifies in an unsupervised way a fringe in its wrapped form into useful or non-useful, and that can be used to facilitate the creation of large datasets;
- a model for detection and classification of fringes from square, semi-circular or triangular defects using wrapped images as input;
- the fact that both methods (OD and AUTO-UFP use fringe wrapped images, dispensing with the phase jump removal step for quantitative defect analysis.

5.1 FUTURE WORKS AND PERSPECTIVE

There is a lot to be done from the results obtained. Initially, as the object detection with deep learning performed well, it is necessary to validate the construction of the defects with actual defects of adhesion failures. This work can be done using a reference method that can deal with deboundings.

Also, a standardization of the creation of these air cushions could benefit further studies, as this defect creation was done in a handcrafted way. Controlled air volume and fixed thickness would be two possibilities.

Once this has been done, it is suggested to create a dataset with shearography images from plates with defects at different rotations to emulate the behavior of a field acquisition. This can be done either by manufacturing new plates or by a robust method of artificially rotating the images from this work. Such a rotation method should not simply rotate the image as a whole but rather each fringe independently while maintaining the influence of this fringe on its periphery.

Related to the detector results, it suggests the development of confirmation rules to remove false positives with a low score without harming those classes that have already been penalized for having observations with low-quality fringes.

In some cases, the classification of the detected defect shape may not be necessary. A simplified model is sufficient if the defect shape does not matter, but only its localization. Models obtained in this work can be adapted for this task by adjusting the dataset for this more straightforward case or changing the detection rules.

As seen during the justification of the model choice, newer detectors can provide better results. Retraining a detection model is also a suggestion for possible work.

No prediction was made about the size of the defects. It is evident the usefulness of a model that receives a bounding box containing a fringe and can estimate the defect area that originated it. Perhaps this model is easy to build, but there must be a dataset with fringes from defects of various areas, covering all possibilities for its training.

There is also the possibility of a framework development that uses spatial-temporal analysis and detection. The idea is include the whole acquisition set to detect the defect.

Finally, there are developments related to the automatic fringe evaluation method (AUTO-UFP). The hyperparameters obtained in this work for this method come from the analysis of only fringes derived from adhesion faults. There is the possibility of extending this analysis to other types of defects, mainly delamination.

About its use, AUTO-UFP leaves gaps between closed intervals of fringe patterns considered useful. It is proposed to create rules to fill these gaps to make the dataset larger and more coherent. Still, this method receives an image with only one fringe, but in a field acquisition, more defects may be present. It is feasible to combine this method with the fringe detector for use during application and not only for dataset creation.

Despite the similar mean average precision results, the model's performance concerning the classes differed from that of the manually built model. Therefore, it is suggested to study the relationship between AUTO-UFP and the model learning parameters.

5.2 PUBLICATIONS

Two articles were published in journals and four in conferences during the doctoral period, three of them as the first author. They were mainly related to machine learning applications and computer vision. Publications on the results of this work, among others, are in production.

5.2.1 Journal articles

Construction of small sets of reference images for feature descriptors fitting and their use in the multiclassification of parts in industry (FRÖHLICH et al., 2020b).

Hybrid model approaches for compensating environmental influences in machine tools using integrated sensors (DAHLEM et al., 2020).

5.2.2 Conference papers

Defect classification in shearography images using convolutional neural networks (FRÖHLICH et al., 2018).

Comparison of principal component analysis and multi-dimensional ensemble empirical mode decomposition for impact damage segmentation in square pulse shearography phase images (FRÖHLICH et al., 2020a).

Impact damage characterization in CFRP plates using PCA and MEEMD decomposition methods in optical lock-in thermography phase images (OLIVEIRA et al., 2019).

Defect inspection in stator windings of induction motors based on convolutional neural networks (OLIVEIRA et al., 2018).

REFERENCES

AEBISCHER, H. A.; WALDNER, S. A simple and effective method for filtering speckle-interferometric phase fringe patterns. **Optics Communications**, v. 162, n. 4, p. 205–210, 1999.

AHMAD, H.; MARKINA, A.A.; POROTNIKOV, M.V.; AHMAD, F. A review of carbon fiber materials in automotive industry. In: 3. v. 971.

ALLEVI, G.; PANDARESE, G.; REVEL, G. M. Improvement of defect size and morphological estimation in shearography inspection by wavelet transform. **Review of Scientific Instruments**, v. 90, n. 10, 2019.

BARRERA, E. S. **Shearografia instantânea multidirecional**. 2018. PhD thesis – Universidade Federal de Santa Catarina.

BARRERA, E. S.; FANTIN, A. V.; WILLEMANN, D. P.; BENEDET, M. E.; ALBERTAZZI GONÇALVES, A. Jr. Multiple-aperture one-shot shearography for simultaneous measurements in three shearing directions. **Optics and Lasers in Engineering**, v. 111, p. 86–92, Dec. 2018.

BARROS, S. De; BANEA, M. D.; BUDHE, S.; SIQUEIRA, C. E. R. De; LOBÃO, B. S. P; SOUZA, L.F.G. Experimental analysis of metal-composite repair of floating offshore units (FPSO). **The Journal of Adhesion**, Taylor & Francis, v. 93, n. 1-2, p. 147–158, 2017.

BEITZEL, S. M.; JENSEN, Eric C.; FRIEDER, Ophir. MAP. In: **Encyclopedia of Database Systems**. Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009. P. 1691–1692.

BĚLSKÝ, P.; KADLEC, M. Capability of non-destructive techniques in evaluating damage to composite sandwich structures. **International Journal of Structural Integrity**, v. 10, n. 3, p. 356–370, 2019.

BENCHMARK. **Carbon fiber woven cloth**. 2021. Available from: %5Curl%7Bhttps://www.bmssi.ca/carbon-fiber-woven-cloth%7D. (accessed: 04.18.2021).

BENEDET, M. E.; WILLEMANN, D. P.; FANTIN, A. V.; SCHMITZ, C. R.; SILVA, F. A.; ALBERTAZZI, A. Inspeção com shearografia em corpos de prova tubulares reparados com material compósito - estudo de caso para o dimensionamento de defeitos. **Revista Abendi**, v. 9, n. 95, p. 46–55, Feb. 2020.

BENEDET, M. E.; WILLEMANN, D. P.; FANTIN, A. V.; SILVA, F. A. A.; ALBERTAZZI GONÇALVES JR., A. **Inspeção com shearografia em corpos de provas tubulares reparados com material compósito - estudo de caso para o dimensionamento de defeitos**. [S.l.: s.n.], 2017. P. 1–13. 14 COTEQ - Conferência sobre tecnologia de equipamentos.

BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning Long-Term Dependencies with Gradient Descent is Difficult. **Transactions on Neural Networks**, IEEE Press, v. 5, n. 2, p. 157–166, Mar. 1994.

BERTHELOT, J. **Composite Materials: Mechanical Behavior and Structural Analysis**. 1. ed. [S.l.]: Springer-Verlag New York, 1999. (Mechanical Engineering Series).

BIG, Bank Industrial Group. **Storage Tank Repair Service**. 2021. Available from: %5Curl%7Bhttps://www.banksindustrial.com/storage-tank-repair%7D. (accessed: 04.18.2021).

BLAIN, P.; VANDENRIJT, J.-F.; LANGUY, F.; KIRKOVE, M.; THÉROUX, L.-D.; LEWANDOWSKI, J.; GEORGES, M. Artificial defects in CFRP composite structure for thermography and shearography nondestructive inspection. In: ASUNDI, Anand Krishna (Ed.). **Fifth International Conference on Optical and Photonics Engineering**. [S.l.]: SPIE, 2017. International Society for Optics and Photonics, p. 562–571.

BOCHKOVSKIY, A.; WANG, C.; LIAO, H. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. **ArXiv**, 2020.

BORTOLI, T.; FANTIN, A. V.; BARRERA, E. S.; BENEDET, M. E.; WILLEMANN, D. P.; ALBERTAZZI GONÇALVES JR., A. Diameter quantification of through holes in pipelines hidden by protective layers of composite materials using instantaneous shearography simultaneously in three shearing directions. In: LEHMANN, Peter; OSTEN, Wolfgang; JR., Armando Albertazzi Gonçalves (Eds.). **Optical Measurement**

Systems for Industrial Inspection XI. [S.l.]: SPIE, 2019. International Society for Optics and Photonics, p. 525–532.

BUDHE, S.; BANEA, M.D.; BARROS, S. de; ROHEM, N.R.F. Assessment of failure pressure of a GFRP composite repair system for wall loss defect in metallic pipelines. **Materialwissenschaft und Werkstofftechnik**, v. 49, n. 7, p. 902–911, 2018.

C&CN. **What's fiberglass, and how does the delicate material reinforce thousands of products?** 2021. Available from: %5Curl%7Bhttps://cen.acs.org/materials/inorganic-chemistry/s-fiberglass-does-delicate-material/96/i38%7D. (accessed: 04.18.2021).

CHALMERS, D. W. Experience in design and production of FRP marine structures. **Marine Structures**, v. 4, n. 2, p. 93–115, 1991.

CHANG, C.; SRINIVASAN, K.; WANG, W.; GANAPATHY, G. P.; VINCENT, D. R.; DEEPA, N. Quality Assessment of Tire Shearography Images via Ensemble Hybrid Faster Region-Based ConvNets. **Electronics**, v. 9, n. 1, 2020.

CHOLLET, F. **Deep Learning with Python**. [S.l.]: Manning, Nov. 2017.

DAHLEM, P.; SANDERS, M. P.; FRÖHLICH, H. B.; SCHMITT, R. H. Hybrid model approaches for compensating environmental influences in machine tools using integrated sensors. **at - Automatisierungstechnik**, v. 68, n. 6, p. 465–476, 2020. DOI: doi:10.1515/auto-2020-0007. Available from: https://doi.org/10.1515/auto-2020-0007.

DAI, J.; LI, Y.; HE, K.; SUN, J. R-FCN: Object Detection via Region-based Fully Convolutional Networks. In: LEE, D.; SUGIYAMA, M.; LUXBURG, U.; GUYON, I.; GARNETT, R. (Eds.). **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2016.

DE ANGELIS, G.; MEO, M.; ALMOND, D.P.; PICKERING, S.G.; ANGIONI, S.L. A new technique to detect defect size and depth in composite structures using digital shearography and unconstrained optimization. **NDT and E International**, v. 45, n. 1, p. 91–96, 2012.

DI MARTINO, A. **Spotting anomalies in carbon fibre composites using deep convolutional neural networks**. University of Edinburgh: [s.n.], 2015. Dissertação.

DNVGL-RP-C301. **Design, fabrication, operation and qualification of bonded repair of steel structures**. Geneva, CH, July 2015.

DONOHO, D. L.; JOHNSTONE, I. M. Ideal spatial adaptation by wavelet shrinkage. **Biometrika**, v. 81, n. 3, p. 425–455, Sept. 1994.

EGBO, M.K. A fundamental review on composite materials and some of their applications in biomedical engineering. **Journal of King Saud University - Engineering Sciences**, 2020.

ENGELBRECHT, A. P. **Computational Intelligence: An Introduction**. 2nd. [S.l.]: Wiley Publishing, 2007.

EVERINGHAM, M.; VAN GOOL, L.; WILLIAMS, C. K. I.; WINN, J.; ZISSERMAN, A. **The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results**. [S.l.: s.n.].
http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

FANTIN, A.V.; WILLEMANN, D.P.; VIOTTI, M.; ALBERTAZZI, A. A computational tool to highlight anomalies on shearographic images in optical flaw detection. In: v. 8788.

FARUK, O.; SAIN, M.; TJONG, J. **Lightweight and sustainable materials for automotive applications**. 1. ed. [S.l.]: CRC Press, 2017.

FELIPE, D. S.; BENEDET, M.; FANTIN, A. V.; WILLEMANN, D. P.; MARTINS, F. D. F.; D'ALMEIDA, A. L. F. S.; SOARES, S. D.; ALBERTAZZI GONÇALVES JR., A.; LENZI, A. Procedure for detecting the shape and size of defects on metallic substrates under composite repairs using shearography. **Applied Optics**, OSA, v. 59, n. 27, p. 8089–8097, Sept. 2020.

FRÖHLICH, H. B.; DE OLIVEIRA, B. C. F.; BARRERA, E. S.; DE BORTOLI, T. J.; ALBERTAZZI, A. Comparison of principal component analysis and multi-dimensional ensemble empirical mode decomposition for impact damage segmentation in square pulse shearography phase images. In: TESCHER, Andrew G.; EBRAHIMI, Touradj (Eds.). **Applications of Digital Image Processing XLIII**. [S.l.]: SPIE, 2020a. International Society for Optics and Photonics, p. 283–291. DOI: 10.1117/12.2566634.

FRÖHLICH, H. B.; FANTIN, A. V.; DE OLIVEIRA, B. C. F.; WILLEMANN, D. P.; IERVOLINO, L. A.; BENEDET, M. E.; GONÇALVES JUNIOR, A. A. Defect

classification in shearography images using convolutional neural networks. In: 2018 International Joint Conference on Neural Networks (IJCNN). [S.l.: s.n.], 2018. P. 1–7.

FRÖHLICH, H. B.; GROZMANI, N.; WOLFSCHLÄGER, D.; ALBERTAZZI, A.; SCHMITT, R. H. Construction of small sets of reference images for feature descriptors fitting and their use in the multiclassification of parts in industry. **The International Journal of Advanced Manufacturing Technology**, v. 108, p. 105–116, 2020b. DOI: doi:10.1515/auto-2020-0007. Available from: https://doi.org/10.1515/auto-2020-0007.

GÉRON, A. **Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems**. Sebastopol, CA: O'Reilly Media, 2017.

GHOBADI, A. Common type of damages in composites and their inspections. **World Journal of Mechanics**, n. 9, p. 24–33, 2017.

GIRSHICK, R. **Fast R-CNN**. [S.l.: s.n.], 2015.

GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. **Rich feature hierarchies for accurate object detection and semantic segmentation**. [S.l.: s.n.], 2014.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. http://www.deeplearningbook.org.

HAYKIN, S. S. **Neural networks and learning machines**. Third. Upper Saddle River, NJ: Pearson Education, 2009.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.: s.n.], 2016. P. 770–778.

HECHT, E. **Optics**. [S.l.]: Pearson Education, Incorporated, 2017.

HERRÁEZ, M. A.; BURTON, D. R.; LALOR, M. J.; GDEISAT, M. A. Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. **Applied Optics**, OSA, v. 41, n. 35, p. 7437–7444, 2002.

HESLEHURST, R. Defects and Damage in Composite Materials and Structures. In: [s.l.]: CRC Press, Inc., 2014.

HUNG, Y. Y. Shearography: A New Optical Method For Strain Measurement And Nondestructive Testing. **Optical Engineering**, SPIE, v. 21, n. 3, p. 391–395, 1982.

IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: PROCEEDINGS of the 32nd International Conference on International Conference on Machine Learning - Volume 37. Lille, France: JMLR.org, 2015. (ICML'15), p. 448–456.

In: IRVING, Philip; SOUTIS, Constantinos (Eds.). **Polymer Composites in the Aerospace Industry (Second Edition)**. Second Edition. [S.l.]: Woodhead Publishing, 2020. (Woodhead Publishing Series in Composites Science and Engineering).

JACCARD, P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. **Bulletin del la Société Vaudoise des Sciences Naturelles**, v. 37, p. 547–579, 1901.

JIANG, X.; HADID, A.; PANG, Y.; GRANGER, E.; FENG, X. **Deep Learning in Object Detection and Recognition**. 1st. [S.l.]: Springer Singapore, 2019.

JIAO, L.; ZHANG, F.; LIU, F.; YANG, S.; LI, L.; FENG, Z.; QU, R. A Survey of Deep Learning-Based Object Detection. **IEEE Access**, v. 7, p. 128837–128868, 2019.

KARAMI, M. Review of Corrosion Role in Gas Pipeline and Some Methods for Preventing It. **Journal of Pressure Vessel Technology**, v. 134, n. 5, Sept. 2012. DOI: 10.1115/1.4006124.

KARBHARI, V. M. **Non-destructive Evaluation (nde) of Polymer Matrix Composites: Techniques and Applications**. [S.l.]: Woodhead Pub., 2013. (Woodhead publishing series in composites science and engineering).

KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization. In: BENGIO, Yoshua; LECUN, Yann (Eds.). **3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings**. [S.l.: s.n.], 2015.

KRISHAN, K. C. **Composite Materials: Science and Engineering**. 3. ed. [S.l.]: Springer-Verlag New York, 2012.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. ImageNet Classification with Deep Convolutional Neural Networks. In: PEREIRA, F.; BURGES, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. (Eds.). **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2012.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. **Nature**, v. 521, p. 436–444, May 2015.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-Based Learning Applied to Document Recognition. In: 11. PROCEEDINGS of the IEEE. [S.l.: s.n.], 1998. P. 2278–2324.

LECUN, Y.; KAVUKVUOGLU, K.; FARABET, C. Convolutional Networks and Applications in Vision. In: PROC. International Symposium on Circuits and Systems (ISCAS'10). [S.l.]: IEEE, 2010.

LIN, T.; DOLLÁR, P.; GIRSHICK, R.; HE, K.; HARIHARAN, B.; BELONGIE, S. Feature Pyramid Networks for Object Detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.: s.n.], 2017. P. 936–944.

LIN, T.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLÁR, P. Focal Loss for Dense Object Detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 42, n. 2, p. 318–327, 2020.

LIN, T.i; GOYAL, P.; GIRSHICK, R. B.; HE, K.; DOLLÁR, P. Focal Loss for Dense Object Detection. **CoRR**, abs/1708.02002, 2017.

LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S.; FU, C.; BERG, A. C. **SSD: Single Shot MultiBox Detector**. [S.l.: s.n.], 2015.

LIU, Z.; GAO, J.; XIE, H.; WALLACE, P. NDT capability of digital shearography for different materials. **Optics and Lasers in Engineering**, v. 49, n. 12, p. 1462–1469, 2011.

MA, Q.; REJAB, R.; ZHANG, B.; MERZUKI, M.; MANOJ KUMAR, N. Wireless technology applied in 3-axis filament winding machine control system using MIT app

inventor. **IOP Conference Series Materials Science and Engineering**, v. 469, Jan. 2019.

MACEDO, F.J.; BENEDET, M.E.; FANTIN, A.V.; WILLEMANN, D.P.; SILVA, F.A.A. da; ALBERTAZZI, A. Inspection of defects of composite materials in inner cylindrical surfaces using endoscopic shearography. **Optics and Lasers in Engineering**, v. 104, p. 100–108, 2018.

MALLICK, P. K. **Composites Engineering Handbook**. Boca Raton: CRC Press, 1997.

MARINUCCI, G. **Materiais Compósitos Poliméricos**. 1st. São Paulo: Artliber Editora LTDA, 2011.

MARTINEZ, J.L.; CYRINO, J.C.R.; VAZ, M.A.; HERNÁNDEZ, I.D.; PERRUT, V.A. Composite patch repair of damaged tubular members from flare boom structures subjected to compressive loads. **Composite Structures**, v. 257, 2021.

MARUSCHAK, P.; PRENTKOVSKIS, O.; BISHCHAK, R. Defectiveness of external and internal surfaces of the main oil and gas pipelines after long-term operation. **Journal of Civil Engineering and Management**, v. 22, p. 279–286, Feb. 2016.

MITCHELL, T. M. **Machine Learning**. 1. ed. USA: McGraw-Hill, Inc., 1997.

NEUBECK, A.; VAN GOOL, L. Efficient Non-Maximum Suppression. In: 18TH International Conference on Pattern Recognition (ICPR'06). [S.l.: s.n.], 2006. P. 850–855.

NOCEDAL, J.; WRIGHT, S. J. **Numerical Optimization**. 2nd. New York, NY, USA: Springer, 2006.

OLIVEIRA, B. C. F. de; FRÖHLICH, H. B.; BARRERA, E. S.; BALDO, C. R.; JR., A. Albertazzi G.; SCHMITT, R. H. Impact damage characterization in CFRP plates using PCA and MEEMD decomposition methods in optical lock-in thermography phase images. In: LEHMANN, Peter; OSTEN, Wolfgang; JR., Armando Albertazzi Gonçalves (Eds.). **Optical Measurement Systems for Industrial Inspection XI**. [S.l.]: SPIE, 2019. International Society for Optics and Photonics, p. 674–684.

OLIVEIRA, B. C. F.; SEIBERT, A. A.; FRÖHLICH, H. B.; COSTA, L. Ro. C.; LOPES, L. B.; IERVOLINO, L. A.; DEMAY, M. B.; FLESCH, R. C. C. Defect inspection in stator windings of induction motors based on convolutional neural networks. In: 2018 13th IEEE International Conference on Industry Applications (INDUSCON). [S.l.: s.n.], 2018. P. 1143–1149.

PADILLA, R.; NETTO, S. L.; DA SILVA, E. A. B. A Survey on Performance Metrics for Object-Detection Algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). [S.l.: s.n.], 2020. P. 237–242.

PAVLOPOULOU, S.; GRAMMATIKOS, S.; KORDATOS, E.; WORDEN, K.; PAIPETIS, A.; MATIKAS, T.; SOUTIS, C. Continuous debonding monitoring of a patch repaired helicopter stabilizer: Damage assessment and analysis. **Composite Structures**, v. 127, p. 231–244, Sept. 2015.

RAVISHANKAR, B.; NAYAK, S. K.; KADER, M. A. Hybrid composites for automotive applications – A review. **Journal of Reinforced Plastics and Composites**, v. 38, n. 18, p. 835–845, 2019.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You Only Look Once: Unified, Real-Time Object Detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.: s.n.], 2016. P. 779–788.

REDMON, J.; FARHADI, A. YOLOv3: An Incremental Improvement. **ArXiv**, 2018.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: CORTES, C.; LAWRENCE, N.; LEE, D.; SUGIYAMA, M.; GARNETT, R. (Eds.). **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2015.

REVEL, G. M.; PANDARESE, G.; ALLEVI, G. Quantitative defect size estimation in shearography inspection by wavelet transform and shear correction. In: 2017 IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace). [S.l.: s.n.], June 2017. P. 535–540.

ROHEM, N.R.F.; PACHECO, L.J.; BUDHE, S.; BANEA, M.D.; SAMPAIO, E.M.; BARROS, S. de. Development and qualification of a new polymeric matrix laminated composite for pipe repair. **Composite Structures**, v. 152, p. 737–745, 2016.

RUBINO, F.; NISTICÒ, A.; TUCCI, F.; CARLONE, P. Marine Application of Fiber Reinforced Composites: A Review. **Journal of Marine Science and Engineering**, v. 8, n. 1, 2020.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning Representations by Back-Propagating Errors. In: NEUROCOMPUTING: Foundations of Research. Cambridge, MA, USA: MIT Press, 1988. P. 696–699.

RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision (IJCV)**, v. 115, n. 3, p. 211–252, 2015.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 3rd. USA: Prentice Hall Press, 2009.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, p. 1627–1639, 1964.

SCHAFER, R. W. What Is a Savitzky-Golay Filter? [Lecture Notes]. **IEEE Signal Processing Magazine**, v. 28, n. 4, p. 111–117, 2011.

SEICA, M.V.; PACKER, J.A.; GUEVARA RAMIREZ, P.; BELL, S.A.H.; ZHAO, X.-L. Rehabilitation of tubular members with carbon reinforced polymers. In: p. 365–373.

SILVA, F. A. A. da. **Colar de múltiplos sensores de Shearografia para inspeção de reparos e uniões de materiais compósitos**. Florianópolis, Brazil: Programa de pós-graduação em engenharia mecânica - Universidade Federal de Santa Catarina, 2017. P. 168. Dissertação de mestrado.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: INTERNATIONAL Conference on Learning Representations. [S.l.: s.n.], 2015.

SMITH, R. A. Composite defects and their detection. Article 6.36.5.3. In: UNESCO Encyclopaedia of Life Support Systems. [S.l.]: UNESCO, 2002.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **The Journal of Machine Learning Research**, v. 15, n. 1, p. 1929–1958, Jan. 2014.

STEVENS, E.; ANTIGA, L.; VIEHMANN, T. **Deep Learning with Pytorch**. Third. Shelter Island: Manning, 2020.

SUSSILLO, D.; ABBOTT, L. F. **Random Walk Initialization for Training Very Deep Feedforward Networks**. [S.l.: s.n.], 2015.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. **Going Deeper with Convolutions**. [S.l.: s.n.], 2014. arXiv: `1409.4842 [cs.CV]`.

TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks**. [S.l.: s.n.], 2020.

UNTERSUCHUNGEN zu dynamischen neuronalen Netzen. PhD thesis – Technische Universität München.

VAN DER WALT, S.; SCHÖNBERGER, J. L.; NUNEZ-IGLESIAS, J.; BOULOGNE, F.; WARNER, J. D.; YAGER, N.; GOUILLART, E.; YU, T. scikit-image: image processing in Python. **PeerJ**, PeerJ Inc., v. 2, e453, 2014.

VANDENRIJT, J-F.; XIONG, H.; LEQUESNE, C.; BLAIN, P.; GEORGES, M. Shearography inspection of monolithic CFRP composites: finite element modeling approach for assessing an adequate strategy of artificial defects representing delamination. In: LEHMANN, Peter; OSTEN, Wolfgang; JR., Armando Albertazzi Gonçalves (Eds.). **Optical Measurement Systems for Industrial Inspection XI**. [S.l.]: SPIE, 2019. International Society for Optics and Photonics, p. 107–113.

VANDENRIJT, J.-F.; GEORGES, M. Automated defect detection algorithm applied to shearography in composites. In: p. 237–240.

VIRTANEN, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. **Nature Methods**, v. 17, p. 261–272, 2020.

VISIN, F.; KASTNER, K.; CHO, K.; MATTEUCCI, M.; COURVILLE, A. C.; BENGIO, Y. ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks. **CoRR**, abs/1505.00393, 2015.

WANG, B.; ZHONG, S.; LEE, T.; FANCEY, K. S.; MI, J. Non-destructive testing and evaluation of composite materials/structures: A state-of-the-art review. **Advances in Mechanical Engineering**, v. 12, n. 4, 2020.

YANG, L.; XIE, X. **Digital Shearography: New Developments and applications**. Bellingham: SPIE Press, 2016.

YE, Y.; MA, K.; ZHOU, H.; AROLA, D.; ZHANG, D. An automated shearography system for cylindrical surface inspection. **Measurement**, v. 135, p. 400–405, 2019.

ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. **Dive into Deep Learning**. [S.l.: s.n.], 2020. `https://d2l.ai`.