eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

INFORMATIKA FAKULTATEA
FACULTAD DE INFORMÁTICA

Artificial Intelligence for Advanced Manufacturing Quality Control

Fátima Aurora Saiz Álvaro

Donostia - San Sebastián 2022

Supervised by:
Prof. Manuel M. Graña Romay - *Computational Intelligence Group, UPV-EHU*
Dr. Iñigo Barandiaran Martirena - *Vicomtech, Basque Research and Technology Alliance (BRTA)*

Acknowledgements

I would like to thank all the people who have directly or indirectly contributed to this Thesis. I am grateful for the opportunities that Vicomtech offered me to carry out applied research in a wide variety of conditions and to get in touch with brilliant and exceptional researchers, especially my esteemed Thesis supervisor at Vicomtech, Dr. Iñigo Barandiaran. The knowledge you have transmitted to me is invaluable, without your guidance I would not be here today. Thanks also to my university advisor, Dr. Manuel Graña, for his final push, his contributions to some work and his excellent mentoring throughout this doctoral process. Their immense knowledge, patience and expertise have encouraged and supported me throughout my research.

I cannot forget my colleagues at Vicomtech, especially the Cognitive Vision group, who have been a continuous source of inspiration and a continuous learning process for the improvement of the quality of the research. Honestly, more than colleagues, I take with me a great family from whom I will never stop learning and will never tire of admiring.

Infinite thanks to my family, and especially to my parents, because without their education, love and ethics I would not have reached the halfway point. It's all thanks to you, you are my inspiration. Thank you Ibon for your patience and unconditional love. Last but not least, thanks to my friends for being a constant support and a source of energy to face the challenges of doing research at PhD level in a very competitive environment full of daily pressure. Thanks for always being there for me and making each day happiest, I hope to always keep you close.

Partial funding from Elkartek projects KK-2021/00070 and KK-2022/00051 is acknowledged by advisor M Graña.

Fátima Aurora Saiz Álvaro

Abstract

Quality control is a centerpiece of any manufacturing industry, regardless of the industrial sector. This complex and demanding process must be carried out with a high degree of precision and rigour. Intelligent automatic systems have attracted the interest of the scientific community for the fulfilment of this critical and complex task. This Thesis addresses the challenge of AI-based image quality control systems applied to the manufacturing industry, with the aim of improving this field through the use of advanced data acquisition and processing techniques, in order to obtain robust, reliable and optimal systems.

Throughout this Thesis, research contributions are presented in the different stages of the design and development of inspection systems. These contributions begin with the use of complex data acquisition techniques, specifically focused on the capture of specular surfaces that are typical in the manufacturing industry. Other contributions are also presented in advanced processing techniques to maximise the information available in the images, in order to facilitate and improve the learning of neural networks. Further contributions include the application and design of specialised neural networks for defect detection, and finally the integration and validation of these systems in production processes. It was developed in the context of several applied research projects that provided practical feedback on the usefulness of the proposed computational advances, as well as real-life data for experimental validation.

Keywords:
Industrial Manufacturing, Quality Control, Deep Learning, Photometric Stereo, Synthethic Data, In-line Superficial Inspection, Defect Detection.

Contents

List of Figures

CHAPTER 1

Introduction

1.1 Introduction

Manufacturing industry is a driver of employment and economic development in a country. Specifically, in the European economy, manufacturing industry represented 14.5% of EU GDP between 2009 and 2019 [5]. Due to an increasingly globalized and competitive economy, this industry sector faces a diverse set of challenges that forces it to be constantly adapting. This adaptation is possible thanks to innovation, which allows production systems to be flexible in order to meet the changing demands of consumers and thus remain a global benchmark. Thanks to the integration of new technologies, the industry moves towards an advanced manufacturing paradigm.

Advanced manufacturing is an important area for learning and process improvement, based increasingly on R&D, becoming one of Europe's key enabling technologies (KETs) [6]. The use of innovative technologies and methodologies to improve competitiveness in the manufacturing sector is becoming more and more established. The aim of this type of manufacturing is to improve services, quality, the ability to meet market demand trends, and to increase the value of the company, including advances in all aspects of the value chain. To remain a basis and driver for the future, manufacturing industry must lead the technological and ecological transition to Industry 5.0 [7]. This means going a step beyond productivity and efficiency, prioritising the well-being of workers in the production process and sustainability.

For this transition, the European Commission defines the following key technologies: human-centric and human-machine-interaction solutions, bio-inspired technologies and smart materials, real time based digital twins and simulation, cyber safe technologies, Artificial Intelligence and technologies for energy efficiency and trustworthy autonomy.

Nowadays, manufacturing companies no longer just want to identify a problem and take corrective action, but they also want to identify in advace all the problems that could occur, convert their processes towards maximum resource efficiency, and to eliminate the factors that prevent these goals from being achieved. For this reason, they are investing more and more in the application of new technologies, as can be seen since the Industry 4.0 [8] revolution.

The inspection stage in the quality processes in the manufacturing industry is critical to avoid defective final products. The exigence of high standard quality control systems in order to satisfy customer requirements, reduce waste of materials, energy and time, and to improve product safety and quality is increasing dramatically [9]. Improvements in quality control systems must also comply with the new requirements that manufacturing industry is undergoing in terms of reducing lead times under increasing market competitiveness.

However, quality should not only be offered in the final products, it should also be present throughout the entire production process. Techniques such as statistical process control (SPC) are often used in production processes [10]. The use of automated vision systems and Artificial Intelligence (AI) makes possible to inspect each item in the entire production process in real time. The benefits for the organisation are less waste, fewer reworks, fewer rejects, fewer complaints, and fewer returns, resulting in lower costs and higher productivity [11].

Currently, many inspection processes are carried out manually thus only a few samples of each production batch are inspected. The lack of full process automation has severe consequences such as no data traceability, human error caused by fatigue and other factors, lack of data to generate process knowledge, and limited process analysis. The promise of employing automated machine vision and AI systems is that data is collected accurately and systematically. The move of organisations towards integrating AI into their processes is a reality [12]. The impact on work productivity, product personalisation, time saving, and product quality is expected to continue to grow in the coming years.

The aim of this Thesis is to develop machine vision and AI systems that are capable of improving current systems, throughout all stages of the production process. For this purpose, our work is focused on the challenges of automating inspection systems, and more specifically in the challenges from the point of view of Deep Learning techniques [13].

1.1.1 Challenges of automated quality inspection systems in the manufacturing industry

Quality control is a complex process whose success is based on a clear understanding of the function, characteristics and requirements of the product being manufactured. All this knowledge is necessary to design a robust and reliable system that will repeat over time and provide confidence to the human operators.

By endowing a system with the ability to see and understand, via image capturing devices, it can almost instantly determine whether or not a product meets certain pre-defined requirements in terms of functionality or aesthetics. Passing each product item in front of a sensor or a camera, the system can obtain accurate readings of its size, colour, shape or dimensional defects, missing parts, internal or surface defects.

Unlike the human eye, a machine can perform this task uninterruptedly and make decisions immediately, based on the data that it receives. In this way, machine vision quality control of any industrialised area can be applied efficiently, quickly and extremely cost-effectively.

Machine vision based visual quality inspection systems based on intelligent technologies are one of the most cost-effective solutions available to an automated industry [14]. They eliminate subjectivity in inspection, compared with manual or visual inspection. All

products follow the same standards with the same quality, removing the uncertainty of the human factor from the equation. They also provide adaptability and flexibility as these systems are mutable, allowing them to adapt to different products on different or the same production line. This means that the same equipment can be reconfigured to work in another area without major complications. Another advantage is that quality control is performed in a non-invasive way. This is an important feature, as vision systems are minimally invasive techniques, allowing for cleaner and safer inspections.

However this task is challenging, as it involves a number of time-consuming and costly steps. The design, development and deployment of automated systems are usually processes that are not carried out quickly, and where the involvement of the company is mandatory for the definition of requirements, the formation about the production process, the generation of the database, and the final validation of the system. Throughout this process different challenges often appear, such as: the system must be robust in hostile environments, the total availability and repeatability of the system must be maximised, the system must generalise with different references or finding an acquisition system for complex materials and geometries, among others.

On the positive side, once this phase is over, companies will be ready to meet the demands of new customers, who are usually looking for more customised products and new services. At the same time, by becoming smarter, factories will also become more operationally and energy efficient [15] [16]. Furthermore, the industrial sector will be able to produce 24/7, to achieve greater precision, to reach excellent quality levels and to reduce production and reaction time to a change or a need.

These systems bring great benefits to overall product quality inspection processes due to their low error margins. These systems remove defective items from the production line, which in general greatly improves the final quality perceived the customers.

This Thesis addresses the challenge of AI-based image quality control systems applied to manufacturing industry, aiming to improve this field through the use of advanced techniques for data acquisition and processing, in order to obtain robust, reliable and optimal systems. This Thesis presents contributions on the use of complex data acquisition techniques, the application and design of specialised neural networks for the defect detection, and the integration and validation of these systems in production processes.

1.2 Publications

This Thesis is presented as a compilation of the following papers in compliance with the rules defined by the UPV/EHU's school of doctorate.

- Saiz, F.A.; Barandiaran, I.; Arbelaiz, A.; Graña, M. Photometric Stereo-Based Defect Detection System for Steel Components Manufacturing Using a Deep Segmentation Network. Sensors 2022, 22, 882. https://doi.org/10.3390/s22030882. **[JCR (2021): 3.847, 5-year: 4.050, Q2]**

- Saiz, F.A.; Alfaro, G.; Barandiaran, I.; Graña, M. Generative Adversarial Networks to Improve the Robustness of Visual Defect Segmentation by Semantic Networks in Manufacturing Components. Appl. Sci. 2021, 11, 6368. https://doi.org/10.3390/app11146368 **[JCR (2021): 2.838, 5-year: 2.921, Q2]**

- Saiz, F.A., & Barandiaran, I. (2020). COVID-19 detection in chest X-ray images using a deep learning approach. Int. J. Interact. Multimedia Artif. Intell., 6 (Regular Issue) (2020), p. 4. http://dx.doi.org/10.9781/ijimai.2020.04.003 **[JCR (2020): 4.936, 5-year: 3.730, Q2]**

Additionally, the following publications are the direct result of the works reported in this Thesis:

- Saiz, F. A., Barandiaran, I.; Arbelaiz, A.; Graña, M. (2022, May). Artificial Intelligence for Advanced Manufacturing Quality Control. In Basque Conference on Cyber Physical Systems and Artificial Intelligence (pp. 111-118). Zenodo.
- Garcia, A., Franco, J., Sáiz, F., Sánchez, J. R., Bruse, J. L. (2022). Containerized edge architecture for manufacturing data analysis in Cyber-Physical Production Systems. Procedia Computer Science, 204, 378–384. doi:10.1016/j.procs.2022.08.046
- Saiz, F.A.; Alfaro, G.; Barandiaran, I. An Inspection and Classification System for Automotive Component Remanufacturing Industry Based on Ensemble Learning. Information 2021, 12, 489. https://doi.org/10.3390/info12120489. **[JCR (2021): Q3]**
- Saiz, F. A., Alfaro, G., Barandiaran, I., Garcia, S., Carretero, M. P., & Graña, M. (2021). Synthetic Data Set Generation for the Evaluation of Image Acquisition Strategies Applied to Deep Learning Based Industrial Component Inspection Systems.
- Saiz, F.A., & Barandiaran, I. (2020). COVID-19 detection in chest X-ray images using a deep learning approach. Int. J. Interact. Multimedia Artif. Intell., 6 (Regular Issue) (2020), p. 4. http://dx.doi.org/10.9781/ijimai.2020.04.003 **[JCR (2020): 4.936, 5-year: 3.730, Q2]**
- Saiz, F. A., Serrano, I., Barandiarán, I., & Sánchez, J. R. (2018, September). A robust and fast deep learning-based method for defect classification in steel surfaces. In 2018 International Conference on Intelligent Systems (IS) (pp. 455-460). IEEE.

The following publications are an indirect result of the works carried out during this Thesis:

- Presa, S., Saiz, F.A., Barandiaran, I. (2022). A Fast Deep Learning Based Approach for Unsupervised Anomaly Detection in 3D Data. In 2022 7th International conference on Frontiers of Signal Processing. Proceedings pending to be published.
- Ojer, M.; Alvarez, H.; Serrano, I.; Saiz, F.A.; Barandiaran, I.; Aguinaga, D.; Querejeta, L.; Alejandro, D. Projection-Based Augmented Reality Assistance for Manual Electronic Component Assembly Processes. Appl. Sci. 2020, 10, 796. https://doi.org/10.3390/app10030796 **[JCR (2020): 2.679, 5-year: 2.736, Q2]**
- Ojer, M., Serrano, I., Saiz, F. et al. Real-time automatic optical system to assist operators in the assembling of electronic components. Int J Adv Manuf Technol 107, 2261–2275 (2020). https://doi.org/10.1007/s00170-020-05125-z **[JCR (2020): 3.226, 5-year: 3.320, Q2]**

CHAPTER 2

Foundations and Context

This chapter presents the theoretical concepts underlying the research developed in this Thesis. First, we provide an introduction to quality control in industrial environments, its importance and challenges that intelligent systems must overcome. Subsequently, we describe typical inspection pipeline, discussing in some detail all the theoretical concepts of each stage, from the design of the acquisition system to the final validation. Finally, the metrics that were used throughout the experiments for the evaluation of the trained models are explained.

2.1 The relevance of quality control

Quality control is a the key process that has a direct impact on the final manufactured product. Quality control is a testing process that attempts to measure and guarantee product compliance within the final specifications. In order to ensure the quality during the whole manufacturing process, a series of tests are applied at each stage of the process.

The aim of a quality control is to help any type of company or industry to meet the demand for better products, ensuring zero defects and compliance with the ISO 9000 standards that regulate quality [17].

The implementation of a quality control system in a production environment requires planning and effort to ensure that it is done correctly. In this regard, there is a wide variety of quality control techniques that can be applied in industrial processes, such as:

- **Six Sigma:** The Six Sigma [18] approach is the customer satisfaction method developed at Motorola. It is based on two premises: (1) A product is built in the shortest time and at the lowest cost if no mistake is made in the process; and (2) if no defect can be found anywhere in the process of building a product for the customer, then the customer probably will not find one either.
- **Statistical Process Control (SPC):** This technique [19] is based on tracking production metrics by monitoring quality at each step. SPC focuses on optimising continuous improvement by using statistical tools to analyse data, make inferences about process behaviour and then make appropriate decisions.

The basic assumption of SPC is that all processes are subject to variation. Variation measures how the data is distributed around the central tendency. This allows a better identification of problems and their resolution at an early stage.

- **Total Production Maintenance (TPM):** Total Production Maintenance is a working method originated in the 1970s. It ensures and improves production integrity through the use of machine, equipment, process and material innovation [20]. By using a proactive approach, problems are identified at an early stage and eliminated as soon as possible. It is a well-established way of improving manufacturing production systems and works well as part of an overall quality control strategy in most environments. The ultimate goal is to achieve the so-called 3 zeros: Zero defects, Zero breakdowns and Zero accidents.

- **Total quality management (TQM):** TQM implies the presence of quality control in all areas: in products, designs, procedures and systems [21]. A fundamental factor considering the importance that quality has achieved as a means of guaranteeing customer satisfaction and reducing costs. Their principles are: Customer-focused, Total employee involvement, process-centered, Integrated system, strategic and systematic approach, continual improvement, Fact-based decision making and Communications.

There are different areas where actions can be taken to improve the quality of total production. This Thesis focuses on image based product quality inspection, and more precisely, on defect detection at early stages of manufacturing.

Nowadays, the market demands quality products and the industry has been updating its procedures and methods in order to meet this requirement. Traditionally, human inspection was the method on which quality was based, discarding defective products before being sent to the market. It is well known that this procedure, despite being exhaustive, does not fully comply with the objective, mainly due to human operator fatigue and uncertainty, resulting products being accepted or rejected incorrectly. We must also note that each item rejected as defective has consumed company resources, labour, raw material, energy, etc., resulting in an effective increase in the final cost of the product.

Artificial Intelligence has taken off globally in recent years, being applied in several contexts and sectors and the manufacturing industry and quality control specially is not an exception. During the process of analysing and collecting data from companies, AI can recognise regular patterns not perceivable to the human eye.

AI and machine learning in particular become possibly the most important factor in complex contexts such as quality control. Using AI is required when there are numerous unclear requirements, a multitude of variables or where conditions change over time. These methods will be introduced in the following chapters.

2.2 Key points for solving the challenges of automated inspection systems using Artificial Intelligence techniques

AI in industry refers to predictive analytics methodologies, artificial vision systems or decision support systems based on intelligent systems [22]. Machine vision intelligent

systems have become an essential element of technology, reporting highly accurate results in multiple tasks applied to industrial manufacturing quality control processes, such as surface inspection or the automatic handling of components in robotics [23]. Thanks to the advances in AI, manufacturing industry can benefit from the improvements in automation, productivity and safety that these systems offer.

Among many existing techniques used in Machine Learning, there is a branch that is currently being researched and applied, known as is Deep Learning. Deep Learning consists of an automatic learning process of computational architectures such as artificial neural networks, originally developed trying to simulate the way in which the human brain works. These architectures process data, extract features, and learn patterns to be used afterwards during the decision-making process.

In a conventional machine vision method, a developer must define and verify individual features manually. However, with deep learning, learning algorithms are used to automatically find and extract unique patterns in order to identify and differentiate between classes.

Many industrial applications are using these techniques in different fields. For example, in the inspection process carried out in the electronics industry, it is possible to detect different non-conceivable product defects effectively and automatically [24]. Another industrial application is the localisation of paint defects very difficult for the human eye but are able to be identified by these learning systems [25].

Within industrial quality control there are different inspection tasks ranging from simple to more complex ones, mostly being characterised by the feasibility to define the features that describes the problem. Simple tasks such as the correct placement of components or the presence or absence of components are easily quantifiable. However, not all the parameters that determine the proper quality of a product are easy to determine. Moreover, due to the diversity of components that can be handled on the same production line and their different characteristics, the task of identifying and classifying them is complicated.

It is important to clarify the relationship between Machine Learning and Deep learning. Machine Learning is a subset of Artificial Intelligence that allows building intelligent machines that learn from a set of data. Deep Learning is a subset of Machine Learning, which uses Machine Learning algorithms to train deep neural networks achieving greater accuracy in cases where conventional Machine Learning fails to perform adequately.

Conventional Machine Learning methods are limited by the need to process the original data from its natural form to a form of representation that is appropriate for the solution of each specific problem [26]. In real industrial environments this task is even more complex as the data comes from different sources of information and with very particular characteristics. Deep Learning techniques change this paradigm completely as the network itself is responsible, from the raw data, for estimating the internal representation of the data by means of optimisation techniques, simultaneously with the learning of the classifier or inference system. This estimation is carried out in a single process and through the use of layers specifically designed to learn filters, which allow the signal to be represented in a versatile, hierarchical and a derivable way.

Deep Learning is most commonly applied to the classification of more complex objects. Deep learning-based architectures vary according to the type of data (numerical, point clouds, images, etc.) or the purpose of use (classify, detect, segment, etc.). Focusing their

use on images, use of convolutional neural networks (CNNs) is expanding due to their adaptation to different applications. However, obtaining quality results depends on proper supervision, especially in the training phase. Although these models employ learning, they depend on many factors such as the quality and size of the dataset and the chosen architecture, where the knowledge of a specialist is important [27].

However, one of the main problems in training these systems is the requirement of a large sample of images to recognise the object [28]. The size or the number of samples varies depending on the characteristics of the object, since the sample must represent as much as possible all the scenarios in which it can be found. This requirement is compounded by the difficulty of sometimes not having these data available, and even if they were available, the capturing process takes considerable time [29]. This is why new trends are emerging trying to mitigate or avoid this problem, such as anomaly detection or synthetic data generation [30] [31]. Throughout this Thesis, the research and developments carried out in this direction are shown.

2.2.1 Automatized quality control systems and their challenges

Visual detection technology based on image processing has been widely used in various fields, such as medicine [32], manufacturing [33], textile [34] and automotive industries [35] for its unique advantages of accuracy and speed. Early defect detection methods can be grouped into two general categories: contact and non-contact detection. Contact detection receives information through direct contact with the surface of the component using a sensor that interacts with it. However, non-contact inspection relies on technologies based on photoelectricity and electromagnetism to obtain information about the surface parameters of the sample without contacting it.

Visual detection based on image sensors is one of the most widely used non-contact detection methods because it is an effective combination of the high speed achieved by contact detection methods and the flexibility of non-contact detection methods [36] [37]. The key feature is the simplicity of implementation of these systems, as they can be deployed using an image acquisition system, a computer or a processing unit and an image processing algorithm. The validation of these technologies is supported by a large amount of related publications [38] [39] [40], that is continuously growing over the last decade. This upward trend is based on the improvements in computer technology, the development of more advanced sensors, and developments in intelligent image processing algorithms.

Although the solution seems straightforward, real-time surface inspection in industry faces a number of challenges. Some of the challenges that the systems must overcome are mentioned below and will be developed in depth throughout the Thesis:

- **Defect catalogue variability and multiple surfaces:** The variability of surface defects is wide, with the location of defects varying depending on the component and the processes applied to it. Some defect categories may have many intra-class differences, and other defect categories may be similar to each other, which complicates system development.
- **Total machine availability and accuracy:** A large number of production lines are in operation 24 hours a day, 7 days a week, so the inspection system must meet a strict KPI of machine availability. In addition, the algorithms to be developed must

meet high demands in terms of accuracy, as they must outperform human detection capabilities. Any system failure, whether of hardware or software, wastes money and time, and negatively impacts the reliability of the system for the company and the operators.

- **High production rate:** Generally, the production rate in the manufacturing industry is high. For example, the production of rolled steel components can be 20m/s. This is an important factor to take into account in all components of the inspection system, as the result must be given in a very limited time to take corrective actions and avoid scrapping as much material as possible.
- **The complexities of deployment within the manufacturing line:** Manufacturing lines can be placed in very hostile conditions. Depending on the component to be manufactured, there are different types of adversities such as high temperatures, corrosive particles in suspension, oil, steam, vibrations, etc. These particularities make difficult to install the acquisition systems properly, as they require extra protection for the hardware to last over time. In addition, the protections should not require daily maintenance of the system, as this would invalidate its usefulness on the production line. In terms of the quality of the data obtained, these factors cause a performance loss that can increase over time between maintenance periods, causing problems for the stability of the algorithms.

2.3 Inspection systems pipeline

Automatic inspection systems combine hardware and software components to perform feature recognition based on image acquisition and processing. Generally, these systems have a common pipeline that can be divided into several stages, such as shown in Figure 2.1, which are detailed in more detail below.

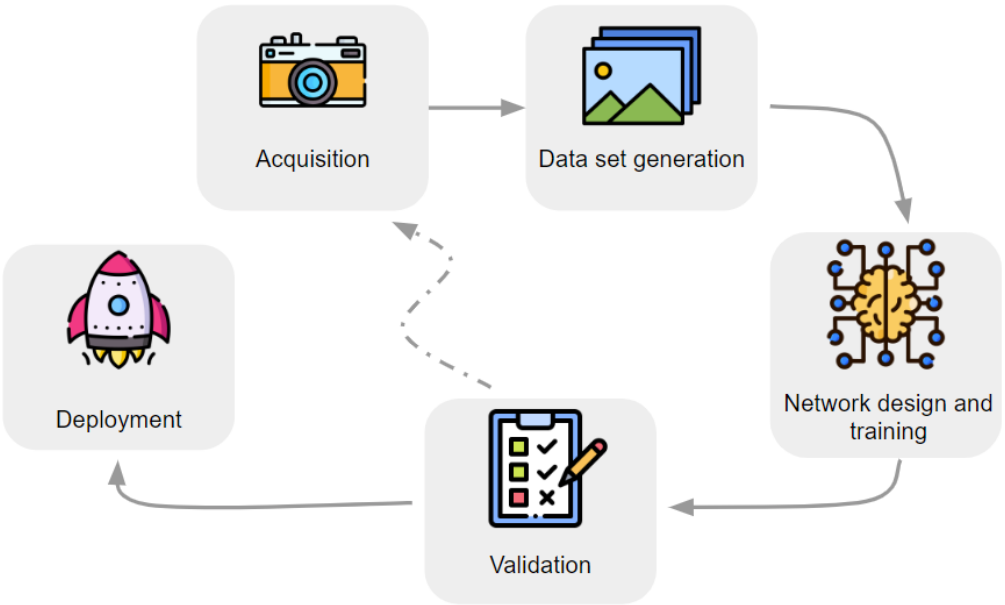


Figure 2.1: Inspection system pipeline

2.3.1 First step: Image Acquisition

The image acquisition system has the function of transforming the optical scene into numerical data that will be received by the processing device. The correct design of this system has an important impact on the acquisition of proper data, and therefore on the correct final system performance.

The sensors that can be used in this type of system are multiple and are chosen according to the defects to be detected and the characteristics of the production line. When designing an illumination setup for an inspection system, the task that the inspection system must overcome must be considered: look for a defect in a part, measure the dimensions of a part, or simply determine the presence or absence of a feature. Then, it must be defined the optical components that compose the inspection system. If the defects to be detected are of dimensional nature, the most suitable sensors are 3D sensors, however for aesthetic defects, which are the types of defects that are the focus of this Thesis, the most suitable sensors are 2D cameras.

2.3.1.1 A brief perspective on 3D acquisition

While 2D vision techniques effectively solve many problems related with dimensional inspection, they are severely limited by not being able to acquire all part geometry. 3D acquisition techniques can be classified as active or passive [41] [42]. Passive methods, which are rarely used in industry, use object reflection and scene illumination to derive shape information, so an active device is not required. In the active form, which is more robust, appropriate and specific sources of energy are used to obtain the information.

A distinction is also made between direct and indirect measurements [43]. Direct techniques result in range data, e.g. a set of distances between the unknown surface and the range sensor. Indirect measurements are inferred from monocular images and *a priori* knowledge of the object properties.

The most common technique for obtaining three-dimensional information for continuous products is laser optical triangulation [44] [45], based on the perception of positional variation of a laser projection on the object. In order to obtain a complete scan, either the sensor or the object must be moved. In continuous inspection, the movement is provided by the process itself and it is usually reported to the scanner via an encoder. Lasers with different wavelengths are used depending on the environmental conditions, such as 635nm (red) for cold applications, 470nm (blue) for hot component inspection, or 530nm (green) for underwater and some hot applications. As for the cameras, they are based on high-speed matrix cameras usually based on CMOS sensors.

Another 3D acquisition system that is becoming more widespread in the industrial field are those based on structured light [46]. In contrast to laser-based devices, there is no need to move the system or the part as they are able to digitise a bigger area than a laser line at once.

Structured light reconstruction is a technique that provides very good results in 3D scanning of objects with little texture information, such as metallic objects [47] [48]. They are based on the projection of a pattern or a sequence of patterns that univocally determine the deviation between the projection and the pixel positions. Basically, it consists of using one or several cameras and a projector to illuminate the object or scene to be reconstructed

using one or several coded patterns and to analyse the image captured by the camera to obtain, by triangulation, the information of each point of view of the object or surface reconstructed in three-dimensional space.

Two groups of techniques can be distinguished depending on the nature of the patterns used in the projection stage; those using a continuous pattern structure and those using a discrete structure. Discrete coding methods are those that use patterns with a digital profile, such as binary patterns with black and white pixel values, assigning the same value to the region with the same colour. Patterns with continuous structure, however, present a continuous variation in intensity or colour on the two axes.

The accuracy of these scanners depends mainly on the resolution and sharpness of the projector, as well as the scanning distance. As the scanner is moved further away from the object, the area covered by the projector and cameras becomes larger and more detail can be discerned.

2.3.1.2 2D acquisition: sensors, optics and illumination

The definition of a 2D acquisition system is not a simple and quick task, especially in industrial environments. The systems to be installed in the production shop floor must be robust in adverse conditions, reliable and repetitive, so they have to try to control the maximum possible variables existing in the production line. Typically, this system consists of a sensor (or camera), a lens and an illumination assembly [49].

2.3.1.2.1 Camera sensor types

Each of these components must be carefully chosen depending on the application and the type of part being inspected. The typology of image sensors can be classified according to two main criteria: sensor geometry and spectral sensitivity. According to the geometric criterion, there are two types of sensors:

- **Area:** the sensor is a square pixel array so that in each frame a complete image is obtained. The vertical resolution of the image is therefore limited to the size of the sensor. Furthermore, these sensors are only suitable for low acquisition rate processes.

- **Linear:** The sensor has only one line of pixels, so that several frames have to be acquired in order to build a complete image. The vertical resolution is not limited by the sensor, but requires synchronised motion between the camera and the part being captured. As shown in the Figure 2.2, this synchronisation is normally done by encoders on a conveyor belt. This type of sensor can work in processes with high acquisition rate.

2.3.1.2.2 Optics

The camera's optics perform the function of focusing and directing the light onto the sensor. Its characteristics determine the shape of the final image as well as the amount of light the sensor will receive. Generally there are two main kinds of lenses used in industrial applications: conventional lenses and telecentric lenses.

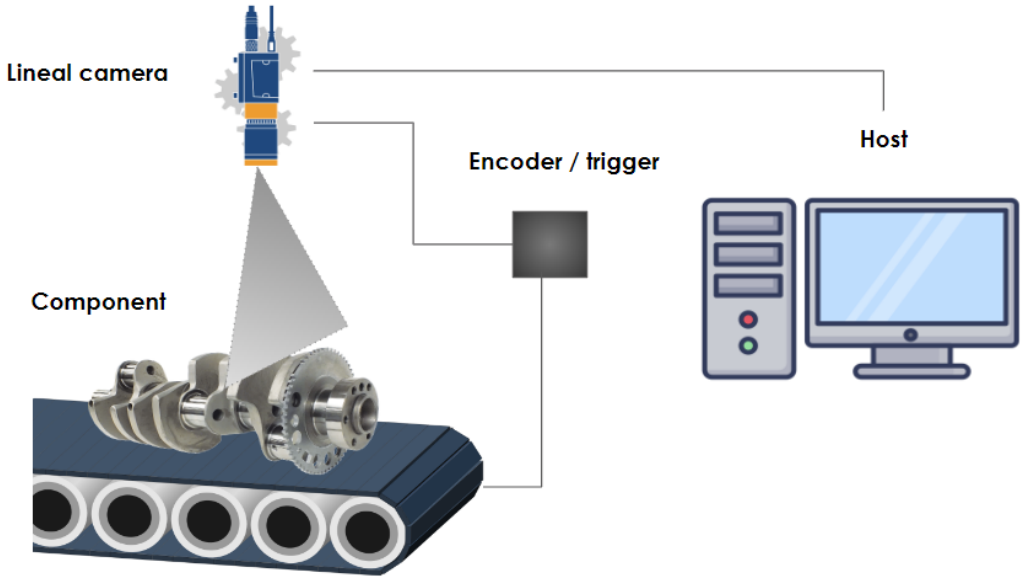


Figure 2.2: Example of typical image acquisition schema using linear cameras.

Conventional lenses have a behaviour that approximates the perspective projection model. In this way, objects have an apparent change in size, appearing smaller when they are further away from the lens. This scaling factor depends directly on the focal length, which also determines the effective working area of the camera.

The main disadvantage of this type of lens is that it is not easy to take measurements from the images directly due to the perspective distortion, without any type of calibration of geometric distortion correction [50].

In order to solve this problem, telecentric lenses achieve a behaviour that approximates a parallel projection model [51]. As can be seen in the Figure 2.3, under this premise the size of the objects in the image does not depend on their distance from the camera, so it is always possible to establish a relationship between pixel size and object size, thus being able to extract very accurate measurements.

However, these lenses have two disadvantages. On the one hand, their depth of field distance is very small, only a few millimetres. On the other hand, they are usually quite large lenses, which can be a handicap in some installations.

As an example, the Figure 2.3 shows a comparison of the image generated by a conventional lens compared to the one generated by a telecentric lens. As can be seen in the case of the telecentric lens, the size of the object is not distorted by its distance from the camera.

The accuracy that can be achieved with a telecentric lens depends on its working area, sensor size, and sensor resolution. With a standard configuration it is possible to achieve accuracies around 10µm.

2.3.1.2.3 Illumination setup

One of the most critical variables in the choice of the acquisition hardware is the ambient lighting and the specularity of the surfaces to be inspected. The surfaces of a component have optical properties that fall into one of three general categories of reflectance: specular,

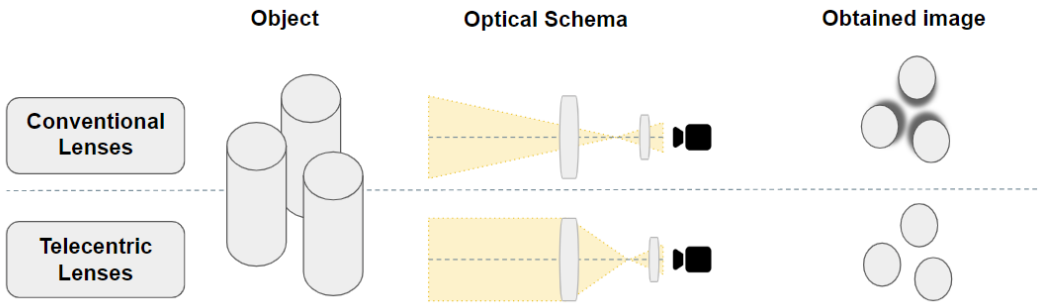


Figure 2.3: Effect of lens distortion on the acquired image.

diffuse or directional, as shown in Figure 2.4. Components or products may have different types of surfaces, so it is important to understand how light interacts with them [52].

Very specular surfaces are mirror-like, flat and highly polished. They reflect light at an angle equal and opposite to the angle of incidence. Diffuse surfaces are rough and have an opaque shine, so they scatter light in all directions, an effect called diffuse reflectance. And finally, directionally reflective surfaces usually contain fine slots that reflect light in direction depending on the angle of incidence.

In addition to the reflective properties of a component, the geometry of the surface and whether it is flat, curved or prismatic must also be taken into account. On a flat surface it is easier to achieve uniform illumination, and the choice of systems is often easier than for other geometries. However, the changing slope of a curved component can often cause a lighting problem that manifests itself as uneven illumination across the surface.

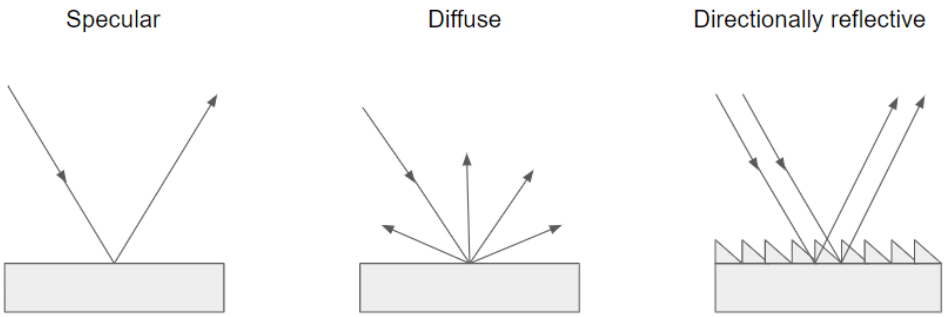


Figure 2.4: Types of light reflections based on surface characteristics.

To ensure that an algorithm is robust and does not suffer from artefacts appearing in the images, such as reflections, the illumination must be controlled as much as possible. For this purpose, as Figure 2.5 shows, there are different illumination schemes [53]:

- **Backlight:** This type of illumination is used to generate a high contrast between the background and the objects to be inspected. In this configuration the illuminator is placed behind the object in the direction of the camera as shown in the Figure 2.5. This technique is particularly effective when evaluating contours or measuring holes.

- **Diffuse light:** This technique consists of indirectly illuminating the object from all directions. It is particularly effective for inspecting curved surfaces and highly specular materials. The most common configuration consists of a hemispherical dome that reflects the light towards the object inside the dome.

- **Bright field light:** This technique consists of illuminating the object from the side so that the angle of incidence of the light is within the cone of vision of the camera. Due to its directionality, it is very useful when highlighting topological details of the piece, such as reliefs. However, it presents problems with highly specular surfaces due to the reflection it produces on its surface.

- **Dark field light:** Similar to the previous technique, this method illuminates the object from the side, but with a very low angle of incidence so that the light is reflected outside the cone of vision of the camera. In this way only the diffuse component of the textured areas reaches the camera, preventing the specular component from saturating the image. This technique is particularly effective with highly specular materials, such as metals.

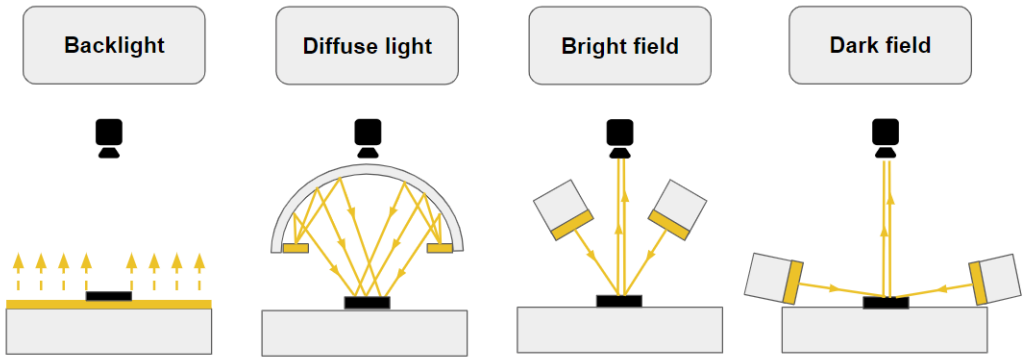


Figure 2.5: Types of light reflections based on surface characteristics.

2.3.1.2.4 Photometric stereo

Due to the difficulty of choosing lighting, these aforementioned illumination schemes serve as a basis for the creation of more a advanced illumination systems, such as photometric stereo, which has special relevance in this Thesis.

Reflectance based shape recovery of non-planar surfaces from several reflectance images obtained under different irradiance sources is a classic task in computer vision [54]. This type of approaches determine the absolute depth of the surfaces by reconstructing the shape of the object under changing illumination conditions such as color or orientation, among others. This problem is called Shape from Shading when just one irradiance image is used for reconstruction process [55]. Photometric stereo methods firstly recover surface orientations and can be combined with an integration method to calculate a height or depth map. Even without a subsequent integration step the surface orientations can be used, for example to determine curvature parameters of object surfaces [56]. To acquire images for photometric stereo the object is consecutively illuminated by several light sources. In one of the applications developed in the gramwork of this Thesis we use four different light

sources from four different orientations around the component to be inspected, as shown in Figure 2.6.

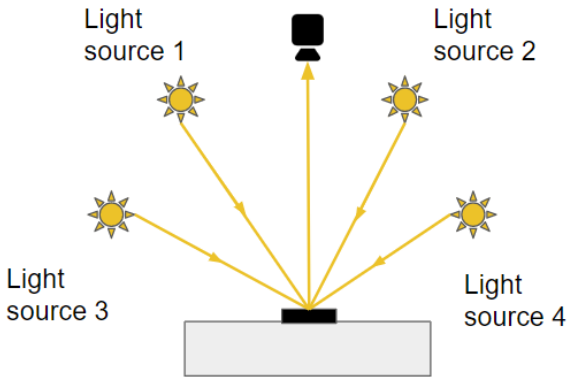


Figure 2.6: Schema of photometric stereo acquisition system with four light sources.

In this settings, we get 5 different photometric images for each acquisition as shown in Figure 2.7, namely:

- Curvature images: Provides the contour lines of the surface topography. May be used for the verification of local defects such as scratches, impact marks, etc.
- Texture images: Provides color or spectral response. They are very suitable for detecting discoloration defects and rust damage.
- Gradient image X: Signal variation in x direction.
- Gradient image Y: Signal variation in y direction.
- Range images: Computed as the image gradient magnitude. It highlights information about the changes in the intensity of the image.

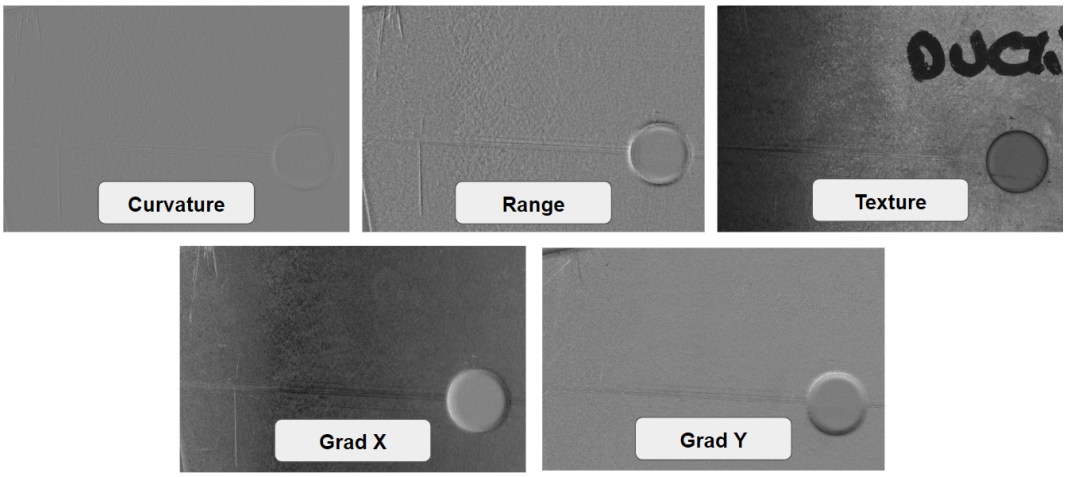


Figure 2.7: An example of obtained images by photometric stereo acquisition.

These images are the result of a system of equations introduced by Woodham [54] formalizing the solution of the problem assuming Lambertian reflectance, and taking over that the distant light sources are known point sources. Given a known vector $\vec{I}$ of i observed intensities, the known matrix of normalised light directions $[\vec{L}] = (L^1, L^2, L^3)^T$, and the reflectivity ρ, the unknown surface normal $\vec{n}$ can be obtained by inverting the following lineal equation:

$$\vec{I} = \rho[\vec{L}]\vec{n} \tag{2.1}$$

If the three illumination vectors L^k do not lie in the same plane, the matrix $[\vec{L}]$ is non-singular and can be inverted, giving the following equation:

$$[\vec{L}]^{-1}\vec{I} = \rho\vec{n} \tag{2.2}$$

As $\vec{n}$ has a unit length, we can estimate the surface normal and the albedo. The problem comes when we have more than three light sources, in this case the illuminations matrix $[\vec{L}]$ would no longer be square, and therefore could not be inverted [57].

When the light sources are more than three, this problem is solved by least squares, using the Moore-Penrose pseudo-inverse [58] in the Equation 2.1, thus obtaining the following solution:

$$[\vec{L}]^T\vec{I} = [\vec{L}]^T[\vec{L}]\rho\vec{n} \tag{2.3}$$

$$\rho\vec{n} = ([\vec{L}]^T[\vec{L}])^{-1}[\vec{L}]^T\vec{I} \tag{2.4}$$

Where $([\vec{L}]^T[\vec{L}])^{-1}[\vec{L}]^T$ is the Moore-Penrose pseudo-inverse. After that, ρ and $\vec{n}$ can be solved as before.

2.3.2 Second step: Dataset generation

Once the appropriate acquisition system design is chosen, the subsequent step is the generation of the dataset. For AI-based system training, the quantity and quality of the data available is decisive for the success or failure of the overall system [59]. This process usually requires three different sets of data: the training set, the validation set and the test set.

The training set is used to teach the model, while the validation set is used to check how well the model has learned in every N iterations. When this process reaches a convergence state, the training is considered to be complete. Then the test set is used to obtain an estimation of how the model will perform on unseen data, so that its performance will be extrapolated to the behaviour of the final system performance.

This data has to represent as far as possible the whole nature of the components to be inspected and their corresponding defects. This task seems straightforward in other areas such as autonomous driving, where data for element classification is abundant, but is not so straightforward in manufacturing industrial environments. Due to the manufacturing rate and the small percentage of defective parts in production, the composition of a robust

and reliable database with variability is often impossible to achieve. To overcome these problems, novel techniques are used to minimise these requirements through the use of advanced data augmentation or synthetic image generation [60].

2.3.2.1 The importance of data and the difficulties in obtaining representative and reliable ones

Every year, new, more complex and more accurate CNNs are designed for image recognition tasks. Typically, these networks increase in width or depth [61], thus gaining representational power. These neural networks learn a hierarchy of visual representations, which were shown to be very effective in computer vision tasks [62] [63] [64]. But learning these representations requires a large amount of data that often cannot be achieved.

For the comparison and evaluation of networks, public datasets are often used, which typically are of the order of hundreds of thousands of images. Some of the best known ones are mentioned below:

- **The Pascal Visual Object Classes (PASCAL VOC)**
  The PASCAL VOC 2012 dataset [65] contains 20 object categories including vehicles, household, animals, and other objects. This dataset is usually used for object detection, semantic segmentation or classification purposes, so each image in this dataset has pixel-level segmentation annotations, bounding box annotations, and object class annotations. The train/val data has 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations.

- **ImageNet**
  ImageNet [66] is an image dataset organised according to the WordNet hierarchy, where there are more than 100,000 categories. ImageNet contains 1000 images of each category, and the quality of each image and its corresponding annotation is controlled by humans. In total it contains approximately 15 million images.

- **COCO dataset**
  MS COCO (Microsoft Common Objects in Context) [67] is a large-scale image dataset containing 328,000 images of 91 object types with a total of 2.5 million labelled instances. It is usually used for object detection, segmentation, and captioning tasks.

These datasets have images of all kinds, which do not resemble in any way any likely scenario in the manufacturing industry. However, one of the most powerful approaches is to use these images in a pre-training stage. In this way, the original model learns visual representations using millions of labelled images, and at a later stage is trained on target tasks to perform better. In this way, the network transfers the acquired knowledge and needs fewer images of the target environment to converge. Transfer learning improves performance in multiple vision tasks, and during this Thesis it was applied in several projects due to the lack of data available for training.

Although transfer learning is a very useful tool for obtaining robust models and filling this lack of data, it is not a universal solution if still there is insufficient data of the target scenario. It is well known that performance increases logarithmically as a function of the amount of data available for training [68]. There is a logarithmic relationship between

performance on vision tasks and the amount of training data used for representation learning.

Finally, the amount and complexity of the data available has to be related to the capacity of the neural networks. For example, in the case of ResNet-50 [69] the gain in COCO object detection is lower (1.87%) compared to when using ResNet-152 (3.87%), that has more layers and more parameters [61].

2.3.2.2 Data augmentation techniques

Data augmentation aims to increase the number of samples in training in order to reduce overfitting and to overcome the problem of lack of data. There are multiple strategies for data augmentation starting from more traditional image operations to more modern and novel approaches such as the use of GANs. In the scheme shown in Figure 2.8, an outline of proposed techniques is shown, whose branches will be explained below.

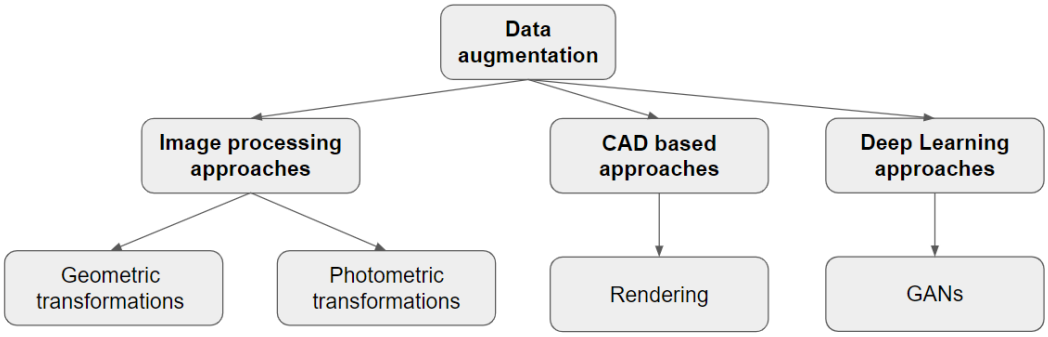


Figure 2.8: Data augmentation techniques schema.

2.3.2.2.1 Image based approaches: Geometric and Photometric transformations

One of the most traditional data enhancement techniques is to apply geometric and photometric transformations to the image. Geometric transformations involve a transformation in the image coordinate system. There are many applicable operations as shown in Figure 2.9, such as translation, rotation, resizing, affine transformation, or perspective transformation.

Photometric transformations alter the RGB (red, green and blue) channels by shifting each pixel value (r, g, b) to new values (r', g', b') according to predefined rules; these transformations adjust the lighting and colour of the image but leave the geometry unchanged. For example, these transformations can highlight edges and contours, can influence colour such as discoloration, or can simplify the image with detexturing. These transformations have to be meaningful, and they have to represent situations that can occur in reality. It should be noted that all the image transformations applied are to facilitate network learning, so it is important to design a good strategy in terms of the operations to be applied.

2.3.2.2.2 CAD based approaches: Rendering

Synthetic data sets applied to machine learning have been used in different areas, such as

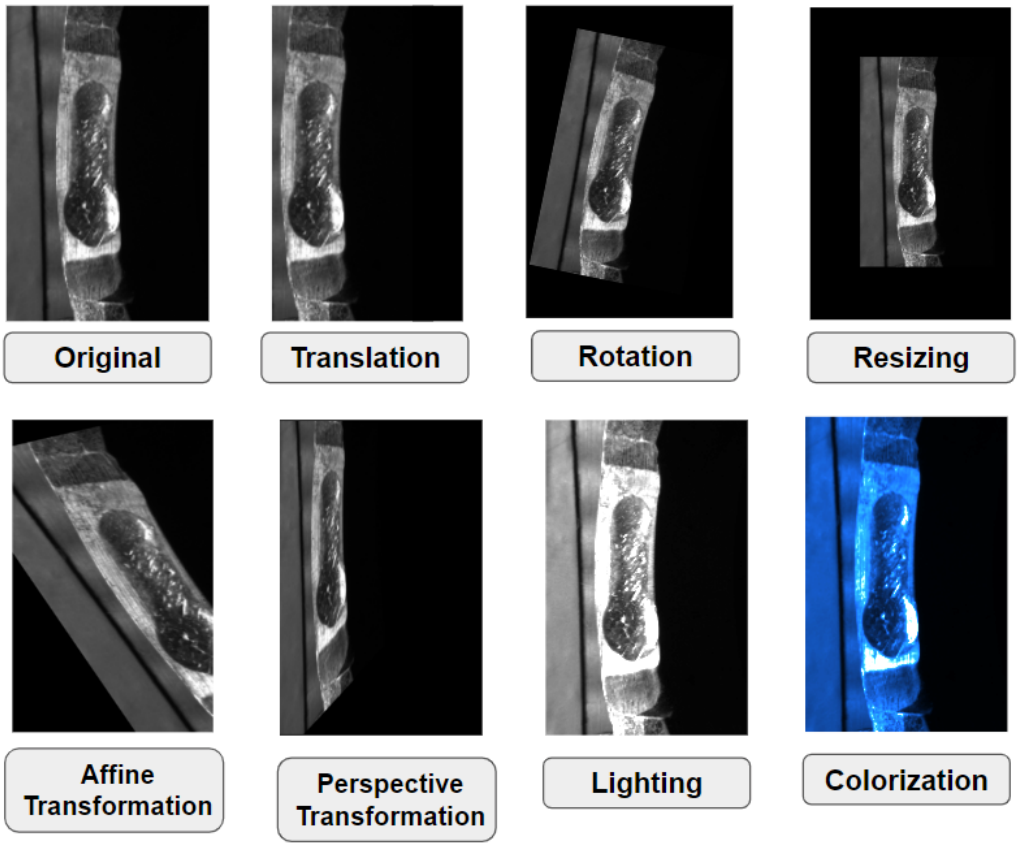


Figure 2.9: Examples of data augmentation techniques.

object detection [70], 3D object position recognition [71] or text recognition [72]. Non-photorealistic image data sets are easy to generate but usually tend to obtain poor results when used for training a machine learning models for detection or classification. When those trained models are applied to real images they use to fail because the generated synthetic data does not correctly represent the reality and therefore the trained models are not able to generalize well [73, 74]. Recent advances in computer graphics [75] allow these systems to be more photo-realistic, being able to generate images that better approximate those obtained in real scenarios [76]. These new advances in computer graphics techniques are an increasingly popular tool for training deep learning models. Indeed, some deep learning methods have obtained good performance on complex real-world images when trained only with synthetically generated data [77]. However, the deployment of deep learning models in multiple sectors is still limited by the difficulty and high consuming task for collecting high-quality data sets for the training phase. The lack of real data may require the use of synthetic rendered images to train neural networks. Specifically in the manufacturing sector, CAD models of many industrial components are often available. Therefore, it is feasible to generate rendered images to use Deep Learning in production environments, often with excellent results, as discussed in the articles [78, 79, 80].

Real-time rendering engines like Unreal Engine, Unity or CryEngine are now used to generate realistic synthetic data sets [81] in real-time thus accelerating the data sets generation and therefore accelerating the whole training models process [74].

2.3.2.2.3 Deep Learning approaches: GANs

When techniques that apply geometric or photometric transformations to the images are used, the main drawback is that no new data are introduced into the model, merely the same samples in a different state are included. Therefore, the model has already seen these samples and the impact on generalisation is limited. Generating new realistic synthetic data is a difficult task that includes learning to mimic the original distribution of the dataset, for which complex task GANs are being used.

Generative approaches to model building can be applied to learn how to synThesise realistic images from data without an underlying detailed physical model (CAD model). Recent works use Generative Adversarial Networks(GANs) for this task [82] [30]. This method demands less human involvement because both the learning process and the posterior image generation are fully automatic. However, the image annotation task needs to be done manually in a separate process.

The capabilities of GANs are impressive for data augmentation, as they can effectively learn the underlying distribution of the input data and generate very realistic samples. However, there are some limitations, such as not having a metric for assessing the quality of the generated samples. In addition, the training of GANs is often unstable and requires a lot of computational resources.

2.3.3 Third step: Network design and training

The design and choice of a neural network depends on the problem to be solved and on the available data. There are different problems that are usually solved by vision in industrial quality control systems, which can be grouped into three main classes: defect detection, defect classification and component detection [49]. Depending on the problem, the choice of the network will be different, and this will be detailed in depth in Subsection 2.3.3.1.

Regardless of the type of network chosen, training is an iterative process. While training, there is no way to estimate the exact number of images that are needed for the model convergence, nor is it possible to guess the ideal architecture and hyper-parameters that will make the training an absolute success. Each training instance will encounter problems or shortcomings that will be solved in a new iteration until the desired results are achieved.

2.3.3.1 A taxonomy of detection methods

An object detection algorithm aims to detect multiple objects and give their position in the image. These types of networks are based on basic classification networks, such as Alexnet [83], Resnet [69] and VGG [84], as their strategy consists of sampling and searching through the image in search of possible objects. This process is not as simple as simply iterating and testing, as there are multiple variables that would vary the result, such as the size of the sliding window, whether it uses one or several sizes, the displacement distance between samples, the overlap between the analysis windows, etc. In addition, the computational time that would be required if this search is configured with many parameters to test must be added. As a result of these problems, there are well-known strategies to solve them. The most relevant ones are mentioned below:

2.3.3.1.1 Object detection networks

- **Region Based Convolutional Neural Networks (R-CNN)**
  R-CNN [85] arises with the proposal to determine regions of interest within the image and then perform image classification on these regions using a pre-trained network, as shown in Figure 2.10. The region extraction algorithm could propose up to two thousand regions of various sizes, based on image characteristics such as colour, line detection, textures, or sharp changes in contrast and brightness.

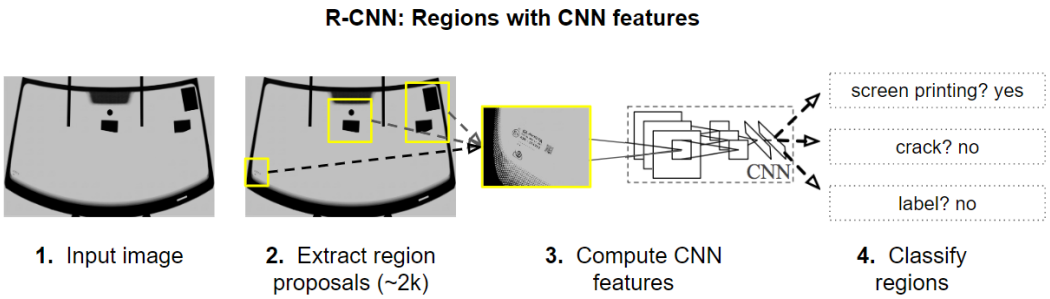


Figure 2.10: Proposed regions using CNN features in R-CNN.

The overlap between regions is controlled by using the Intersection over Union (IoU) and Non-Maximum-Suppression (NMS) concepts. The IoU provides a percentage of accuracy of the prediction area versus the actual bounding box to be detected, and it will be explained more in detail in Section 2.4. The NMS allows the selection of the bounding box that best fits the result among the multiple bounding boxes that detected the same object, eliminating those that were not chosen.

This type of network was an improvement over the iteration algorithm initially proposed, but its training and inference time was extremely slow. R-CNN served as the basis for the design of new and faster networks that follow a similar approach: Fast R-CNN and Faster R-CNN.

Fast R-CNN [86] improves the algorithm by employing the features extracted by CNN to classify object propositions more efficiently. In addition, it features improvements in the IoU and loss function to better position the bounding boxes.

Faster R-CNN [1] addresses the main bottleneck of previous models, which is selective search. This new architecture employs an algorithm called Region Proposal, shown in Figure 2.11. In addition, it introduces the pre-setting of anchors, which allow defining sizes and scales of bounding boxes to be used in the search based on the size characteristics of the objects.

These types of architectures are used in several applications of industrial manufacturing quality control systems, especially the Faster R-CNN due to its improvements and speed. The first outstanding application is the detection of defects in various types of industry, such as the textile [87] and automotive industries [88], where different types of CNNs are combined to extract features that serve as the basis for Faster R-CNN. This strategy has been used for the analysis of different surfaces, among the relevant ones for this Thesis, metal surfaces with complex and polished geometries [89] [90]. The results obtained by this network are in the range between 75% and 80% of mAP, which may be insufficient to meet some strict quality controls.

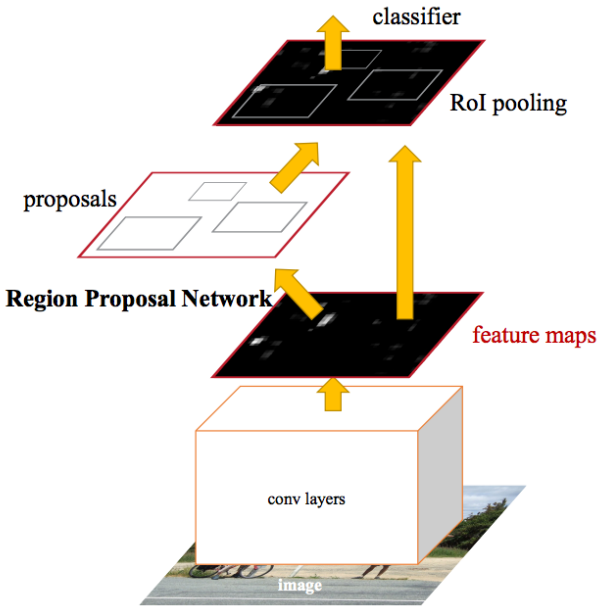


Figure 2.11: Faster R-CNN Region Proposal Algorithm [1]

Another notable application is the verification of component assembly, where the mission of the network is to identify parts and inspect their assembly [91] [92]. In this case, the model identifies and segments the part in the image, to subsequently obtain its category. Thanks to the fact that this model obtains the positions of the objects in the image, the correct position of the components in the assembly can be verified. The results presented in the work ensure that this type of model offers a robust solution in industrial environments.

- **You Only Look Once**
  You Only Look Once (YOLO) [93] is an architecture created in 2016. This network detects and classifies in a single step all the objects for which it has been trained, as shown in Figure 2.12. Thanks to this single step strategy, it achieves speeds never before achieved with no very powerful computers. This allows real-time video detection of hundreds of objects simultaneously and even execution on mobile devices.

  Yolo defines a fixed-size grid over the image, divided in several cells. On these cells it tries to detect objects using fixed anchors. Like R-CNN, it uses IoU and Non-Max-suppression techniques. It also has an associated regression network in the last step for the bounding-boxes position calculation. The main advantage of YOLO is due to its CNN network. While R-CNN used some additional algorithm to select the regions of interest on which it makes predictions, YOLO uses the same CNN classification network with an algorithm whereby it does not need to iterate through the grid and does the detection simultaneously. In terms of its architecture, YOLO uses a CNN network called Darknet [94], although it can also be trained with any other Convolutional network. It embeds in its output both the classification and the positioning and size of the detected objects.

  Over time, this network has evolved and new versions have been created, such as:

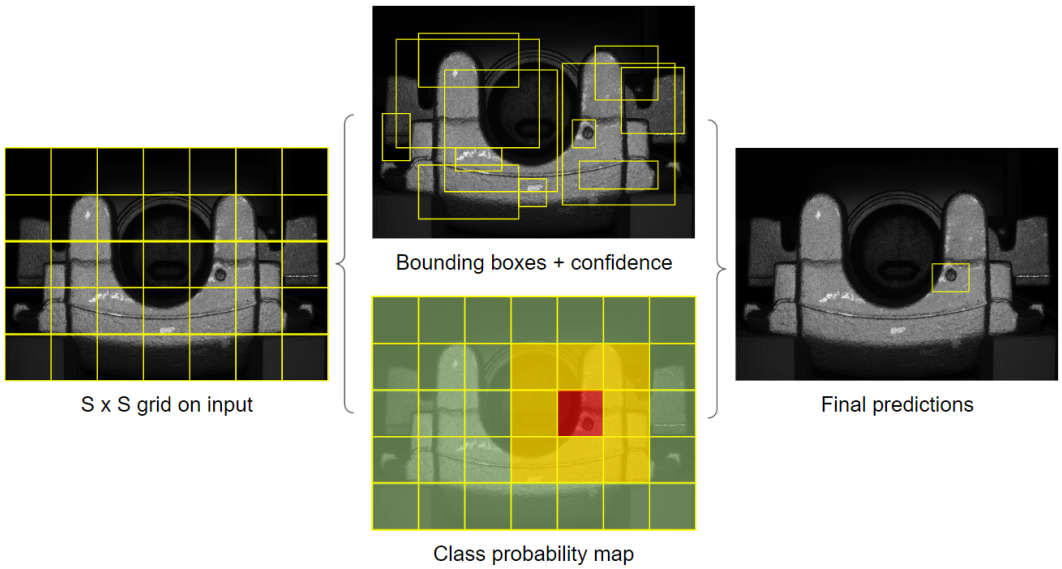


Figure 2.12: Schema of YOLO detection and classification algorithm.

YOLOv2 [95], YOLOv3 [94], YOLOv4 [96], YOLOv5 [97] and YOLOv6 [98], which are focused on improving the accuracy of the bounding boxes, while maintaining their speed.

Thanks to the inference speed offered by this architecture, it is widely used in quality control applications in manufacturing industry. It is widely used in the detection of defects in metallic components [99] [100] [101], obtaining results of over 90% in mAP. In addition, it is also used in component assembly testing tasks, where the network's capacity to learn a large number of different classes stands out [102].

- **Single Shot Detector**
  SSD [2] has a pyramidal structure in its CNN in which the layers gradually decrease. This allows it to detect both large and small objects. It does not use a predefined grid, but has anchors of different proportions that scale as go down the pyramid (map of smaller features, with proportionally larger anchors).

  The main architecture, as shown in Figure 2.13, is an FPN and a backbone that can be chosen, although in its first versions it used a VGG19.

  This network is also relevant for its speed, but it is not as fast as YOLO, so it is not a good choice for inspection applications in high throughput environments [103].

- **RetinaNet**
  The Facebook AI Research team presented a new object detection algorithm, RetinaNet [104], which, despite performing the detection task in a single step, obtained results comparable to those of Faster R-CNN.

  The authors of the paper found that, in object detection and instance segmentation algorithms, the vast majority of windows created by the network during the initial phases of its training were negative (they did not capture any objects). As a result, there was a decompensation in the training of the loss function in favour of these

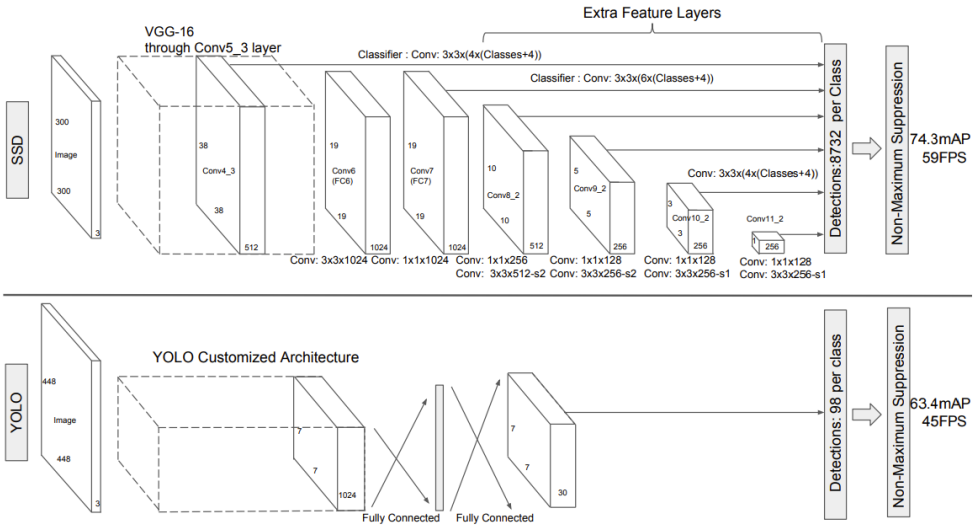


Figure 2.13: SSD architecture [2]

negatives. In two-stage algorithms (such as Mask R-CNN), this problem did not occur, as the auxiliary network had previously discarded these false negatives.

For this reason, the researchers proposed to modify the cross-entropy loss function in such a way that the easy-to-detect negatives would contribute very little to the cross-entropy loss. The new function, called the focal loss (FL), has the form shown in Equation 2.5.

$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma} \log(p_t) \tag{2.5}$$

where FL is the loss function, p_t is the probability attributed by the network to the correct category and γ a constant. When p_t is very high, it means that the network has got it right.

2.3.3.1.2 Segmentation networks

Segmentation neural networks obtain the pixel mask of the objects of interest to be extracted from an image, performing a pixel-level classification. These networks are an alternative that provides superior performance in complex applications compared to more classical techniques such as thresholding, growing regions or Watershed's algorithm. Despite many traditional techniques, the emergence of deep neural networks has implied a radical paradigm shift in terms of object segmentation, achieving optimal results in terms of accuracy and speed.

They have encoder and decoder modules. The encoder is typically a pre-trained classification network. The decoder, however, is in charge of projecting the features learned by the encoder, until a dense classification is achieved.

Image segmentation can be divided into two problems: the classification of pixels with semantic labels or the segmentation of individual objects. Semantic segmentation performs pixel-level labelling of a set of object classes, however, instance segmentation

further extends the scope of semantic segmentation by detecting and delineating all objects of interest in an image. The following is a review of the most commonly used segmentation architectures:

- **Fully Convolutional Network**
  [105] have been the first to develop a Fully Convolutional Network (FCN) trained end-to-end for image segmentation. This network is based on classical convolutional networks such as AlexNet [83], VGG [84] or GoogleNet [106]. It is mainly used for semantic segmentation.

  In terms of architecture, FCN has only locally connected layers, such as convolutional, pooling and oversampling layers. By avoiding dense layers, it prescends parameters, thus achieving a lighter network. Its architecture extracts image features in down-sampling, and performs object localisation in up-sampling as shown in Figure 2.14.

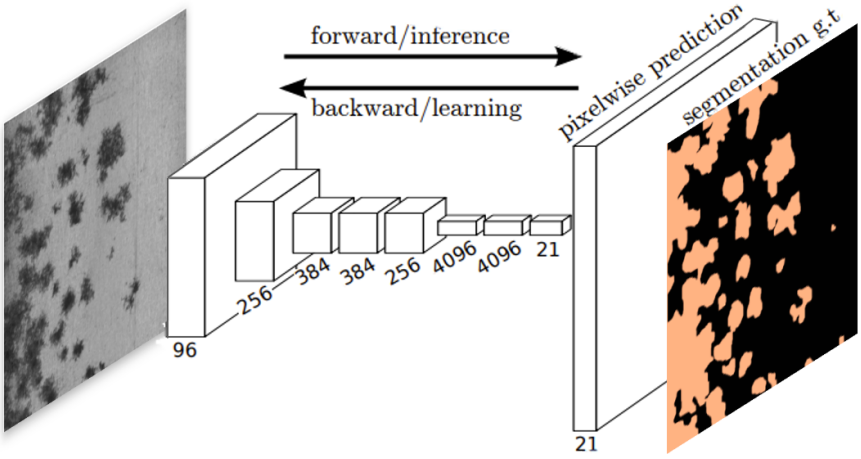


Figure 2.14: FCN architecture.

  The results obtained by this network in industrial quality control applications are promising [107]. It has been used for defect segmentation in large images, where defects are a small part of the image, so it proved to be stable in detecting small defects [108]. It has also obtained very good results in segmentation with complex backgrounds, where problems of noise caused by illumination or irregular geometries and textures are presented [109].

- **Feature Pyramid Network**
  The Feature Pyramid Network (FPN) allows extracting features from the input image at various levels. This process is independent of convolutional backbone architectures, so it is a generic solution that builds feature pyramids within deep convolutional networks.

  The pyramid construction is performed in two directions: ascending and descending. In the ascending direction, the usual feature extraction is performed, as the spatial resolution decreases but the semantic value increases. In the descending direction, higher resolution layers are constructed using the semantic layer. To improve the details lost in subsampling, lateral connections linking both directions are provided to improve the network output and facilitate training.

- **Mask R-CNN**
  Mask R-CNN [3] is a convolutional neural network that performs image segmentation and instance segmentation. This network is based on Faster R-CNN [1], but differs in the outputs it has. While Faster R-CNN gets two outputs, a class label and a bounding box that locates the object in the image, Mask R-CNN also generates the object mask. The additional mask output is distinct from the class and box outputs, requiring the extraction of a much finer spatial representation of an object.

  This network is a hybrid model of segmentation and object detection, so its category should not be a single one. It is placed in this section to provide a preliminary grounding in the fundamentals of the networks on which it is based.

  In terms of its architecture, the key element of this network is pixel-to-pixel alignment, which is the main shortcoming of Faster R-CNN. It has two stages. The first one generates proposals of regions where an object could be present. In the second stage, the object class is predicted and the bounding box is refined. Additionally, it generates a pixel-level mask of the object based on the proposed regions from the first stage as shown in Figure 2.15. The communication between the two stages is done with different connections along its structure.

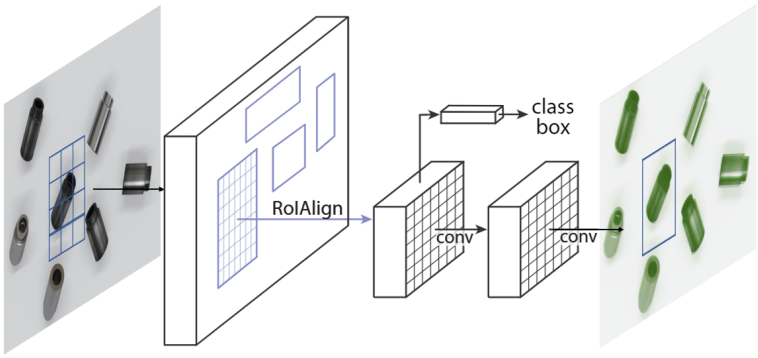


Figure 2.15: The Mask R-CNN framework for instance segmentation [3]

  This network has many industrial applications that have a higher complexity and need to be solved with a combination of techniques to achieve the desired results. Different studies have been found that demonstrate its ability to detect objects in complex environments. This network has been applied in different scenarios with lack of data, where the lack of data was overcome with synthetic data [110] or transfer learning [111], or in non-conventional surfaces such as sheepskin [112], obtaining promising results in all scenarios.

- **DeepLab**
  DeepLab is an advanced segmentation model designed by Google. The novelties of this network are given by the increase of the output sampling of the last convolution layer and the modification in the calculation of the loss, which is now done at pixel level. In addition, DeepLab introduces atrous convolution for up-sampling.

  Atrous convolution means an alternative to the use of deconvolution, which increases memory requirements and computation time. This type of convolution employs sampled filters, thus extending the field of view of the filters without increasing the number of parameters or computation requirements.

DeepLab was improved over time, having different versions. The most relevant features of each version are explained below:

DeepLabv1 [4]: This new version eliminates the down-sampling operator of the last layers and instead increases the sampling of the atrous filters in later convolutional layers, thus obtaining feature maps calculated with a higher sampling. They introduce the Conditional Random Field (CRF) concept in order to capture fine details. Deeplabv1 takes the input images and after passing through atrous layers obtains an approximate score map. This map is sampled at the original image size by bilinear interpolation. Finally, in order to improve the segmentation result, the previously mentioned CRF is applied. The outline of this process is shown in Figure 2.16.

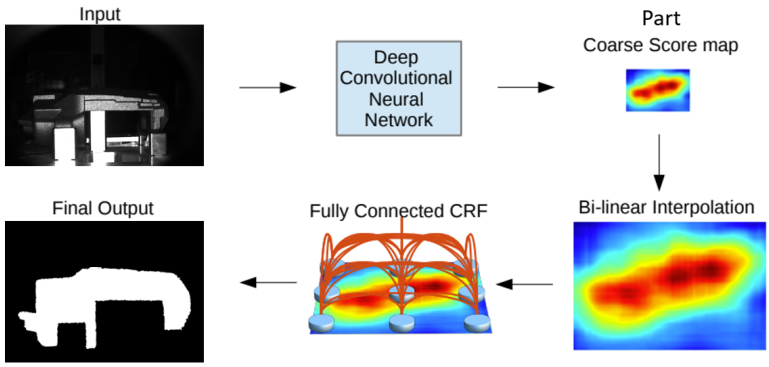


Figure 2.16: Deeplabv1 model outline illustration [4]

DeepLabv2 [113]: Regarding DeepLabv2, its main objective was to improve the performance of DeepLabv1, focusing on improving the detection of multi-scale objects using Atrous Special Pyramid Clustering (ASPP). The strategy is based on applying atrous convolutions with different sampling rates to the input feature map and then merging them. ASPP allows taking into account different scales of objects, thus improving the accuracy of the network.

DeepLabv3 [114]: The DeepLabv3+ model corresponds to a deep neural network that applies several parallel atrous convolutions with different rates (specifically called Atrous Spatial Pyramid Pooling, or ASPP). This is useful since the important semantic information is encoded in the last feature map, and the atrous convolution allows extracting denser feature maps.

One of the most important and innovative parts of the network is the Atrous convolution, which is a powerful tool that allows explicitly controlling the resolution of the features computed by deep convolutional neural networks and adjusts the field of view of the filter to capture multi-scale information, i.e. it generalises the standard convolution operation. In particular, in the case of two-dimensional signals, for each location i in the output feature map y and a convolution filter w, an atrous convolution is applied on the input feature map x, following the Equation 2.6.

$$y[i] = \sum_{k} x[i + r \cdot k] w[k] \tag{2.6}$$

Where r represents the atrous rate which determines the step at which the input signal is sampled. In the case of $r = 1$ is the special case of standard convolution.

2.3.3.1.3 Anomaly detection networks

In the field of autonomous learning, unsupervised algorithms are becoming more and more relevant, because most of the available data and data that can potentially be collected are not labelled or with a defined structure, making this type of artificial intelligence algorithms very useful.

Among the deep learning architectures or Deep Learning, the best known models are Autoencoders, Long Short Term memory networks (LSTM) or generative networks (GANs).

Anomaly detection is a sub-area of unsupervised learning, where the objective is to detect patterns within the data, which deviate from an established norm.

- **Autoencoders**
  Autoencoders are neural networks that detect anomalies based on the reconstruction they generate. Anomalies are detected in data whose reconstruction is erroneous. Autoencoders are neural networks that are trained by means of back propagation. They consist of two stages, encoding, where a reduction of the data set dimensions takes place, and then subsequent decoding, where reconstruction of the original data takes place. By comparing the input data with the reconstruction generated at the output, it is possible to identify anomalies in the data set.

  There are different types of autoencoders: Convolutional AutoEncoders (CAEs) and Contractive AutoEncoders.

  In CAEs the input data is considered as a sum of different types of signals, which can be filtered by means of convolutional operations. These neural networks are able to learn which is the most suitable filter to represent the input data in smaller dimensions and also to reduce the error rate for reconstruction [31] [115].

  Contractive AutoEncoders [116] [117] are a variation where the network is intended to be sufficiently robust to small variations in the input data. This is achieved by applying a type of regularisation to the network that corresponds to the Frobenius norm of the Jacobian matrix [118] of the encoder activation function.

- **Convolutional LSTMs**
  LSTMs [119] [120] are a special type of recurrent neural networks (RNNs) in which certain knowledge acquired by the network is persistently maintained and used as a basis for predictions.

  In convLSTM networks, convolutional neural networks (CNNs) are used to extract features from spatial data in images or videos, and then use similar features extracted in the past to make predictions.

  This allows the creation of models that generate a correlation of space and time in a given video. A typical architecture made using LSTM models has two branches, one in charge of predicting future frames and the other of reconstructing current or past frames. With these two metrics, a probability of the presence of anomalies in the data can be determined.

In general, in the case of generative models, the way to detect anomalies is through the probability that a certain element is reproduced or generated by the network; if this probability is low, then this element may be an outlier.

- **Variational AutoEncoders (VAEs)**
  VAEs are a generative model that are able, by means of the probability density function, to create images, or missing portions of images, during the reconstruction of the data [121] [122].

  Generally the data that are generated remain close to the ensemble norm, so that by checking them against the input data it is possible to detect anomalies.

- **Generative adversarial networks (GANs)**
  These are networks consisting of two parts, the generative network and the deterministic network. These networks compete with each other; the generative network produces candidates and the deterministic network evaluates them.

  In the context of anomaly detection, these can be useful when trying to generate a given sample; if it is not possible to generate it, this may indicate that anomalous data is present. [123] [124]

2.3.4 Fourth step: Validation and Deployment

In order to know when a model is fit or poor, it must be validated on the test set. This set must be well constructed, so that it contains examples of what the model may encounter when it is in production. If this set is properly constructed, the results obtained in the validation stage will be similar to those that the model will have in the future.

In this stage, different classification performance metrics are computed to obtain numerical evidence of the behaviour of the model in images not seen in the training stage. When these metrics meet the objectives defined for the automatic inspection system, the next and final step is the deployment of the model. These metrics are explained in more detail in Section 2.4 below.

Deep Learning based systems need special computing hardware components that are not common in industrial machines, e.g. a GPU. The quality control, depending on the component being inspected, has very demanding speed requirements, and this, together with the computational requirements of Deep Learning models, means that having a GPU in the computer assigned to quality control is necessary. Furthermore, these computers not only make a quality assessment, they also have to keep a traceability of the defect detections for other purposes such as productive maintenance or for quality certification requirements. Therefore, the software architecture and the way in which these types of systems are deployed in the production line is not a simple task.

2.4 Defect detection performance measures

In the field of surface defect detection, the following performance measures are often used to evaluate the results obtained:

- True Positives (TP): the defect is detected as defect.

- True Negatives (TN): the background is detected as background.
- False Positives (FP): the background is mistakenly detected as defect.
- False Negatives (FN): the defect is mistakenly detected as background.

Besides the statistics described above, the following metrics are commonly used to evaluate the quality of the segmentation:

- Mean Dice Value: is a spatial overlap based metric. It focuses on measuring the similarity between two samples through the union and intersection of sets of predicted and ground truth pixels. This index takes values between 0 and 1, and is better when it is closer to 1, since this means to have more surface in common between the ground truth and the result of the segmentation. This coefficient is calculated by the Equation 2.7, where A and B are the ground truth mask and the predicted mask respectively.

$$Dice(A, B) = \frac{2\,\|A \bigcap B\|}{\|A\| + \|B\|} \tag{2.7}$$

- Sensitivity: is the ability of the model of not marking a negative sample as positive. It is measured by the formula of Equation 2.8.

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{2.8}$$

- Specificity: is the ability to find all positive samples. It is measured by the formula of Equation 2.9.

$$Specificity = \frac{TN}{(TN + FP)} \tag{2.9}$$

- Pixel accuracy: the percent of pixels in the image that are classified correctly. It is calculated by the Equation 2.10.

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{2.10}$$

- F1 Score: The F1 Score is an harmonic mean of the Precision and Recall Metrics. The closer to 1, the better performer is going to be the DL model. Mathematically, it is expressed in 2.11.

$$f1Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{2.11}$$

- AUROC and ROC: The AUROC (Area Under the ROC curve) and ROC curve (Receiver Operating Characteristic) is one of the most widely used metric in binary classification problems. It indicates the probability that the model classifies a randomly chosen positive sample as positive. The closer to 1 the better performer is the model.

CHAPTER 3

Hypotheses and Objectives

The main objective of this Thesis is to research and develop intelligent systems for image based quality control that overcome the multiple adversities that they face in the harsh environments of production plants. Let us briefly highlight some of these adversities: (1) high manufacturing cadences that hinder the real time data acquisition due to the speed and management of the images, (2) the components under inspection are complex in terms of geometry and/or material so advanced optical sensors are required, and (3) production installations offer a hostile environment with particles in suspension and the influence of ambient lighting that impairs the collection of quality data and affects its operation due to its variable nature over time.

This Thesis reports on the design and experimental evaluation of AI-based inspection systems for reliable final machines. Specifically, (a) we evaluate advanced acquisition systems such as photometric stereo, (b) we evaluate and design convolutional neural networks for the detection of small defects or objects in images, (c) we study the impact of the data on the generalisation and learning of the networks, and (d) finally validate the results experimentally. All the research works have been carried out during the development of to real industry projects. Chapter 4 details how the research result fit into the frame of the applied industrial projects which justifies the industrial labelling of this Thesis.

In the following, we outline the different hypotheses and objectives guiding our research work. We also provide a brief description of the publications supporting this Thesis as a collection of published papers and a recollection of the identified key contributions of our work.

3.1 Hypotheses

- The impact of employing an acquisition system that is robust in terms of illumination and that can highlight defects in the complex and reflective surfaces that are common in industrially manufactured components is critical in the design and behaviour of the final intelligent system that automates quality control.
- Data has a major impact on the learning and generalisation capabilities of neural networks, and the lack of this information can be alleviated by the use of advanced

data augmentation techniques, such as the use of geometric transformations, image rendering from CADs or the use of synthetic samples generated by GANs.

- To maximise the training results, the networks must be adapted to the characteristics of the data and the target scenario, which is the detection of small defects in high-resolution images.
- The validation process of these systems must be done in an iterative way with the help of quality experts, and their deployment in the factory entails a series of requirements in terms of software architecture that must be met.

3.2 Objectives

The main objective of this Thesis is to improve and develop artificial intelligence systems for image based quality control in the manufacturing industry.

The objectives derived from this main objective are as follows:

- Develop advanced acquisition systems for the complex characteristics of the components manufactured in the industry.
- Study the impact of data on the final result of these systems and on the training of neural networks.
- Develop data augmentation methods for the specific task of defect detection.
- Develop neural networks optimised for defect detection and for the special characteristics of the acquired data.
- Validate and deploy these systems in production lines, thus automating quality control.

3.3 Brief description of the publications and contributions

The research work is assembled in a total of four main articles. Each research article focuses on accomplishing of one or several technical objectives, answering the hypothesis listed above.

3.3.1 Article 1: Photometric Stereo-Based Defect Detection System for Steel Components Manufacturing Using a Deep Segmentation Network

This paper presents an automatic system for the quality control of metallic components using a photometric stereo-based sensor and a customized semantic segmentation network. This system is designed based on interoperable modules, and allows capturing the knowledge of the operators to apply it later in automatic defect detection. A salient contribution is the compact representation of the surface information achieved by combining photometric stereo images into a RGB image that is fed to a convolutional segmentation network trained for surface defect detection. We demonstrate the advantage of this compact surface

imaging representation over the use of each photometric imaging source of information in isolation. An empirical analysis of the performance of the segmentation network on imaging samples of materials with diverse surface reflectance properties is carried out, achieving Dice performance index values above 0.83 in all cases. The results support the potential of photometric stereo in conjunction with our semantic segmentation network.

3.3.2 Article 2: An Inspection and Classification System for Automotive Component Remanufacturing Industry Based on Ensemble Learning

This paper presents an automated inspection and classification system for automotive component remanufacturing industry, based on ensemble learning. The system is based on different stages allowing to classify the components as good, rectifiable or rejection according to the manufacturer criteria. A study of two deep learning-based models' performance when used individually and when using an ensemble is carried out, obtaining an improvement of 7% in accuracy in the ensemble approach. The results demonstrate the successful performance of the system in terms of component classification.

3.3.3 Article 3: Generative Adversarial Networks to Improve the Robustness of Visual Defect Segmentation by Semantic Networks in Manufacturing Components

This paper describes the application of Semantic Networks for the detection of defects in images of metallic manufactured components in a situation where the number of available samples of defects is small, which is rather common in real practical environments. In order to overcome this shortage of data, the common approach is to use conventional data augmentation techniques. We resort to Generative Adversarial Networks (GANs) that have shown the capability to generate highly convincing samples of a specific class as a result of a game between a discriminator and a generator module. Here, we apply the GANs to generate samples of images of metallic manufactured components with specific defects, in order to improve training of Semantic Networks (specifically DeepLabV3+ and Pyramid Attention Network (PAN) networks) carrying out the defect detection and segmentation. Our process carries out the generation of defect images using the StyleGAN2 with the DiffAugment method, followed by a conventional data augmentation over the entire enriched dataset, achieving a large balanced dataset that allows robust training of the Semantic Network. We demonstrate the approach on a private dataset generated for an industrial client, where images were acquired by an ad-hoc photometric-stereo image acquisition system, and on a public dataset, the Northeastern University surface defect database (NEU). The proposed approach achieves an improvement of 7% and 6% in an intersection over union (IoU) measure of detection performance respectively over the conventional data augmentation.

3.3.4 Article 4: COVID-19 detection in chest X-ray images using a deep learning approach

The Corona Virus Disease (COVID-19) is an infectious disease caused by a new virus that has not been detected in humans before. The virus causes a respiratory illness like the flu with various symptoms such as cough or fever that, in severe cases, may cause pneumonia.

The COVID-19 spreads so quickly between people, affecting to 1,200,000 people worldwide at the time of writing this paper (April 2020). Due to the number of contagious and deaths are continually growing day by day, the aim of this study is to develop a quick method to detect COVID-19 in chest X-ray images using deep learning techniques. For this purpose, an object detection architecture is proposed, trained and tested with a public available dataset composed with 1500 images of non-infected patients and infected with COVID-19 and pneumonia. The main goal of our method is to classify the patient status either negative or positive COVID-19 case. In our experiments using SDD 300 model we achieve a 94.92% of sensibility and 92.00% of specificity in COVID-19 detection, demonstrating the usefulness application of deep learning models to classify COVID-19 in X-ray images.

3.4 Key contributions

The following are the technical and methodological key contributions of this Thesis:

Advanced acquisition systems development for complex specular surfaces

- We have developed and tested a system based on diffuse illumination on non-Lambertian surfaces with a white light source to extract surface topographic information and spectral (color) response at the same time.

Advanced image processing techniques to maximise the information present in images in order to facilitate and improve the training of deep neural networks

- We have shown the benefit of using compact RGB images composed by photometric stereo channels to feed segmentation networks.
- We have developed and applied in practical problems a processing pipeline for the generation of synthetic images of industrial environments, for the training and validation of Deep Learning models for inspection systems.
- We have demonstrated the use of GANs as a data augmentation method, validating their contribution to improve defect segmentation on metal surfaces. The method is validated on different well-known segmentation architectures.
- We have demonstrated different advanced image processing methods to enhance the image information and facilitate the training of deep neural networks.

Development and research of deep neural networks for fast, accurate and reliable detection of surface defects in metal components

- We have designed and validated a specialised defect segmentation network for RGB images composed by the different images obtained in the photometric stereo reconstruction.
- We have demonstrated the benefits of model assembly for defect detection and segmentation with high intra-class variability applied to industrial remanufacturing.
- We carried out extensive research about the best convolutional neural networks for the classification of typical defects on metal surfaces.

Validation of automatic inspection systems in real industrial manufacturing production lines

- We have carried out the rollout and validation of inspection systems in different manufacturing production lines as well as a proposed software architecture for the deployment of deep learning-based models in production.
- We have developed automated quality control systems designed for ad-hoc troubleshooting which were validated on real production lines.

CHAPTER 4

Results

In this chapter, we provide the details of how our results answer the requirements of the industrial projects directly related to this Thesis, and their relationship with the achieved publications.

4.1 KEYINSPECT: Machine vision inspection system of reflective components

The business objective of the **KEYINSPECT** project is to automate the inspection of reflective swords to allow the customer to remain competitive in the market. In recent years, the volume of nickel-plated keys in the automotive market has increased dramatically due to a strategic change in the design and aesthetics of the final product. Taking into account the compromise acquired by the manufacturing company with the final customers, it is essential to find a solution in order not to lose competitiveness. On the social side, this project will involve to create new jobs with higher qualifications, without eliminating the current ones, and at the same time increase productivity, raising production towards the industry of the future.

One of the scientific-technological objectives in the inspection of automotive sector parts is mainly to be able to identify all the defects on the component surfaces and especially on the reflective ones. The system developed in this project must be able to verify more than 3000 parts per hour and correctly eliminate 98% of the parts that do not comply with the predetermined characteristics, while the false rejection rate must not exceed 5%. The system is looking for both aesthetic and dimensional defects over parts with and without surface treatment. An example of the variety of these parts is shown in Figure 4.1.

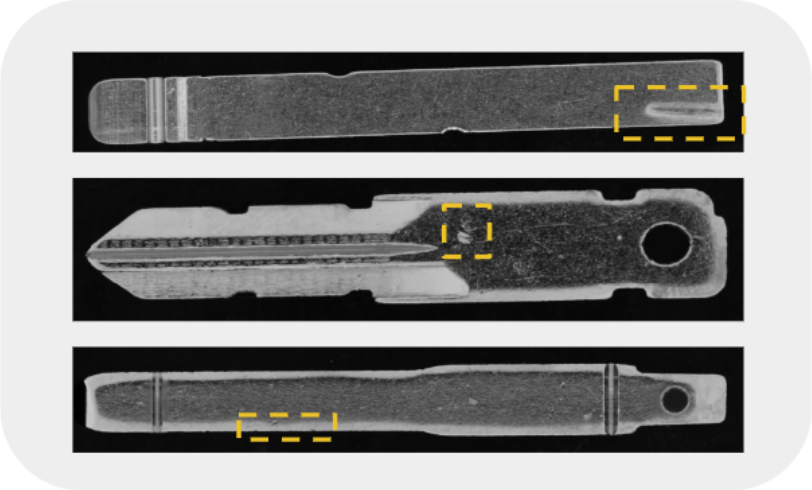


Figure 4.1: An example of different parts dealt with in the **KEYINSPECT** project with their defects highlighted in yellow.

To attack these challenges, a system was designed, built, and validated with a photometric stereo illumination ring able to obtain high resolution images despite the high reflectivity of the surface of the objects to be analysed. A tool has been developed with multi-detectors based on machine learning techniques, which are able to detect all the surface defects defined by the manufacturer. The models developed for this machine was designed and optimised for the type of images obtained with this method of image acquisition, thus achieving the quality requirements imposed by the customer.

The results of this project are published in the paper *"Photometric Stereo-Based Defect Detection System for Steel Components Manufacturing Using a Deep Segmentation Network"*, where it is demonstrated that the use of a photometric stereo sensor is adequate for the application of defect detection on reflective surfaces. The proposed method of stacking multiple sources of data from the photometric stereo sensor into a high resolution multi-channel (RGB) image is a key contribution to the improvement in the accuracy performance achieved with the ad-hoc Convolutional Neural Network. A typical problem with this type of application is the difficulty of detection of very small defective regions with respect to the dimension of the component. This makes it difficult to avoid false positives as the dataset is greatly unbalanced between the amount of normal and defect pixels. The added extra information provided by the multiple data layers from the photometric stereo allows the model to extract new features related with defective regions in a better an optimal way.

In comparison against other inspection methods proposed in the state of the art our method allows the inspection of reflective surfaces in the cycle time required by the manufacturing company. Other methods propose multiple image acquisitions from different angles to avoid surface reflections, thus affecting the acquisition time and also the required space for the installation of the inspection station in the production line. Compared with the methods that employ Deep Learning our method is novel in using multiple combined sources of information allowing a more detailed and robust description of the component surface. This feature gives greater stability in defect detection compared with traditional image acquisition based systems. In addition, a specific segmentation network is created allowing the exploitation of this information in a fast and accurate way. The proposed acquisition

system can be also used on non-planar geometries, such as cylindrical components just by modifying the automation of the component's displacement.

A previously referred challenge in machine vision applications developed for industrial manufacturing processes is the speed at which they must operate to comply with high production rates. Defect segmentation applications are often not fast enough to be integrated into an industrial production line. However, our proposed system is able to process images in the required cycle time. We demonstrate that if the inference is performed on an industrial grade computer, the required response time would not be achieved. Hence, a distributed computation for parallelizing processes is mandatory. This computing architecture is also a benefit in terms of organisation and scalability, since the same system can inspect components in several production lines and more hardware resources can be easily added to cope with the required compute demand.

4.2 CARJAULAS: Assembly of deep learning models for the classification of CV joints for remanufacturing

The general objective of **CARJAULAS** project is the application of Computer Vision to solve complex problems of remanufacturing companies in the fields of handling, diagnosis and automatic classification of multi-reference parts. The industrial remanufacturing process consists of the recovery of a previously used product or component to the same or larger dimensions in terms of performance and quality of that product/component.

The use case addressed in this project is the recovery of vehicle transmissions and, more specifically, the verification of the surface condition (wear) of the recovered transmission CV joints, classifying them into three groups according to their wear: recoverable without machining, recoverable with machining, non-recoverable. These wear defects occur in the area where the CV joints come into contact with the balls, known as the bearing contact zone. An example of the wears is shown in Figure 4.2. During the process of transmission restoration, it is necessary to check the wear on both sides of all cage races. Currently, this check is done subjectively by an operator based on his sense of touch. This manual process generates a lot of uncertainty, which results in miss-classifications: parts that would be recoverable are rejected and out-of-tolerance parts are recovered and end up failing prematurely when serviced to the final user.

Figure 4.2: CARJAULAS acquisition system and an example of the wear images obtained.

One of the main challenges was the implementation of an illumination/optical assembly

to highlight this defect, as it is very small and difficult to access. This problem is further accentuated when considering a universal sorting system for CV joints of different families with quite different sizes and geometries. It is also necessary to implement a handling system to acquire the top and bottom sides of all their tracks (6-track and 8-track models). Likewise, a model based on Deep Learning was developed to classify the pieces based on supervised training using pre-classified samples. This model was validated by the collaborating company GKN Driveline, using new samples that were not used during the training process and with different references in order to have a significant result that can be extrapolated to the real operation of the system when deployed in the production line.

In this project, more than 80 different parts needed to be inspected. Besides having different geometries, they were also grease-stained and damaged at different levels. In order to obtain a robust and reliable system, it was necessary to have more than one model to analyse each image in parallel. To this end, a study of traditional machine learning techniques and Deep Learning models was carried out to find the best combination of classifiers.

The results obtained with traditional methods were not accurate enough regarding the manufacturing industry requirements. These methods detect more false positives, increasing false rejections. An automatic system based on these techniques would be unreliable and would generate more financial and environmental costs than manual inspection.

Proposed system based on a combination of YOLOv5 and DeepLabV3+, i.e., an ensemble of classifiers, classifies the component based on the individual results per bearing contact point region and following the customer's criteria. This system achieves an accuracy of 100% in the overall performance test, entailing a promising tool to solve the problem presented by the customer.

The publication *"An Inspection and Classification System for Automotive Component Remanufacturing Industry Based on Ensemble Learning"* is the direct result of this project.

4.3 BATTERYCHECK: Automatic quality control system for car batteries

The project **BATTERYCHECK** arises from the need to automate Clarios' battery quality inspection system. Clarios carries out the entire manufacturing process of its batteries and controls the quality at different stages of production using different vision processes. However, in some complex processes quality control is carried out manually. This project comes from the need to automatically detect the defects presented in its catalogue, thus reducing the need to perform this manual control.

The batteries have different types of defects in the different components that are part of the battery, such as: lack of filling or burrs in the lead components, plastic separator damage or repetition and missing metal plates. In this project a 2D and 3D acquisition systems was designed and installed, which made possible to study the information coming from both sources of information. The catalogue of manufactured references is extensive, so the system developed is agnostic to the reference being manufactured. The algorithms designed were validated on two production lines at the Clarios factory in Burgos.

In addition to the current production plant, The system is intended to be deployed in

multiple lines of the company, so an easy deployable software architecture was developed. Thanks to this architecture the operators have control over production, have a history of each line, and can take corrective actions in case of detecting deviations. An example of the visualization that operators has in each line is shown in Figure 4.3.

Figure 4.3: Visualization of BATTERYCHECK application where the image and point cloud are shown, with the inspection result bellow.

This project, apart from including the specific acquisition system design necessary to solve the problem, and the use of methods based on Artificial Intelligence to locate the defects, carried out an extensive study and development of the optimal deployment of these systems, known as cyber-physical production systems, in production factories. These cyber-physical production systems (CPPS) integrate both the virtual and physical parts of manufacturing lines. This integration requires information technology (IT) and operations technology (OT), standards and specifications related with Industry 4.0. This leads to an intrinsic complexity that poses a barrier to the deployment of CPPS in real-world manufacturing scenarios. This project has enabled the development of a simple edge architecture, based on micro-services and focused on a common CPPS task: the management of asynchronous manufacturing data analysis jobs. As a result, the paper *"Containerized edge architecture for manufacturing data analysis in Cyber-Physical Production Systems"* was published.

4.4 INN-SURF: Unitary surface inspection system for hot inners

GKN Driveline is the largest division within the Global Engineering Group, GKN PLC. GKN leads the global manufacturing supplier base for the automotive, heavy machinery and aerospace industries. GKN provides technology based, highly engineered products to producers of light vehicles, agricultural construction equipment and aero engines. Nearly 40,000 people work in GKN companies and joint ventures in more than 30 countries.

The scope of the resulting **INN-SURF** inspection system impacts the INNERS line. The inners line includes 2 medium heat presses producing at a rate of 1500 parts per press per hour, with an annual capacity of 14 million components. Defects that typically occur in this component include bumps, burrs and underfills.

The mechanical configuration adopted was designed to meet the following requirements: accuracy in the positioning of the part for the correct acquisition of the defect, cycle time maximising the acquisition time and working in parallel in the image processing for decision making, stability of the illumination in the different working conditions, temperature and part geometry, mechanical robustness and isolation of press vibrations, compact layout given the space limitations at the press exit and mechanical simplicity.

Different software modules were developed to analyse different zones of the inner: upper zone, lower zone and sides, as shown in Figure 4.4. Due to the repetitiveness in terms of appearance of the components and given the difficulty of building a defect database, anomaly detection models were developed. The models were integrated into different software modules of the project, which are supported by an Edge Computing platform.

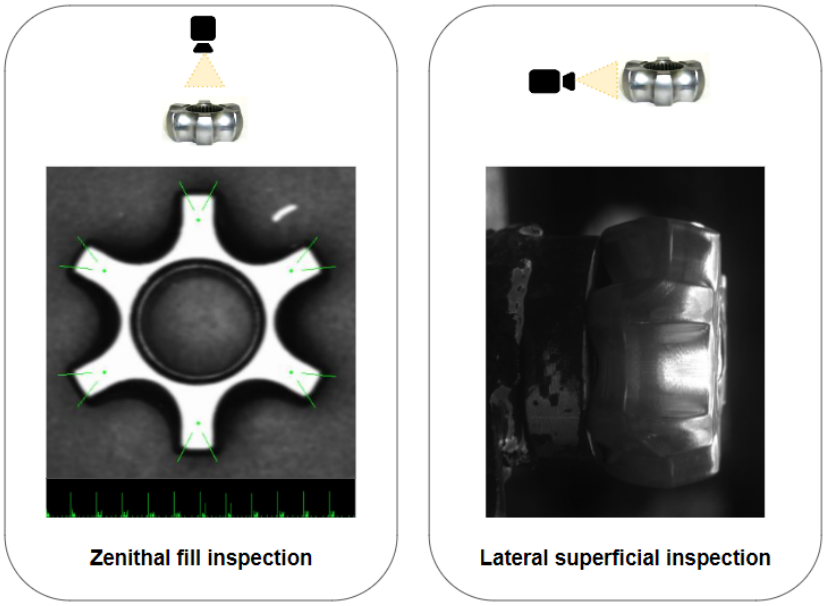


Figure 4.4: CV Joint inner inspection: zenithal for fill inspection and lateral for superficial inspection.

The main problem of this project is the harsh conditions of its manufacturing environment, which limits the ability to test and collect a high quality database for training the neural models. The machine where hot quality control is performed has a very high production rate, which makes it impossible to manually check piece by piece for defects. In addition, it has a large imbalance between non-defective and defective samples.

To alleviate these problems, this project has proposed a pipeline to generate rendered images from CAD models of industrial components and then feed them to a Deep Learning-based anomaly detection model. This approach can simulate the possible geometric and photometric transformations in which parts could be presented to a real camera to faithfully reproduce the image acquisition behaviour of an automatic inspection system. We evaluate the accuracy of several neural models trained with different synthetically generated

datasets simulating different transformations such as part temperature or part position and orientation with respect to a given camera.

The results show the feasibility of the proposed approach during the design and evaluation process of the image acquisition setup and to ensure the success of the future real application, as can be shown in the publication *"Synthetic Data Set Generation for the Evaluation of Image Acquisition Strategies Applied to Deep Learning Based Industrial Component Inspection Systems"*.

4.5 BLANKSURFACE: Inspection of rolled steel using photometric stereo acquisition

The **BLANKSURFACE** project aims to inspect rolled steel surfaces manufactured by Gonvarri. These steel sheets have a highly reflective and large surface to inspect and their manufacturing rate is high.

The defects that can appear on them are scale, rust, friction, corrosion, stains and coatings, among others. To solve this challenge, a stereo photometric illumination system and a linear camera were used. This provides a solution for inspecting large reflective surfaces without the influence of external light sources.

The images obtained are analysed using Resnet that analyses patches as the rolled steel moves along the production line. This specialised neural network detects and classifies them, so that the types of defects that are occurring can be monitored in near real time. An example of this classification is shown in Figure 4.5, where the detected defects are highlighted by red filled rectangles. This information can be exploited to find patterns of behaviour for predictive maintenance.

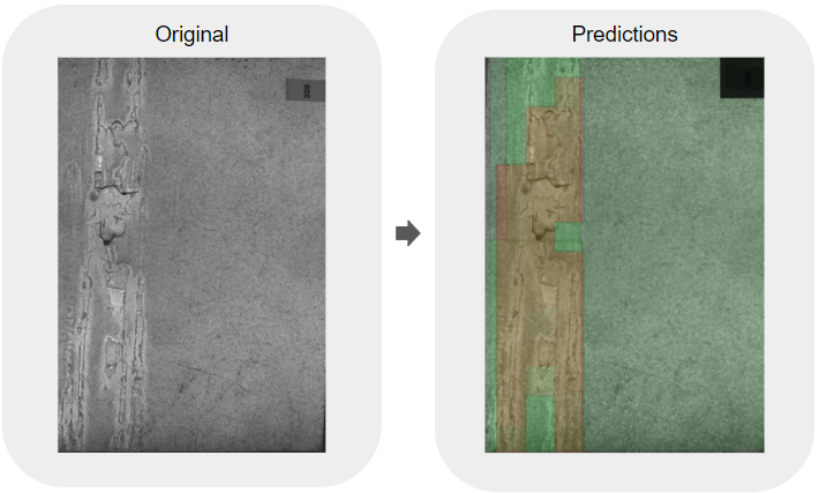


Figure 4.5: Defect classification in rolled steel sheets.

In this project, due to the nature of rolled steel manufacturing, the lightness, speed and robustness of the model were key factors. An AI-based method was developed that combines traditional Machine Learning techniques with convolutional neural networks. Multiple experiments were carried out in order to define the optimal parameters of the

network, and experiments were performed to test the robustness of the model to changes in illumination and occlusions conditions.

This method was also validated with a public dataset, the Northeastern University surface defect database (NEU), achieving a classification rate of 99.95% and taking 0.019 seconds to classify an image. A comparison was made with state-of-the-art methods, outperforming their results. These results can be seen in the publication *"A robust and fast deep learning-based method for defect classification in steel surfaces"*.

However, in the project dataset provided by the company the classes were not well balanced and, therefore, we needed to balance them to obtain a more robust system. In order to overcome this shortage of data, the common approach is to use conventional data augmentation techniques. We resort to Generative Adversarial Networks (GANs) that have shown the capability to generate highly convincing samples of a specific class as a result of a game between a discriminator and a generator module.In this way, we apply GANs to generate more samples of images with specific defects, in order to improve training of Semantic Networks (specifically DeepLabV3+ and Pyramid Attention Network (PAN) networks) carrying out the defect detection and segmentation. Our process carries out the generation of defect images using the StyleGAN2 with the DiffAugment method, followed by a conventional data augmentation over the entire enriched dataset, achieving a larger and well balanced dataset that allows a robust training of the Semantic Network.

We demonstrate that this approach is valid on a private dataset generated for the industrial customer where the images are acquired by an ad-hoc photometric-stereo image acquisition system and on the public NEU dataset. The proposed approach achieves an improvement of 7% and 6% in an intersection over union (IoU) measure of detection performance on each dataset over the conventional data augmentation. The results were published in the paper entitled *Generative Adversarial Networks to Improve the Robustness of Visual Defect Segmentation by Semantic Networks in Manufacturing Components.*

4.6 GLASSINSPECT: Accurate, high-speed automated optical inspection of automotive glass based on vision and Deep Learning techniques

The project **GLASSINSPECT** aims to solve the problem of high-speed and accurate automatic inspection of glass surfaces by means of machine vision and Artificial Intelligence. The project targets the automotive sector, specifically the surface inspection of the translucent part of cars. Obtained results can be easily extrapolated to the quality control of components with glass surfaces in other sectors, such as the aeronautical sector, the construction sector (domestic glass windows) or medical sector.

Surface inspection of translucent components has until now been carried out manually because of the complexity of the design of image acquisition systems and the algorithms required to automate it. Image acquisition systems for this application must cover a very wide field of view and acquire defects with very high detail. The image acquisition system used in this project has a very high resolution (16K) linear camera and a backlight illumination.

The main challenge of this project is that the inspection must be carried out in real time

within the imposed cycle time, which is 8 seconds. During these 8 seconds, it is necessary to acquire images, analyse and decide whether anomalies are found on the component. In addition, each defect on the surface must be located in order to know how it was originated and therefore to be able to correct the manufacturing process. The smallest defect in the catalogue of defects are small pores, measuring 0.4mm, and the largest ones are breaks that cross the entire component surface practically from side to side. An example of some defects is shown in Figure 4.6. The system is designed to be valid for multi-reference inspection, so that defects of different sizes, aspects, position and shapes can be detected on glass components of different sizes and prints.

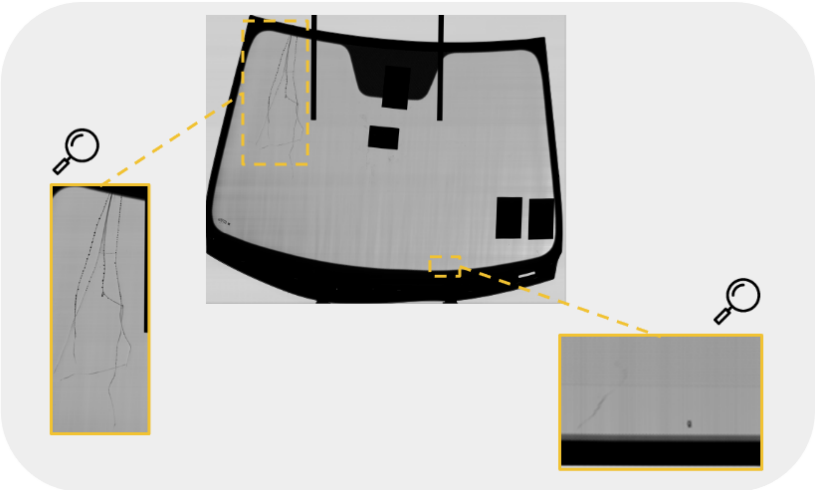


Figure 4.6: An example of car glass with different types of defects highlighted in yellow and zoomed.

In this project, transfer learning and data augmentation techniques were applied because of the limited number of available defective samples that does not allow a full training of the network. On the one hand, the training weights of YOLOv5 in Imagenet were used to obtain a pre-trained model to improve learning. On the other hand, the training database was augmented by following two strategies: the first one was to use images of similar defects thus adding these images to the training set and the second one was to cut out defects from other images and overlay them into background images to generate synthetic images that add variety to the set.

With this augmented set, YOLOv5 defect detection network initialised with pre-trained weights from Imagenet was trained. Thanks to the applied techniques, an improvement in defect detection was obtained, compared to the training performed with the original available samples, of approximately 15% in terms of accuracy.

4.7 COVID-19 detection

The SARS-CoV-2 coronavirus, which produces the disease known as COVID-19, kept the whole world confined in lockdowns during the first months of 2020.

During that time, access to COVID-19 tests was only available to healthcare professionals and their reliability was not very high. The scientific community began to investigate

multiple tasks such as the prediction of new waves of infection, new methods of detecting COVID, or medicines to alleviate its symptoms.

In that sense, it was decided to work on COVID detection using chest X-rays and AI-based techniques. Access to data was limited, and the available annotated databases were very small and from various sources around the world. To address this and minimise false positives for pneumonia, a new dataset combining COVID-19 and pneumonia radiographs was proposed to obtain a robust and reliable model. In addition, advanced image processing techniques were applied to normalise the images and improve the model learning process. Promising results were obtained, 94.92% of sensibility and 92.00% of specificity, which could be improved as the database is enriched.

Although this work is not directly related to industry, it demonstrates the versatility of the methods in both fields. Furthermore, the X-ray acquisition method is commonly used in industry also for internal quality control of components, as can be seen in the following works [125] [126].

CHAPTER 5

Conclusions

5.1 Conclusions

In this Thesis, several vision systems for metallic and therefore reflective components were developed. In terms of acquisition, it was demonstrated that a good acquisition set up has a direct impact on the results in terms of defect detection. Specifically, the use of a photometric stereo acquisition system has shown to provide additional information on these complex surfaces and, therefore, to extract more details of the defects to be detected. To verify this, a comparison analysis was performed for each photometric stereo image, which allows the most suitable combination of images to be chosen.

The combination of photometric stereo images provides additional information to the neural networks increasing the efficiency and accuracy of defect detection. This approach is confirmed on different materials showing a decrease in the false rejection rate and an increase in the true detection rate.

As well as these images provide extra information, the positive impact of processing and data augmentation has also demonstrated in all its variants: conventional, by using rendered images and by using GANs. The main conclusion is that the use of synthetically generated images significantly improves the performance of the networks. In the case of defect segmentation neural networks, obtaining realistic images to resolve the lack of samples in the training dataset improves the predictive power of the network.

The results show that this approach can be a good solution for dataset augmentation in an industrial environment, where difficulties in obtaining defective parts are very common. Even in the case of unbalanced databases in terms of defect classification, synthetic images can complement the information needed to detect and segment a certain type of defect, especially those related to the dimensional condition of the surface. In addition, the use of these images in the training of the neural model can increase its accuracy and help to generalise the process of detecting surface defects.

Experiments show that a 38% concentration of synthetically generated images improves the segmentation results compared to the original dataset using only conventional data augmentation. If there is any restriction in the real data acquisition process, the use of synthetic images could be a very powerful and effective tool as a data augmentation method

for the optimisation of the neural network learning process. Therefore, it is argued that such synthetically created images are fully compatible to be used in machine vision systems based on Deep Learning, and more specifically in semantic segmentation tasks.

As previously mentioned, the networks used in industries have to be lightweight and robust, in order to detect defects quickly and without error. Therefore, multiple experiments have been carried out to maximise the networks as much as possible. One of them takes advantage of the extra information provided by photometric stereo, for which a specific segmentation network was designed to meet these requirements. A performance comparison with different known segmentation architectures was performed to demonstrate the suitability of the proposed segmentation network. The benefits of assembling different models were also demonstrated, with the aim of diversifying problems and specialising each model in different tasks, so that both models support each other and reduce their shortcomings.

Finally, all developments were validated through experimentation and supervision by quality experts. These experiments were based on predicting the future behaviour of the system when it is in production, so that it was validated in realistic scenarios. Another objective of these experiments was to find the perfect fit in terms of sensitivity and specificity, allowing the user to fine-tune the output in terms of false negative and false positive rates.

Performance tests have also carried out in terms of run time. These tests demonstrate the need for accelerated hardware and parallel processing capabilities, such as making use of distributed computing devices and GPUs to meet the speed required on the production line. The need for an agile and easy-to-deploy software architecture, such as the containerised edge architecture, was also demonstrated.

Bibliography

[1] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2015. See pages vii, 21, 22, and 26.

[2] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single shot MultiBox detector. In *Computer Vision – ECCV 2016*, pages 21–37. Springer International Publishing, 2016. See pages vii, 23, and 24.

[3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn, 2017. See pages vii, 26.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs, 2014. See pages vii, 27.

[5] European Commission, Directorate-General for Research, Innovation, and Eurostat. National accounts and gdp, 2022. See page 1.

[6] European Commission, Directorate-General for Research, Innovation, and J Müller. *Enabling Technologies for Industry 5.0 : results of a workshop with Europe's technology leaders*. Publications Office, 2020. See page 1.

[7] European Commission, Directorate-General for Research, Innovation, M Breque, L De Nul, and A Petridis. *Industry 5.0 : towards a sustainable, human-centric and resilient European industry*. Publications Office, 2021. See page 1.

[8] Morteza Ghobakhloo. Industry 4.0, digitization, and opportunities for sustainability. *Journal of cleaner production*, 252:119869, 2020. See page 1.

[9] Amitava Mitra. *Fundamentals of quality control and improvement*. John Wiley & Sons, 2016. See page 2.

[10] T Calvin. Quality control techniques for" zero defects". *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, 6(3):323–328, 1983. See page 2.

[11] Jorge F Arinez, Qing Chang, Robert X Gao, Chengying Xu, and Jianjing Zhang. Artificial intelligence in advanced manufacturing: Current status and future outlook. *Journal of Manufacturing Science and Engineering*, 142(11), 2020. See page 2.

[12] Ricardo Silva Peres, Xiaodong Jia, Jay Lee, Keyi Sun, Armando Walter Colombo, and Jose Barata. Industrial artificial intelligence in industry 4.0-systematic review, challenges and outlook. *IEEE Access*, 8:220121–220139, 2020. See page 2.

[13] Jing Yang, Shaobo Li, Zheng Wang, Hao Dong, Jun Wang, and Shihao Tang. Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges. *Materials*, 13(24):5755, 2020. See page 2.

[14] Ajay Agrawal, Joshua S Gans, and Avi Goldfarb. Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2):31–50, 2019. See page 2.

[15] Raffaele Cioffi, Marta Travaglioni, Giuseppina Piscitelli, Antonella Petrillo, and Fabio De Felice. Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability*, 12(2):492, 2020. See page 3.

[16] Gerald Duft and Pavol Durana. Artificial intelligence-based decision-making algorithms, automated production systems, and big data-driven innovation in sustainable industry 4.0. *Economics, Management and Financial Markets*, 15(4):9–18, 2020. See page 3.

[17] Gopal K Kanji. An innovative approach to make iso 9000 standards more effective. *Total Quality Management*, 9(1):67–78, 1998. See page 5.

[18] B. Smith. Six-sigma design (quality control). *IEEE Spectrum*, 30(9):43–47, 1993. See page 5.

[19] Roland Caulcutt. Statistical process control (spc). *Assembly Automation*, 1996. See page 5.

[20] Robert S Jostes and Marilyn M Helms. Total productive maintenance and its link to total quality management. *Work study*, 1994. See page 6.

[21] Benjamin Osayawe Ehigie and Elizabeth B McAndrew. Innovation, diffusion and adoption of total quality management (tqm). *Management Decision*, 2005. See page 6.

[22] Pei Wang. On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37, 2019. See page 6.

[23] Xin Li and Yiliang Shi. Computer vision imaging based on artificial intelligence. In *2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)*, pages 22–25. IEEE, 2018. See page 7.

[24] Mauricio-Andrés Zamora-Hernández, John Alejandro Castro-Vargas, Jorge Azorin-Lopez, and Jose Garcia-Rodriguez. Deep learning-based visual control assistant for assembly in industry 4.0. *Computers in Industry*, 131:103485, 2021. See page 7.

[25] Jindong Zhang, Jiabin Xu, Linyao Zhu, Kunpeng Zhang, Tong Liu, Donghui Wang, and Xue Wang. An improved mobilenet-ssd algorithm for automatic defect detection on vehicle body paint. *Multimedia Tools and Applications*, 79(31):23367–23385, 2020. See page 7.

[26] Neha Sharma, Reecha Sharma, and Neeru Jindal. Machine learning and deep learning applications-a vision. *Global Transitions Proceedings*, 2(1):24–28, 2021. See page 7.

[27] Anouar Dalli. Impact of hyperparameters on deep learning model for customer churn prediction in telecommunication sector. *Mathematical Problems in Engineering*, 2022, 2022. See page 8.

[28] Jayme Garcia Arnal Barbedo. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and electronics in agriculture*, 153:46–53, 2018. See page 8.

[29] Steven Euijong Whang and Jae-Gil Lee. Data collection and quality challenges for deep learning. *Proceedings of the VLDB Endowment*, 13(12):3429–3432, 2020. See page 8.

[30] Fabio Henrique Kiyoiti dos Santos Tanaka and Claus Aranha. Data augmentation using gans. *arXiv preprint arXiv:1904.09135*, 2019. See pages 8, 20.

[31] Jun Kang Chow, Zhaoyu Su, Jimmy Wu, Pin Siang Tan, Xin Mao, and Yu-Hsing Wang. Anomaly detection of defects on concrete structures with the convolutional autoencoder. *Advanced Engineering Informatics*, 45:101105, 2020. See pages 8, 28.

[32] Petru Manescu, Priya Narayanan, Christopher Bendkowski, Muna Elmi, Remy Claveau, Vijay Pawar, Biobele J Brown, Mike Shaw, Anupama Rao, and Delmiro Fernandez-Reyes. Automated detection of acute promyelocytic leukemia in blood films and bone marrow aspirates with annotation-free deep learning. *arXiv preprint arXiv:2203.10626*, 2022. See page 8.

[33] Sujata Kaushik, Anjali Jain, Tina Chaudhary, and NR Chauhan. Machine vision based automated inspection approach for clutch friction disc (cfd). *Materials Today: Proceedings*, 2022. See page 8.

[34] Mina Boluki and Farahnaz Mohanna. Inspection of textile fabrics based on the optimal gabor filter. *Signal, Image and Video Processing*, 15(7):1617–1625, 2021. See page 8.

[35] Donggyun Im and Jongpil Jeong. R-cnn-based large-scale object-defect inspection system for laser cutting in the automotive industry. *Processes*, 9(11):2043, 2021. See page 8.

[36] Samira Gholizadeh. A review of non-destructive testing methods of composite materials. *Procedia structural integrity*, 1:50–57, 2016. See page 8.

[37] Robert E Green Jr. Non-contact ultrasonic techniques. *Ultrasonics*, 42(1-9):9–16, 2004. See page 8.

[38] Mohammad Javad Shafiee, Mahmoud Famouri, Gautam Bathla, Francis Li, and Alexander Wong. Tinydefectnet: Highly compact deep neural network architecture for high-throughput manufacturing visual quality inspection. *arXiv preprint arXiv:2111.14319*, 2021. See page 8.

[39] Dóra Farkas, Lajos Madarász, Zsombor K Nagy, István Antal, and Nikolett Kállai-Szabó. Image analysis: A versatile tool in the manufacturing and quality control of pharmaceutical dosage forms. *Pharmaceutics*, 13(5):685, 2021. See page 8.

[40] Elena Trajkova, Jože M Rožanec, Paulien Dam, Blaž Fortuna, and Dunja Mladenić. Active learning for automated visual inspection of manufactured products. *arXiv preprint arXiv:2109.02469*, 2021. See page 8.

[41] Chris H Bahnsen, Anders S Johansen, Mark P Philipsen, Jesper W Henriksen, Kamal Nasrollahi, and Thomas B Moeslund. 3d sensors for sewer inspection: A quantitative review and analysis. *Sensors*, 21(7):2553, 2021. See page 10.

[42] C Marc Bastuscheck. Techniques for real-time generation of range images. In *1989 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 262–263. IEEE Computer Society, 1989. See page 10.

[43] Giovanna Sansoni, Marco Trebeschi, and Franco Docchio. State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, 9(1):568–601, 2009. See page 10.

[44] Richard Fox-Ivey, Benoit Petitclerc, and John Laurent. 3d laser triangulation and deep learning approach to tunnel inspection. *Transportation research record*, 2676(4):393–406, 2022. See page 10.

[45] Finn Renken, Jan Schubnell, Matthias Jung, Markus Oswald, Klemens Rother, Sören Ehlers, Moritz Braun, et al. An algorithm for statistical evaluation of weld toe geometries using laser triangulation. *International Journal of Fatigue*, 149:106293, 2021. See page 10.

[46] Tyler Bell, Beiwen Li, and Song Zhang. Structured light techniques and applications. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–24, 1999. See page 10.

[47] Zemin Mao, Dahua Li, and Xuewen Zhao. Structured light-based dynamic 3d measurement system for cold-formed steel hollow sections. *International Journal of Mechatronics and Manufacturing Systems*, 15(2-3):203–225, 2022. See page 10.

[48] Iker Garmendia, Joseba Pujana, Aitzol Lamikiz, Mikel Madarieta, and Josu Leunda. Structured light-based height control for laser metal deposition. *Journal of Manufacturing Processes*, 42:20–27, 2019. See page 10.

[49] H Golnabi and A Asadpour. Design and application of industrial machine vision systems. *Robotics and Computer-Integrated Manufacturing*, 23(6):630–637, 2007. See pages 11, 20.

[50] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000. See page 12.

[51] Bing Pan, Liping Yu, and Dafang Wu. High-accuracy 2d digital image correlation measurements with bilateral telecentric lenses: error analysis and experimental verification. *Experimental Mechanics*, 53(9):1719–1733, 2013. See page 12.

[52] Julie Lemesle, Frederic Robache, Gaetan Le Goic, Alamin Mansouri, Christopher A. Brown, and Maxence Bigerelle. Surface reflectance: An optical method for multiscale curvature characterization of wear on ceramic–metal composites. *Materials*, 13(5), 2020. See page 13.

[53] Amir Novini. Fundamentals of on-line gauging for machine vision. In Armin Gruen and Emmanuel P. Baltsavias, editors, *Close-Range Photogrammetry Meets Machine Vision*, volume 1395, page 13952M. International Society for Optics and Photonics, SPIE, 1990. See page 13.

[54] Robert J Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Image Understanding Systems and Industrial Applications I*, volume 155, pages 136–143. International Society for Optics and Photonics, 1979. See pages 14, 16.

[55] Berthold KP Horn and Michael J Brooks. *Shape from shading*. MIT press, 1989. See page 14.

[56] Reinhard Klette, Ryszard Kozera, and Karsten Schlüns. Shape from shading and photometric stereo methods. *Handbook of computer vision and applications*, 2:531–590, 1998. See page 14.

[57] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25:1239–1252, 2003. See page 16.

[58] Roger Penrose. A generalized inverse for matrices. In *Mathematical proceedings of the Cambridge philosophical society*, volume 51, pages 406–413. Cambridge University Press, 1955. See page 16.

[59] Saman Fattahi, Takuya Okamoto, and Sharifu Ura. Preparing datasets of surface roughness for constructing big data from the context of smart manufacturing and cognitive computing. *Big Data and Cognitive Computing*, 5(4):58, 2021. See page 16.

[60] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019. See page 17.

[61] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era, 2017. See pages 17, 18.

[62] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation, 2013. See page 17.

[63] Pulkit Agrawal, Ross Girshick, and Jitendra Malik. Analyzing the performance of multilayer neural networks for object recognition, 2014. See page 17.

[64] Mihir Jain, Jan C Van Gemert, and Cees GM Snoek. What do 15,000 object categories tell us about classifying and localizing actions? In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 46–55, 2015. See page 17.

[65] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, January 2015. See page 17.

[66] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. See page 17.

[67] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2014. See page 17.

[68] Trond Linjordet and Krisztian Balog. Impact of training dataset size on neural answer selection models. In *European Conference on Information Retrieval*, pages 828–835. Springer, 2019. See page 17.

[69] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. See pages 18, 20.

[70] Bojan Pepik, Rodrigo Benenson, Tobias Ritschel, and Bernt Schiele. What is holding back convnets for detection? In *German conference on pattern recognition*, pages 517–528. Springer, 2015. See page 19.

[71] Mona Jalal, Josef Spjut, Ben Boudaoud, and Margrit Betke. Sidod: A synthetic image dataset for 3d object pose recognition with distractors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. See page 19.

[72] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Reading text in the wild with convolutional neural networks. *International journal of computer vision*, 116(1):1–20, 2016. See page 19.

[73] Yair Movshovitz-Attias, Takeo Kanade, and Yaser Sheikh. How useful is photo-realistic rendering for visual learning? In *European Conference on Computer Vision*, pages 202–217. Springer, 2016. See page 19.

[74] Saul Alexis Heredia Perez, Murilo Marques Marinho, Kanako Harada, and Mamoru Mitsuishi. The effects of different levels of realism on the training of cnns with only synthetic images for the semantic segmentation of robotic instruments in a head phantom. *International Journal of Computer Assisted Radiology and Surgery*, 15:1257–1265, 2020. See page 19.

[75] Daniele Evangelista, Marco Antonelli, Alberto Pretto, Christian Eitzinger, Michele Moro, Carlo Ferrari, and Emanuele Menegatti. Spirit-a software framework for the efficient setup of industrial inspection robots. In *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*, pages 622–626. IEEE, 2020. See page 19.

[76] Qi Wang, Junyu Gao, Wei Lin, and Yuan Yuan. Learning from synthetic data for crowd counting in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. See page 19.

[77] Nerea Aranjuelo, Sara García, Estíbaliz Loyo, Luis Unzueta, and Oihana Otaegui. Key strategies for synthetic data generation for training intelligent systems based on people detection from omnidirectional cameras. *Computers & Electrical Engineering*, 92:107105, 2021. See page 19.

[78] Igor Garcia Ballhausen Sampaio, Luigy Machaca, José Viterbo, and Joris Guérin. A novel method for object detection using deep learning and CAD models. 2021. See page 19.

[79] SEGMENTING UNSEEN INDUSTRIAL COMPONENTS IN A HEAVY CLUTTER USING RGB-D FUSION AND SYNTHETIC DATA Seunghyeok Back, Jongwon Kim, Raeyoung Kang, Seungjun Choi, Kyoobin Lee Gwangju Institute of Science and Technology (GIST), Republic of Korea. 2020. See page 19.

[80] Jan Lehr, Alik Sargsyan, Martin Pape, Jan Philipps, and Jorg Kruger. Automated Optical Inspection Using Anomaly Detection and Unsupervised Defect Clustering. pages 1235–1238, 2020. See page 19.

[81] Roey Ron and Gil Elbaz. Expo-hd: Exact object perception usinghigh distraction synthetic data. *arXiv preprint arXiv:2007.14354*, 2020. See page 19.

[82] Fabio Henrique Kiyoiti dos Santos Tanaka and Claus Aranha. Data augmentation using gans. *arXiv preprint arXiv:1904.09135*, 2019. See page 20.

[83] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger,

editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. See pages 20, 25.

[84] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014. See pages 20, 25.

[85] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation, 2013. See page 21.

[86] Ross Girshick. Fast r-cnn, 2015. See page 21.

[87] Zhiyong Zhao, Kang Gui, and Peimao Wang. Fabric defect detection based on cascade faster r-cnn. In *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, CSAE 2020, New York, NY, USA, 2020. Association for Computing Machinery. See page 21.

[88] Yuanbin Wang, Minggao Liu, Pai Zheng, Huayong Yang, and Jun Zou. A smart surface inspection system using faster r-cnn in cloud-edge computing environment. *Advanced Engineering Informatics*, 43:101037, 2020. See page 21.

[89] Donggyun Im and Jongpil Jeong. R-cnn-based large-scale object-defect inspection system for laser cutting in the automotive industry. *Processes*, 9(11), 2021. See page 21.

[90] Ming-Wei Liu, Yu-Heng Lin, Yuan-Chieh Lo, Chih-Hsuan Shih, and Pei-Chun Lin. Defect detection of grinded and polished workpieces using faster r-cnn. In *2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 1290–1296, 2021. See page 21.

[91] Gang ZHAO, Jingyu HU, Wenlei XIAO, and Jie ZOU. A mask r-cnn based method for inspecting cable brackets in aircraft. *Chinese Journal of Aeronautics*, 34(12):214–226, 2021. See page 22.

[92] Haisong Huang, Zhongyu Wei, and Liguo Yao. A novel approach to component assembly inspection based on mask r-cnn and support vector machines. *Information*, 10(9), 2019. See page 22.

[93] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection, 2015. See page 22.

[94] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement, 2018. See pages 22, 23.

[95] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger, 2016. See page 23.

[96] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection, 2020. See page 23.

[97] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, TaoXie, Kalen Michael, Jiacong Fang, imyhxy, Lorna, Colin Wong, [U+66FE][U+9038][U+592B](Zeng Yifu), Abhiram V, Diego Montes, Zhiqiang Wang, Cristi Fati, Jebastin Nadar, Laughing, UnglvKitDe, tkianai, yxNONG, Piotr Skalski, Adam Hogan, Max Strobel, Mrinal Jain, Lorenzo Mammana, and xylieong. ultralytics/yolov5: v6.2 - YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai integrations, August 2022. See page 23.

[98] Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, Yiduo Li, Bo Zhang, Yufei Liang, Linyuan Zhou, Xiaoming Xu, Xiangxiang Chu, Xiaoming Wei, and Xiaolin Wei. Yolov6: A single-stage object detection framework for industrial applications, 2022. See page 23.

[99] Muhieddine Hatab, Hossein Malekmohamadi, and Abbes Amira. Surface defect detection using yolo network. In *Proceedings of SAI Intelligent Systems Conference*, pages 505–515. Springer, 2020. See page 23.

[100] Jing Li, Jinan Gu, Zedong Huang, and Jia Wen. Application research of improved yolo v3 algorithm in pcb electronic component detection. *Applied Sciences*, 9(18):3750, 2019. See page 23.

[101] Longzhen Yu, Jianhua Zhu, Qian Zhao, and Zhixian Wang. An efficient yolo algorithm with an attention mechanism for vision-based defect inspection deployed on fpga. *Micromachines*, 13(7):1058, 2022. See page 23.

[102] Shwe Lamin Aein, Theint Theint Thu, Phyu Phyu Htun, Aung Paing, and Hay Thar Myo Htet. Yolo based deep learning network for metal surface inspection system. In *Proceedings of the 11th International Conference on Robotics, Vision, Signal Processing and Power Applications*, pages 923–929. Springer, 2022. See page 23.

[103] Lu Tan, Tianran Huangfu, Liyao Wu, and Wenying Chen. Comparison of retinanet, ssd, and yolo v3 for real-time pill identification. *BMC Medical Informatics and Decision Making*, 21(1):1–11, 2021. See page 23.

[104] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection, 2017. See page 23.

[105] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2016. See page 25.

[106] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014. See page 25.

[107] Yongjian Zhu, Chuliu Tang, Hao Liu, and Pengchi Huang. End-face localization and segmentation of steel bar based on convolution neural network. *IEEE Access*, 8:74679–74690, 2020. See page 25.

[108] Chuanzhi Dong, Liangding Li, Jin Yan, Zhiming Zhang, Hong Pan, and Fikret Necati Catbas. Pixel-level fatigue crack segmentation in large-scale images of steel structures using an encoder–decoder network. *Sensors*, 21(12):4135, 2021. See page 25.

[109] Zhihang Li, Huamei Zhu, and Mengqi Huang. A deep learning-based fine crack segmentation network on full-scale steel bridge images with complicated backgrounds. *IEEE Access*, 9:114989–114997, 2021. See page 25.

[110] ZHAO Gang, HU Jingyu, XIAO Wenlei, and ZOU Jie. A mask r-cnn based method for inspecting cable brackets in aircraft. *Chinese Journal of Aeronautics*, 34(12):214–226, 2021. See page 26.

[111] Hamed Raoofi and Ali Motamedi. Mask r-cnn deep learning-based approach to detect construction machinery on jobsites. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 37, pages 1122–1127. IAARC Publications, 2020. See page 26.

[112] Agit Çelik, Yağız Uğurlu, Emre Gözüküçük, and Elena Battini Sönmez. Automatic system for sheepskin quality control with convolutional neural network. In *2019 4th International Conference on Computer Science and Engineering (UBMK)*, pages 1–5. IEEE, 2019. See page 26.

[113] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, 2016. See page 27.

[114] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017. See page 27.

[115] Zhaomin Chen, Chai Kiat Yeo, Bu Sung Lee, and Chiew Tong Lau. Autoencoder-based network anomaly detection. In *2018 Wireless telecommunications symposium (WTS)*, pages 1–5. IEEE, 2018. See page 28.

[116] Siti Farhana Lokman, Abu Talib Othman, Shahrulniza Musa, and Muhamad Husaini Abu Bakar. Deep contractive autoencoder-based anomaly detection for in-vehicle controller area network (can). In *Progress in Engineering Technology*, pages 195–205. Springer, 2019. See page 28.

[117] Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2(1):1–18, 2015. See page 28.

[118] Daniel Jakubovitz and Raja Giryes. Improving dnn robustness to adversarial attacks using jacobian regularization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 514–529, 2018. See page 28.

[119] Benjamin Lindemann, Benjamin Maschler, Nada Sahlab, and Michael Weyrich. A survey on anomaly detection for technical systems using lstm networks. *Computers in Industry*, 131:103498, 2021. See page 28.

[120] Benjamin Lindemann, Nasser Jazdi, and Michael Weyrich. Anomaly detection and prediction in discrete manufacturing based on cooperative lstm networks. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, pages 1003–1010. IEEE, 2020. See page 28.

[121] Byeonggeun Choi and Jongpil Jeong. Viv-ano: Anomaly detection and localization combining vision transformer and variational autoencoder in the manufacturing process. *Electronics*, 11(15):2306, 2022. See page 29.

[122] Nejc Kozamernik and Drago Bračun. Visual inspection system for anomaly detection on ktl coatings using variational autoencoders. *Procedia CIRP*, 93:1558–1563, 2020. See page 29.

[123] Arnaud Bougaham, Adrien Bibal, Isabelle Linden, and Benoit Frenay. Ganodip-gan anomaly detection through intermediate patches: a pcba manufacturing case. In *Third International Workshop on Learning with Imbalanced Domains: Theory and Applications*, pages 104–117. PMLR, 2021. See page 29.

[124] Ta-Wei Tang, Wei-Han Kuo, Jauh-Hsiang Lan, Chien-Fang Ding, Hakiem Hsu, and Hong-Tsu Young. Anomaly detection neural network with dual auto-encoders gan and its industrial inspection applications. *Sensors*, 20(12):3336, 2020. See page 29.

[125] Jitendra Singh Rathore and Prateek Saxena. Non-destructive quality assessment of bio-engineering parts using industrial micro x-ray computed tomography: A review. *Materials Letters*, 287:129252, 2021. See page 46.

[126] Dhruv Gamdha, Sreedhar Unnikrishnakurup, KJ Rose, M Surekha, Padma Purushothaman, Bikash Ghose, and Krishnan Balasubramaniam. Automated defect recognition on x-ray radiographs of solid propellant using deep learning based on convolutional neural networks. *Journal of Nondestructive Evaluation*, 40(1):1–13, 2021. See page 46.

Appendix

applied sciences

MDPI

Article

Generative Adversarial Networks to Improve the Robustness of Visual Defect Segmentation by Semantic Networks in Manufacturing Components

Fátima A. Saiz [1,2,*], Garazi Alfaro [1], Iñigo Barandiaran [1] and Manuel Graña [2]

[1] Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Donostia, 20009 San Sebastián, Spain; galfaro@vicomtech.org (G.A.); ibarandiaran@vicomtech.org (I.B.)
[2] Computational Intelligence Group, Computer Science Faculty, University of the Basque Country, UPV/EHU, 20018 San Sebastián, Spain; manuel.grana@ehu.es
* Correspondence: fsaiz@vicomtech.org

Abstract: This paper describes the application of Semantic Networks for the detection of defects in images of metallic manufactured components in a situation where the number of available samples of defects is small, which is rather common in real practical environments. In order to overcome this shortage of data, the common approach is to use conventional data augmentation techniques. We resort to Generative Adversarial Networks (GANs) that have shown the capability to generate highly convincing samples of a specific class as a result of a game between a discriminator and a generator module. Here, we apply the GANs to generate samples of images of metallic manufactured components with specific defects, in order to improve training of Semantic Networks (specifically DeepLabV3+ and Pyramid Attention Network (PAN) networks) carrying out the defect detection and segmentation. Our process carries out the generation of defect images using the StyleGAN2 with the DiffAugment method, followed by a conventional data augmentation over the entire enriched dataset, achieving a large balanced dataset that allows robust training of the Semantic Network. We demonstrate the approach on a private dataset generated for an industrial client, where images are captured by an ad-hoc photometric-stereo image acquisition system, and a public dataset, the Northeastern University surface defect database (NEU). The proposed approach achieves an improvement of 7% and 6% in an intersection over union (IoU) measure of detection performance on each dataset over the conventional data augmentation.

Keywords: defect segmentation; data augmentation; generative adversarial networks; industrial manufacturing; quality inspection; photometric stereo

Citation: Saiz, F.A.; Alfaro, G.; Barandiaran, I.; Graña, M. Generative Adversarial Networks to Improve the Robustness of Visual Defect Segmentation in Manufacturing Components by Semantic Networks. *Appl. Sci.* **2021**, *11*, 6368. https://doi.org/10.3390/app11146368

Academic Editor: Manuel Armada

Received: 25 May 2021
Accepted: 5 July 2021
Published: 9 July 2021

1. Introduction

In the manufacturing sector, any process needs continuous monitoring to ensure their behavior and performance. In the case of components manufacturing for sub-sectors such as the automotive industry, quality inspection is an imperative as the final quality of the components can affect their functionality. Component manufacturers are continuously looking for innovative quality inspection mechanisms that are the most efficient and cost-effective strategy in order to survive in a very competitive market.

The Industry 4.0 paradigm has brought great opportunities to improve and manage quality control processes through the application of new visual computing related technologies such as computer vision or artificial intelligence [1].

This work focuses on steel quality control. Steel has multiple uses in the construction of buildings and infrastructures or in the manufacturing sector [2]. In the manufacturing sector, the quality control processes carried out on steel products comprises the following phases according to the defects to be located:

- A visual analysis by the operator (Superficial defects);

- Studies on its chemical composition (Structural defects);
- Studies on its mechanical characteristics (Structural defects);
- Study of geometric characteristics (Dimensional defects).

The component quality control process has traditionally been done by applying statistical process control (SPC), where various evaluation tests are applied to a component sampled from the production line at a given sampling frequency. Test results are then compared with acceptance/rejection criteria established according to the regulations or customer requirements. Nowadays, due to the competitiveness of the markets, customer requirements are becoming more and more demanding, aiming to achieve 100% inspection of the production instead of SPC.

In the steel component production workflow, surface defects can occur at several stages. The wide variety of surface defects that can occur for all different existing steel products makes classification difficult. However, most of these defects involve some type of surface roughness.To cope with the needs to inspect 100% of the production at high production rates, and the new requirement of maximum precision in the detection of defects, Deep Learning (DL) based models are becoming the predominant approach to tackle the automatic defect inspection problem [3]. Surface inspection data from the manufacturing industry are usually highly imbalanced because defects are typically low frequency events. It is easy to understand that, for a company to be profitable, most of the manufactured products must be free from defects. Therefore, a database extracted directly from the production line will be composed mostly of non-defective products. Using such type of datasets can lead to unwanted results, due to the extreme imbalance of the classes in the data [4,5]. Machine learning model building approaches generally have a strong bias towards the majority class in the case of imbalanced datasets. In order to correct this effect, the generation of synthetic samples of the minority class is often used to obtain a better balanced dataset [6]. The generation of realistic images containing surface defects can not be easily achieved by conventional data augmentation methods.

1.1. Deep Learning Based Methods for Defect Detection

There are different types of DL techniques that are used to detect defects in images [7,8]. Some approaches such as the *Improved You Only Look Once* (YOLO) network [9] focus on object detection, providing a short response time which is very important in the industrial manufacturing environment in order to cope with high production rates. YOLO reported a 99% detection accuracy with a speed of 83 FPS on NVIDIA GTX 1080Ti GPU. Another example of real-time detection using these types of architectures is [10], which uses an Inception-based MobileNet-SSD architecture, reporting an accuracy of 96.1% on the DAGM 2007 [11] test dataset at 73FPS on a NVIDIA GTX 1080Ti, outperforming a comparable state-of-the-art fully convolutional network (FCN) model.

Depending on the application, a lot of precision regarding the location and the size or the geometry of the defects is required. Many approaches use deep segmentation networks for this purpose. For instance, the authors in [12] propose an end-to-end UNet-shaped fully convolutional neural network for automated defect detection in surfaces of manufactured components. They report results over a publicly available 10 class dataset applying a real-time data augmentation approach during the training phase, achieving a mean intersection over union (IoU) of 68.35%. There are different works analyzing the performance of segmentation networks for superficial quality control [13,14], where DeepLabv3+ reaches the best performance metric values in the defect segmentation.

In general, defect detection results worsen as the number of training samples decreases because the similarities between defects are usually small so the trained models do not generalize well from small datasets. For example, the authors in [5] present two regularization techniques via incorporating abundant defect-free images into the training of a UNet-like encoder–decoder defect segmentation network. With their proposed method, they achieve a 49.19% IoU on a 1-shot setting (K = 1) and 64.45% IoU on a 5-shot setting (K = 5).

1.2. The Issue of Dataset Generation

A major issue is the difficulty of obtaining good quality images due to the complex geometries of the parts or very reflective or shiny materials, and the lighting variations that occur in a normal industrial environment. Thus, creating a stable robust machine vision solution is always a complex task [15–17].

In addition, getting samples of products with defects is not always possible. Due to the high production quality achieved by the manufacturing companies, most of the components will be defect-free. It must also be considered that sometimes the production rate is very high and it makes it difficult to locate defective parts. This is a common problem when collecting datasets for training defect detection models applied to industrial parts [4]. Therefore, the generation of annotated synthetic sample sets is an increasingly promising research area for the machine vision community.

1.3. Advances in Training Data Synthesis

Synthetic data can be obtained in several ways. On the one hand, realistically rendered images can be produced from accurate geometrical and photo-metrical, i.e., physical, models of the real environment where the artificial vision system would be installed [18,19]. Building such an accurate model requires a high amount of human involvement. However, the samples generated are usually very good in terms of visual fidelity. In addition, the annotation masks are automatically generated along with the images due to the rendering process; therefore, they are highly useful for machine learning model training.

On the other hand, generative approaches can be applied to learn how to synthesize realistic images from data without an underlying detailed physical model. Recent works use Generative Adversarial Networks (GANs) [20] for this task, which demand less human involvement because both the discriminator learning process and the image generation are fully automatic. Contrary to rendering based methods, the image annotation task needs to be done manually in a separate process.

For example, semantic networks for image segmentation improve their performance when data augmentation of the training dataset is carried out with synthetic images rendered from 3D models [21]. Specifically, authors report a significant increase in IoU from 52.80% to 55.47% when training the system over the PASCAL 2012 dataset enriched with 100 rendered artificial images.

In [22], the authors propose a new GAN based architecture for evaluating the effectiveness of defective image generation using different public datasets for surface defect segmentation. Another study [23] on defect detection in steel plates using deep learning architectures also applies GANs for training set extension. They report improved detection results with Fast R-CNN [24] and YOLO neural networks as surface defect detectors.

1.4. Contributions in This Paper

In this paper, GANs will be applied for augmenting the dataset. More precisely, we propose to use StyleGAN2+DiffAugment [25] precisely, showing improvements over conventional approaches. The benefit of this augmentation approach will be validated in a semantic deep convolutional network training, specifically with DeepLabv3+ [26]. We will also evaluate the repeatability of the obtained benefits by replicating the process with another segmentation network, specifically Pyramid Attention Network (PAN) [27] and on a well-known public dataset such as the NEU dataset [28].

2. Materials and Methods

2.1. Defects in Steel Surfaces

There are various types of steel that offer different results depending on their composition, thus being able to obtain a material that is more resistant to temperature, impact, corrosion, etc. In our case, we focus on rolled steel. The rolling process is a metal forming process by plastic deformation that consists of being passed between two or more rolls that rotate in opposite directions and apply pressure. There are two types of rolling: cold

rolling and hot rolling. These two types of processes obtain different results, so the choice of the type depends on the intended application. The main difference between both types is the temperature at which the steel is produced, varying from more than 927 ºC in the case of hot rolling to ambient temperature in the case of cold rolling.

In this rolling process, different surface defects can occur such as finishing roll printing, oxide, patches, scratches, crazing, or inclusions. An example of these defects can be shown in Figure 1, shown in red and dotted lines.

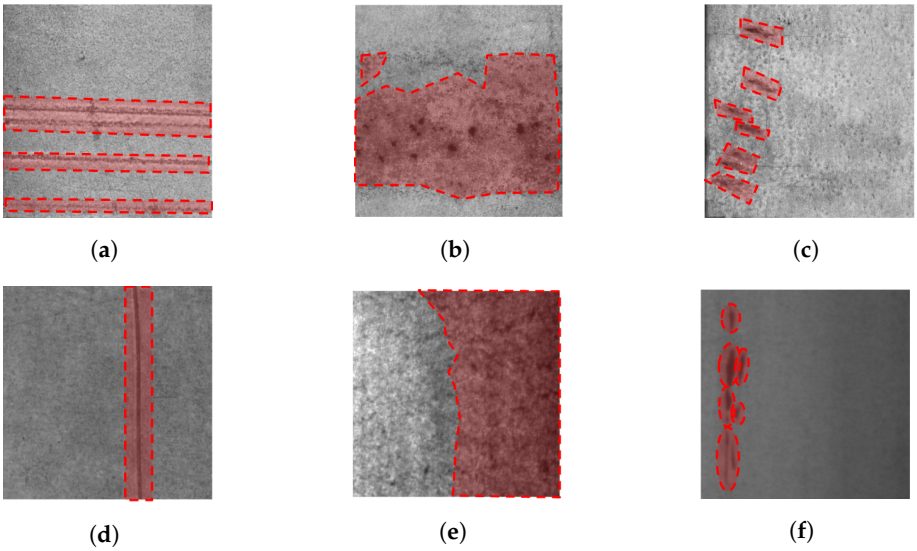


Figure 1. Samples of four types of steel surface defects. (**a**) finishing roll printing; (**b**) oxide; (**c**) patches; (**d**) scratches; (**e**) crazing; and (**f**) inclusions.

2.2. Galvanized Steel Dataset Description

For the experimental validation of the proposed approach, we use a dataset of 204 images of galvanized steel surfaces provided by a steel manufacturing company, which was seeking technological advice on their specific defect detection problems. The imaged material samples show natural or manufacturing defects discussed in Section 2.1.

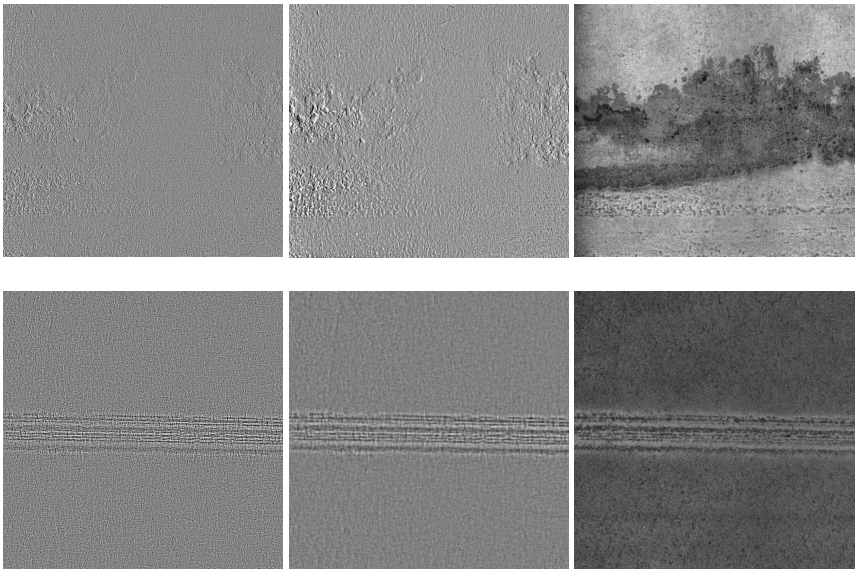
Images were acquired using a photometric-stereo system [29,30], as shown in Figure 2. Image acquisition consists of taking several images with different lighting orientations. From these images, it is possible to compute a three-dimensional geometrical model of the object using a shape-from-shading approach.

Figure 2. Photometric-stereo image acquisition system.

The photometric-stereo system obtains three high resolution computed images for each steel sheet, as shown in Figure 3. A brief description of every computed image follows:

- Curvature images: Provides topographic information of the two gradient images in the x- and y-directions. May be used for the verification of local defects such as scratches, impact marks, etc.
- Texture images: Provides information on surface gloss. They are very suitable for detecting discoloration defects and rust damage.

- Range images: Computed as the image gradient magnitude. It highlights information about the changes in the intensity of the image.



(**a**) (**b**) (**c**)

Figure 3. Different image types obtained by the photometric-stereo system of two different steel sheet samples (one sample in each row); (**a**) curvature image; (**b**) range image; and (**c**) texture image.

From Figure 3, it can be ascertained that each of the three images is suitable for detecting a specific type of defect. Therefore, the system provides the necessary visual information to reveal different imperfections that may be present in the surface of the galvanized steel products. We propose to combine these three single channel images in a RGB format or three layer based images in which each layer corresponds to one such image, i.e., red channel corresponds to curvature image, green channel corresponds to range image, and, finally, the blue channel stores the texture image. In this way, we are collecting the information given by each individual image in one single image.

The dataset annotation task was carried out manually by a human operator who is an expert in the inspection of defects in the galvanized steel products. The annotations have different shapes and sizes, ideally fitting or resembling the geometry of the defects. Two samples of the texture channel with their corresponding ground truth masks in red identifying the defects are shown in Figure 4.

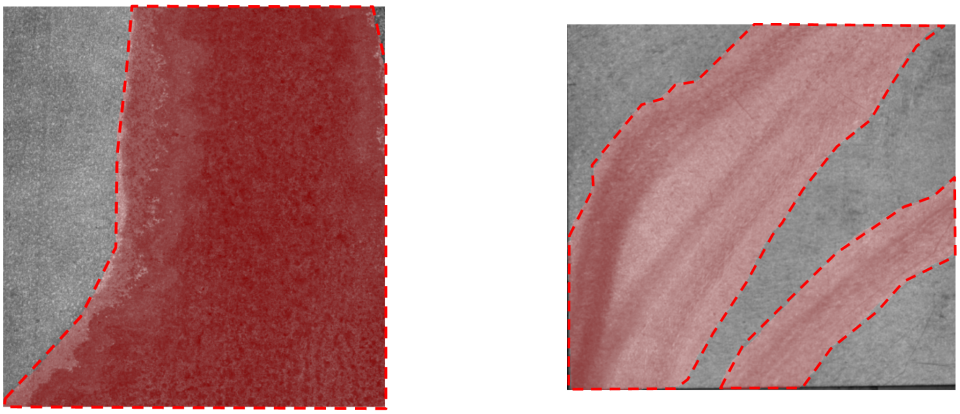


Figure 4. Texture channel of RGB photometric-stereo images with corresponding ground truth masks in red and dotted lines.

2.3. Experimental Design

We use a GAN based approach to generate new photometric-stereo RGB images that will augment the available dataset. Specifically, we use StyleGAN2 in combination with the DiffAugment method. In order to obtain the ground-truth defect masks for the GAN-based generated images, we apply a DeepLabV3+ segmentation network [26] already trained on the original dataset of photometric-stereo RGB images. Once the defect masks are obtained, we then re-train the DeepLabV3+ segmentation network by combining synthetic and original images in different ratios, reporting the impact on the dataset enrichment on the segmentation results. In order to assess the general applicability of our approach, we have (a) trained a different architecture on our private dataset, and (b) applied the entire process on a publicly available dataset that allows independent confirmation of our results.

2.3.1. Conventional Data Augmentation

As mentioned before, in each channel of the RGB image, the texture image, the curvature image, and the range image are stored. We apply a conventional data augmentation approach to obtain a fifteen-fold augmentation of the training dataset obtaining a total of 2445 images. Conventional data augmentation consists of randomly applying a series of geometrical and photometrical transformations to original images. The chosen transformations are:

- Horizontal and Vertical Shift transformation
- Horizontal and Vertical Flip transformation
- Random image rotation transformation (in a range between 0 to 180 degrees)
- Random scale transformation (maximum value 2x)
- Addition of Gaussian noise
- Brightness modification

2.3.2. Defect Segmentation Using DeepLabV3+

Semantic segmentation Deep Learning architectures are usually based on an encoder–decoder structure [31]. The encoder reduces the size of the image after passing through the Maxpool grouping layers in the convolution stage. After this stage comes the decoding phase, which consists of gradually recovering the spatial information until it reaches the same dimensions as the input image. The network output is an image corresponding to the pixelwise classification of the image into defect and defect-free classes, where each intensity value encodes a label of a particular class. In our case, the classification is binary, between the "defect" label with a numerical value of 1 and the "background" or "non-defect" label with a value of 0. To evaluate the semantic segmentation results, we use the Intersection-Over-Union (IoU) metric measuring the overlap between predicted and ground truth defect masks. Accuracy is also reported, though it is biased by the greater number of pixels in the non-defect or background class.

Figure 5 depicts the DeepLabv3+ architecture [26]. It uses a CNN called Xception with Atrous Convolution layers to get the coarse score map and then a conditional random field is used to produce the final output. This architecture has some peculiarities such as the use of the aforementioned Atrous Spatial Pyramid Pooling (ASPP) [32–37] based on the Atrous Separable convolution [38,39]. This network has a simple yet effective decoder module to refine the segmentation results, especially along object boundaries.

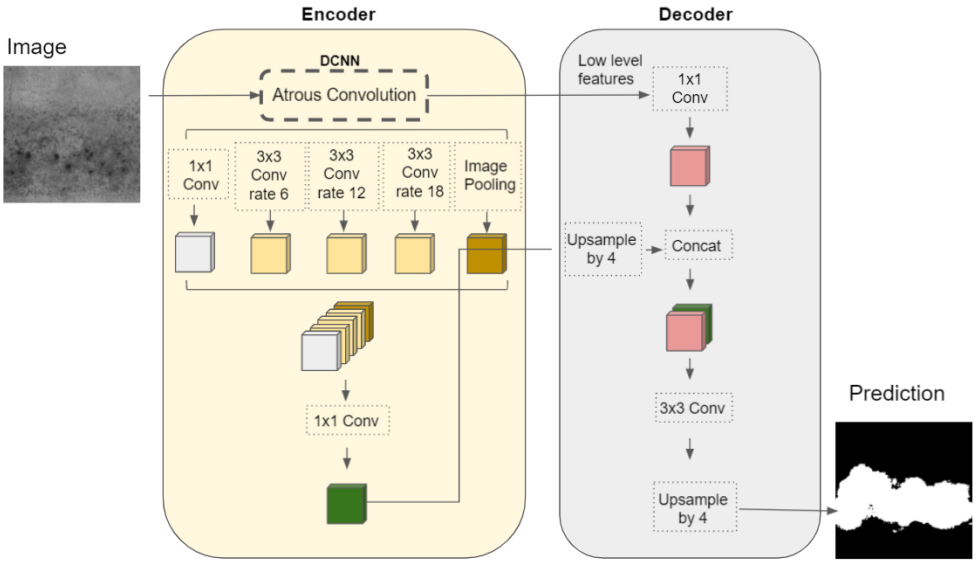


Figure 5. Architecture diagram of the DeepLabV3+ network [26].

2.3.3. GAN Based Image Generation

The basic architecture of GANs is composed of two networks that pursue opposite optimization goals regarding a loss function. On the one hand, the generating network produces artificial data as close as possible to reality trying to mislead the discriminant network. On the other hand, the discriminating network tries to determine whether the input image is real or fake (i.e., generated by the generating network). In this way, the training process pursues the minimization of an overall loss function that measures recognition performance. If the process is successful, the fake images are quite convincing surrogates of the real ones.

We use the StyleGAN2 architecture [40] combined with the DiffAugment [25] method. StyleGAN2 is an improved GAN network in terms of existing distribution quality metrics as well as perceived image quality. DiffAugment uses a simple method that improves the data efficiency of GANs by imposing various types of differentiable augmentations on both real and fake samples. This combination allows for stabilizing training and leads to better convergence in comparison with previous attempts that directly augment the training data, manipulating the distribution of real images.

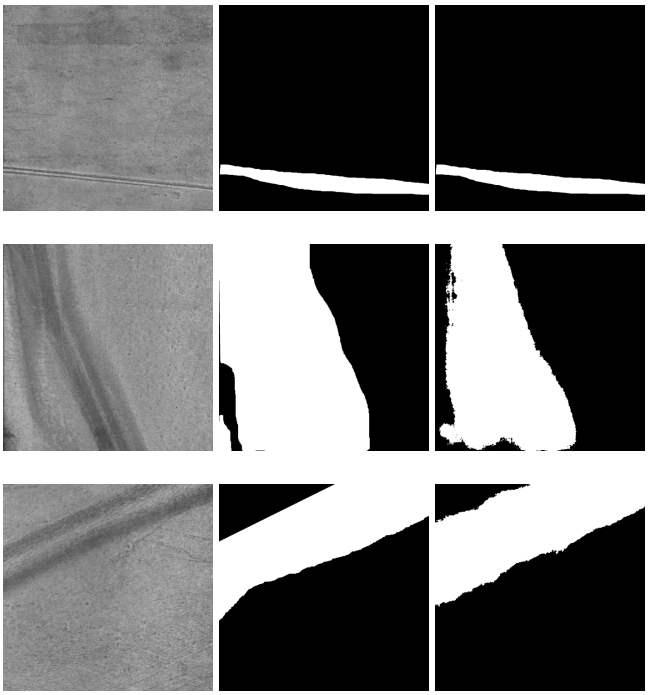
3. Results and Discussion

In this section, different experiments are presented in order to demonstrate whether data augmentation using GANs is a viable and robust solution for the training of steel defect segmentation models. First, we study how the segmentation performance varies along with changes in the synthetic image ratio in the dataset in different training processes. With this experiment, we want to find the optimal ratio and prove the robustness of the segmentation model. Secondly, we validate if the metrics obtained with the optimal image ratio using DeepLabV3+ are repeatable with another segmentation network, such as PAN [27]. Finally, with the same optimal image ratio, we replicate the training process of StyleGAN2+DiffAugment and PAN network with the NEU public database in order to validate if the benefit obtained in our dataset also correlates with another dataset.

3.1. DeepLabV3+ Trained with Real Images and Conventional Data Augmentation

In this experiment, we trained the DeepLabV3+ defect segmentation model using only original images augmented with conventional data augmentation methods as described in Section 2.3.1. The training process was carried out by using 80% of the augmented database and the remaining 20% is set aside for later evaluation. Training was carried out on two Tesla GPUs with 16 GB of VRAM each.

Figure 6 shows some visual detection results obtained with this model and dataset. In this case, we obtain a mean IoU of 68.3%. These results represent the baseline benchmark for the rest of the experiments.

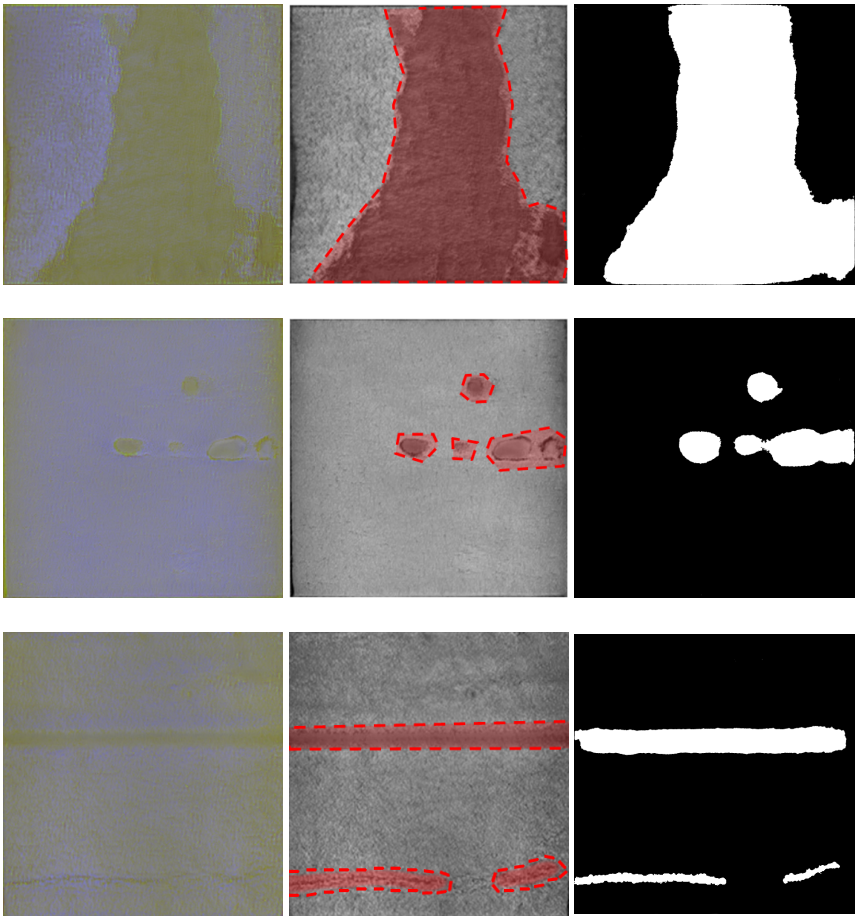


(**a**) (**b**) (**c**)

Figure 6. Some visual results of defect detection by the DeepLabv3+ trained on the conventional augmented dataset. (**a**) photometric-stereo texture channel images; (**b**) ground truth defect masks; and (**c**) predicted defect masks.

3.2. StyleGAN2+DiffAugment for Defect Images Generation

For this training stage, the network requires that the training set has at least 100 real images, as described in [25]. For the StyleGAN2+DiffAugment training process, the original 204 photometric-stereo RGB images were used without any transformation or processing. Using the trained StyleGAN2+DiffAugment, we generated 100 new synthetic images. Generated images capture the general appearance of the components as well as the particular characteristics of the defects of the galvanized steel material. However, generated data need to be annotated. Thus, the annotation of the new StyleGAN2+DiffAugment generated images was done by the DeepLabv3+ semantic segmentation network trained in the previous computational experiment. Subsequently, in order to avoid the holes inside the predicted defect masks, a correction consisting of mathematical morphology was applied. The result can be seen in the set of images shown in Figure 7, where the predicted mask fits the defects of the generated steel sheets.



(**a**) (**b**) (**c**)

Figure 7. (**a**) Synthetic RGB images; (**b**) texture channel of Synthetic images with the defects in red (**c**) ground truth masks.

3.3. Variations in the Ratio of GAN Generated Images in the Training Dataset

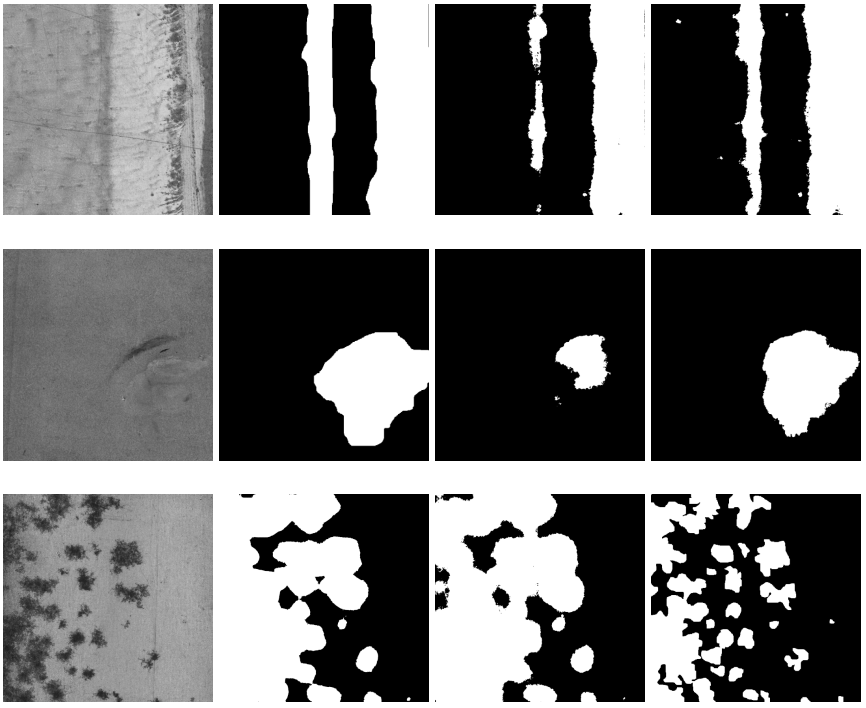
In this computational experiment, six instances of the DeepLabv3+ architecture were trained for semantic segmentation. The basic dataset for the experiment consists of the 204 original images, and a second set with 41 images for the validation process. In addition to the training set, 100 generated images by the GAN were added.

The strategy that was carried out during the process of augmenting the dataset with synthetic images was to modify the number of real images in the original database, and to keep the number of synthetic images constant. Therefore, the volume of the real images was varied three times, one for each training instance, as can be shown in the first column of Table 1. Consequently, the ratio of synthetic versus real image on the training set changes. Table 1 shows different volumes of learning images, along with the corresponding percentage of synthetic images ratio, and the average IoU metric provided by the segmentation model over the validation dataset.

These results show that the mean IoU decreases as the number of real images decreases and the number of synthetic images remains constant. In addition, the number of original images is not very high, so this trend may not be as significant in a more complete dataset.

After evaluating the result of the trained DeepLabv3+ networks in both scenarios, i.e., the scenario using only real images and the scenario combining real and GAN generated images, a significant improvement in the predictions of the second scenario can be perceived in Figure 8—specifically, when the database is composed of 38% of GAN generated images. It can be seen in Figure 8d that the predicted masks are closer to the perimeter of the defects. A comparison between the prediction masks obtained with a DeepLabv3+ network trained using only the original database and with the dataset including synthetic

images generated by the GAN is shown in Figure 8c,d, respectively. It can be assured that the performance of semantic segmentation network performance increases using GAN synthetic images on the basis of the improved visualized defect location and on the obtained metrics in the confusion matrices, shown in Figure 9. It can be observed that the detection of defects is more accurate when using data enhancement with synthetic images, as shown by the value of true positives. The obtained metrics in the confusion matrix for DeepLabV3+ trained using only real images and the training combining real and GAN generated images is shown in Figure 9a,b, respectively.



(**a**) (**b**) (**c**) (**d**)

Figure 8. Detection of defects using two instances of the DeepLabv3+ semantic segmentation neural network. (**a**) photometric-stereo texture channel images; (**b**) ground truth masks; (**c**) predictions with the original RGB photometric-stereo database; and (**d**) predictions adding GAN-generated RGB photometric-stereo images.

Table 1. Different variations of the GAN generated images ratio in the data augmentation of the original RGB training set and the respective mean IoU values (%) of the validation process for each instance.

Different Training Instances of the DeepLabv3+ Model					
Number of Training Set Real Images	**Number of Training Set GAN Generated Images**	**Total Number of Training Set Images**	**Conventional Fifteen-Fold Data Augmentation of the Training Set Images**	**GAN Generated Data Ratio (%)**	**Mean IoU (%) during Evaluation**
163	0	163	2445	0	68.3
163	25	188	2820	13	69.02
163	50	213	3195	23.5	69.16
163	75	238	3570	31.5	69.77
163	100	263	3945	38	70
130	100	230	3450	44	60.09
100	100	200	3000	50	59.93

Figure 11 shows that we get 70% of IoU thanks to the data augmentation done by GAN, compared with 68.3% IoU using only real images. Similarly, it is observed that the IoU curve shown in Figure 10 with the increase of data and the 38% ratio is placed in a higher position compared to the IoU curve with the original dataset. It should be noted that the training set is unbalanced in terms of the type of surface defect of the galvanized steel. It is observed that defects that alter the base color of the material (such as oxidation) are more abundant in the dataset than defects with dimensional affection (such as scratches or impact marks). These images with color based alteration are those included in the range between image 19 and image 35.

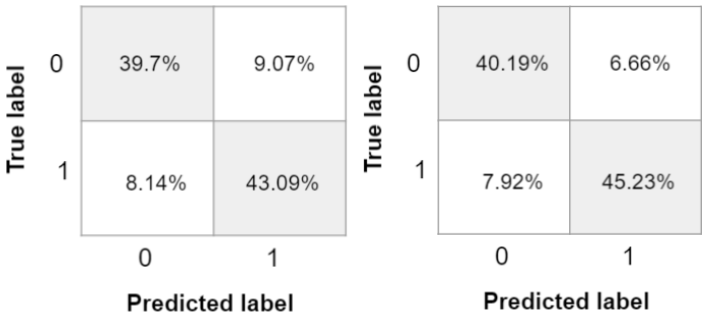


Figure 9. Confusion matrices of the test set, where label 0 represents background and label 1 means defect. (**a**) Confusion matrix of DeepLabV3+ trained only with real images and (**b**) confusion matrix of DeepLabV3+ trained with the original dataset augmented with synthetic images generated by the GAN with a 38% ratio.

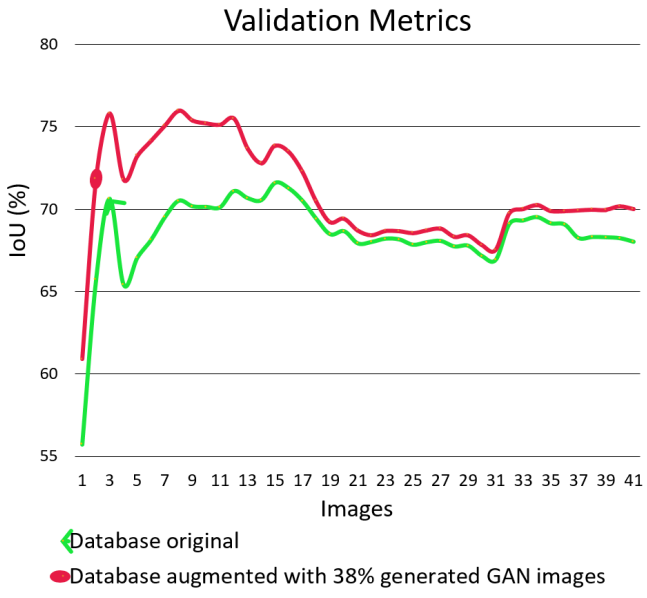


Figure 10. Mean of IoU (%)result with original dataset augmented with a conventional method and the original augmented with synthetic images generated by the GAN with a 38% ratio.

It can be seen in Figure 10 that there are segmentation results in some images, specifically between images 3 and 15 of the test set that obtains more than 6% improvement in the IoU when using a dataset augmented with GAN generated images. We realized that images exhibiting that improvement are those with more severe surface dimensional affection. On the other hand, images that obtain almost the same results are those related with changes only in the color of the material. We argue that segmenting images with surface dimensional affection requires segmentation models with higher information.

Therefore, models that were trained with an augmented dataset using GAN generated images obtain better results compared with those trained with original real images only, due to the addition of new images. Thanks to these new images, we are adding more information to the segmentation model for better distinguishing between affected and non-affected surface regions.

3.4. Using PAN Network for Defect Segmentation

In this experiment, the PAN network [27] was trained with the original real images and with the optimal synthetic augmentation dataset to validate the benefits in a complete different segmentation architecture compared with DeepLabV3+.

PAN is proposed to exploit the impact of global contextual information in semantic segmentation. This network combines attention mechanism and spatial pyramid to extract precise dense features for pixel labeling instead of dilated convolution and artificially designed decoder networks. PAN uses a Feature Pyramid Attention (FPA) module to perform spatial pyramid attention structure on high-level output and combining global pooling to learn a better feature representation. This network also has a Global Attention Upsample module on each decoder layer to provide global context as a guidance of low-level features to select category localization detail [27].

With this network and the original dataset, we obtain an IoU metric of 75.31%, while, in the case of the augmented dataset, we get 82.29%. This repetition in the IoU increment shows that the benefits of augmenting the dataset using synthetic images are also repeated using another segmentation network. The comparison of the obtained results with both segmentation networks with the previously-mentioned datasets is shown in Figure 11. Moreover, apart from the metric, the segmentation quality results are also better. The segmentation performed by the network trained with synthetic data are better adapted to the geometry of the defects. Due to the FPA module, this architecture uses the information at different scales, so the segmentation output is better in terms of pixel-level attention.

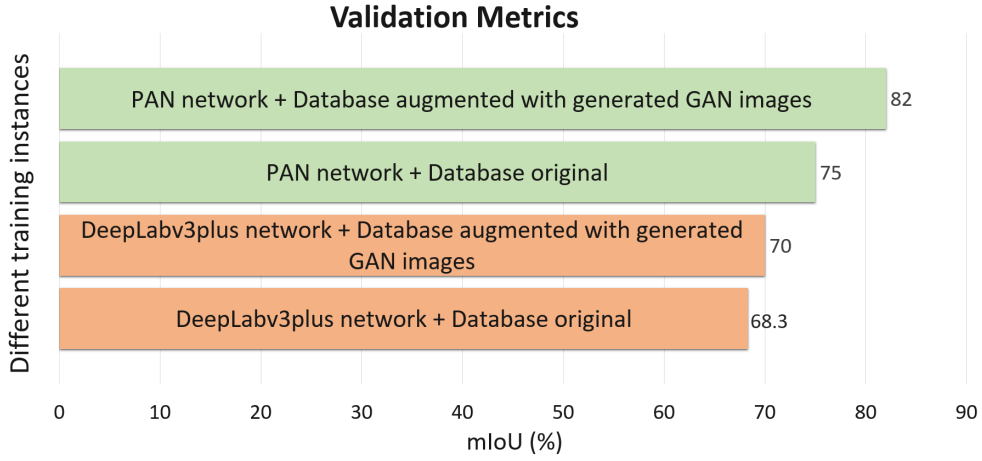


Figure 11. Mean of IoU (%) during evaluation with the original photometric-stereo RGB augmented database and the original one plus synthetic images generated by the GAN with a 38% ratio.

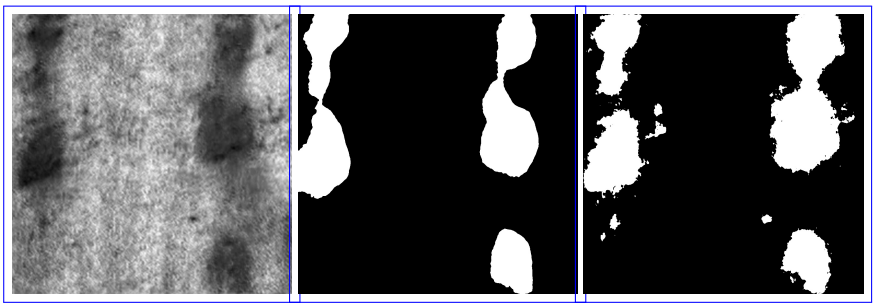
3.5. Validation of the Proposed Method on the NEU Dataset

In order to validate that the benefits obtained by introducing synthetically generated images with the GAN is repeated with other datasets, we replicated the process but using an NEU publicly available dataset [28]. The NEU dataset is a defect database composed of six types of typical defects in hot rolled steel surfaces. These defects are: rolled-in scale (RS), patches (Pa), crazing (Cr), pitted surface (PS), inclusion (In), and scratches (Sc). This database has a total of 1800 greyscale images, where each class has 300 different samples.

As described in Section 3.2, we also trained a StyleGAN2 network to generate 100 new synthetic images but from the NEU dataset. As for the segmentation process, PAN

was also trained with both an original NEU dataset and with an augmented dataset with StyleGAN2 generated images.

In the training with the original NEU dataset augmented with conventional methods, a total of 2445 images were used. In the training with the original images plus the StyleGAN2 synthesized images, as stated in the experiment of image ratio variation, the number of the training set amounts to 3930. The results on the validation set, which has 42 un-augmented images, was 69.11% in IoU with the original dataset and 75.32% in IoU with the synthetic images. A sample image, its corresponding ground truth, and the network segmentation output are shown in Figure 12.



(**a**) (**b**) (**c**)

Figure 12. Results of defect segmentation by the PAN trained on the synthetically augmented NEU dataset. (**a**) StyleGAN2 generated synthetic image; (**b**) ground truth defect mask; and (**c**) predicted defect mask.

The results obtained show that data augmentation by generating synthetic images gives good results in terms of segmentation quality and in the IoU metric. This shows that this data augmentation technique is an additional benefit in network training. It is true that it does not match the results obtained in training using more real images, but it is a good tool to apply variability in training, which traditional data augmentation methods cannot provide.

4. Conclusions

Our main conclusion is that using GAN generated images significantly improves the performance of the semantic network trained for defect segmentation. These types of generative networks allow for obtaining realistic images that are useful to solve the lack of samples in the training dataset, thus improving the predictive power of the semantic segmentation network.

The results show that this approach may be a good solution for the dataset augmentation in an industrial environment, where the difficulties to obtain defective parts is very common. Even for unbalanced databases in terms of defect typology, the generated GAN images can complement the information needed for segmenting a particular type of defect, especially those related with dimensional surface affection. Moreover, the use of GAN images in the model training can increase its accuracy, helping to generalize the surface defect detection process.

Our experiments show that 38% of the GAN generated images ratio improves the segmentation results compared with the original dataset using only conventional data augmentation. If there is any restriction in the real data acquisition process, the use of GANs could be a very powerful and effective tool as a data augmentation method for the optimization of the neural network learning process. It has even been observed that the addition of GAN generated images into the database improves segmentation results of both the DeepLabv3+ and PAN neural models. Therefore, it is argued that such images created by the GANs are fully compatible to be used in machine vision systems based on Deep Learning, and more precisely in semantic segmentation tasks.

Regarding the results obtained in IoU with the two networks in our dataset, the one with the best results was obtained by a PAN approach. We argue that PAN architecture

is able to fuse context information from different scales and produce better pixel level attention, obtaining better results in the segmentation of small defects.

The improvement obtained by augmenting the data with synthetic images was also demonstrated on the public NEU dataset. The same steps as in our dataset were replicated on this dataset, and, by augmenting the training set with synthetic images, we have managed to increase the IoU from 69.11% to 75.32% using the PAN network.

A line of future research is to modify the loss function of the segmentation network into one that allows the user of the system to tune up the output of the networks in terms of false positive and true positive rates, i.e., in terms of the sensitivity and specificity of the network.

Author Contributions: Conceptualization, F.A.S., G.A, I.B. and M.G.; methodology, F.A.S. and G.A.; software, F.A.S. and G.A.; validation, F.A.S. and G.A.; investigation, F.A.S. and G.A.; resources, I.B.; data curation, G.A.; writing—original draft preparation, F.A.S., G.A, I.B. and M.G.; writing—review and editing, F.A.S., G.A, I.B. and M.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Posada, J.; Toro, C.; Barandiaran, I.; Oyarzun, D.; Stricker, D.; de Amicis, R.; Pinto, E.B.; Eisert, P.; Döllner, J.; Vallarino, I. Visual Computing as a Key Enabling Technology for Industrie 4.0 and Industrial Internet. *IEEE Comput. Graph. Appl.* **2015**, *35*, 26–40. [CrossRef]
2. Mallick, P. Advanced materials for automotive applications: An overview. In *Advanced Materials in Automotive Engineering*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 5–27
3. Qi, S.; Yang, J.; Zhong, Z. A Review on Industrial Surface Defect Detection Based on Deep Learning Technology. In Proceedings of the 2020 The 3rd International Conference on Machine Learning and Machine Intelligence, Hangzhou, China, 13–20 September 2020; pp. 24–30
4. Zhang, A.; Li, S.; Cui, Y.; Yang, W.; Dong, R.; Hu, J. Limited data rolling bearing fault diagnosis with few-shot learning. *IEEE Access* **2019**, *7*, 110895–110904. [CrossRef]
5. Lin, D.; Cao, Y.; Zhu, W.; Li, Y. Few-Shot Defect Segmentation Leveraging Abundant Normal Training Samples Through Normal Background Regularization and Crop-and-Paste Operation. *arXiv* **2020**, arXiv:2007.09438.
6. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujscie, Poland, 9–12 May 2018; pp. 117–122. doi:10.1109/IIPHDW.2018.8388338 [CrossRef]
7. Fu, G.; Sun, P.; Zhu, W.; Yang, J.; Cao, Y.; Yang, M.Y.; Cao, Y. A deep-learning-based approach for fast and robust steel surface defects classification. *Opt. Lasers Eng.* **2019**, *121*, 397–405. doi:10.1016/j.optlaseng.2019.05.005. [CrossRef]
8. Luo, Q.; Fang, X.; Liu, L.; Yang, C.; Sun, Y. Automated Visual Defect Detection for Flat Steel Surface: A Survey. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 626–644. [CrossRef]
9. Li, J.; Su, Z.; Geng, J.; Yin, Y. Real-time detection of steel strip surface defects based on improved yolo detection network. *IFAC-PapersOnLine* **2018**, *51*, 76–81. [CrossRef]
10. Zhou, J.; Zhao, W.; Guo, L.; Xu, X.; Xie, G. Real Time Detection of Surface Defects with Inception-Based MobileNet-SSD Detection Network. In Proceedings of the International Conference on Brain Inspired Cognitive Systems, Guangzhou, China, 13–14 July 2019; Springer: Amsterdam, The Netherlands, 2019; pp. 510–519
11. DAGM 2007. Available online: https://hci.iwr.uni-heidelberg.de/content/weakly-supervised-learning-industrial-optical-inspection (accessed on 7 July 2021)
12. Enshaei, N.; Ahmad, S.; Naderkhani, F. Automated detection of textured-surface defects using UNet-based semantic segmentation network. In Proceedings of the 2020 IEEE International Conference on Prognostics and Health Management (ICPHM), Detroit, MI, USA, 8–10 June 2020; IEEE: New York, NY, USA 2020; pp. 1–5
13. Nie, Z.; Xu, J.; Zhang, S. Analysis on DeepLabV3+ Performance for Automatic Steel Defects Detection. *arXiv* **2020**, arXiv:2004.04822.
14. Jun, Z.; Shiqiao, C.; Zhenhua, D.; yuan, Z.; Chabei, L. Automatic Detection for Dam Restored Concrete Based on DeepLabv3. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *571*, 012108. [CrossRef]

15. Xie, X.; Wan, T.; Wang, B.; Xu, L.; Li, X. A Mechanical Parts Image Segmentation Method Against Illumination for Industry. 2020. Available online: https://assets.researchsquare.com/files/rs-119471/v1_stamped.pdf?c=1607614313 (accessed on 7 July 2021)
16. Chen, Y.; Song, B.; Du, X.; Guizani, N. The enhancement of catenary image with low visibility based on multi-feature fusion network in railway industry. *Comput. Commun.* **2020**, *152*, 200–205. [CrossRef]
17. Bagheri, N.M.; van de Venn, H.W.; Mosaddegh, P. Development of Machine Vision System for Pen Parts Identification under Various Illumination Conditions in an Industry 4.0 Environment. *Preprints* **2020**. [CrossRef]
18. Andulkar, M.; Hodapp, J.; Reichling, T.; Reichenbach, M.; Berger, U. Training CNNs from Synthetic Data for Part Handling in Industrial Environments. In Proceedings of the 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE), Munich, Germany, 20–24 August 2018; IEEE: New York, NY, USA 2018; pp. 624–629
19. Hinterstoisser, S.; Lepetit, V.; Wohlhart, P.; Konolige, K. On pre-trained image features and synthetic images for deep learning. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2018.
20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*; Association for Computing Machinery: New York, NY, USA 2014; pp. 2672–2680
21. Goyal, M.; Rajpura, P.; Bojinov, H.; Hegde, R. Dataset augmentation with synthetic images improves semantic segmentation. In Proceedings of the National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics, Mandi, India, 16–19 December 2017; Springer: Amsterdam, The Netherlands, 2017; pp. 348–359
22. Liu, L.; Cao, D.; Wu, Y.; Wei, T. Defective samples simulation through adversarial training for automatic surface inspection. *Neurocomputing* **2019**, *360*, 230–245. doi:10.1016/j.neucom.2019.05.080. [CrossRef]
23. Tang, R.; Mao, K. An improved GANS model for steel plate defect detection. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *790*. [CrossRef]
24. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
25. Zhao, S.; Liu, Z.; Lin, J.; Zhu, J.Y.; Han, S. Differentiable augmentation for data-efficient gan training. *arXiv* **2020**, arXiv:2006.10738.
26. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* **2018**, arXiv:1802.02611.
27. Li, H.; Xiong, P.; An, J.; Wang, L. Pyramid Attention Network for Semantic Segmentation. *arXiv* **2018**, arXiv:1805.10180.
28. Song, K.; Yan, Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl. Surf. Sci.* **2013**, *285*, 858–864. doi:10.1016/j.apsusc.2013.09.002. [CrossRef]
29. Cachero, R.; Abello, C. Técnicas estereo-fotométricas para la digitalización a escala micrométrica. *Virtual Archaeol. Rev.* **2015**, *6*, 72–76. [CrossRef]
30. Salvador Balaguer, E. *Tecnologías Emergentes para la Captura y Visualización de imagen 3D*; Universitat Jaume I: Castellón de la Plana, Spain, 2017.
31. Lateef, F.; Ruichek, Y. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* **2019**, *338*, 321–348. doi:10.1016/j.neucom.2019.02.003. [CrossRef]
32. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
33. He, H.; Yang, D.; Wang, S.; Wang, S.; Li, Y. Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss. *Remote Sens.* **2019**, *11*, 1015. [CrossRef]
34. Lian, X.; Pang, Y.; Han, J.; Pan, J. Cascaded Hierarchical Atrous Spatial Pyramid Pooling Module for Semantic Segmentation. *Pattern Recognit.* **2020**, p. 107622.
35. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
36. Ma, J.; Dai, Y.; Tan, Y.P. Atrous convolutions spatial pyramid network for crowd counting and density estimation. *Neurocomputing* **2019**, *350*, 91–101. [CrossRef]
37. Chen, J.; Wang, C.; Tong, Y. AtICNet: semantic segmentation with atrous spatial pyramid pooling in image cascade network. *EURASIP J. Wirel. Commun. Netw.* **2019**, *2019*, 146. [CrossRef]
38. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding convolution for semantic segmentation. In Proceedings of the 2018 IEEE winter conference on applications of computer vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; IEEE: New York, NY, USA, 2018, pp. 1451–1460
39. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
40. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and Improving the Image Quality of StyleGAN. *arXiv* **2020**, arXiv:1912.04958

Regular Issue

COVID-19 Detection in Chest X-ray Images using a Deep Learning Approach

Fátima A. Saiz*, Iñigo Barandiaran

Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Donostia – San Sebastián (Spain)

Received 10 April 2020 | Accepted 29 April 2020 | Published 30 April 2020

Abstract

The Corona Virus Disease (COVID-19) is an infectious disease caused by a new virus that has not been detected in humans before. The virus causes a respiratory illness like the flu with various symptoms such as cough or fever that, in severe cases, may cause pneumonia. The COVID-19 spreads so quickly between people, affecting to 1,200,000 people worldwide at the time of writing this paper (April 2020). Due to the number of contagious and deaths are continually growing day by day, the aim of this study is to develop a quick method to detect COVID-19 in chest X-ray images using deep learning techniques. For this purpose, an object detection architecture is proposed, trained and tested with a public available dataset composed with 1500 images of non-infected patients and infected with COVID-19 and pneumonia. The main goal of our method is to classify the patient status either negative or positive COVID-19 case. In our experiments using SDD300 model we achieve a 94.92% of sensibility and 92.00% of specificity in COVID-19 detection, demonstrating the usefulness application of deep learning models to classify COVID-19 in X-ray images.

Keywords

COVID-19, Deep Learning, Object Detection, X-ray.

DOI: 10.9781/ijimai.2020.04.003

I. Introduction

THE new SARS-CoV-2 coronavirus, which produces the disease known as COVID-19, kept the whole world on edge during the first months of 2020. It provoked the borders close of many countries and the confinement of millions of citizens to their homes due to infected people, which amounts to 868,000 confirmed cases worldwide at this moment (April 2020). This virus was originated in China in December 2019. From March 2020, Europe was the main focus of the virus sprout, achieving more than 445,000 infected people.

China, with a total of 3,312 deaths and more than 81,000 infected people, has managed to contain the virus almost three months after the start of the crisis in December 2019. Italy, which surpassed the Asian country in death toll on March 2020, became the most affected country, in number of deceased is followed by Spain, with more than 10,000 dead based on a report made on April 2020. This number was constantly growing. There were different studies that predicted the growth of the curves of infections, based on different parameters such as exposed, infected or recovered human's number. These studies allowed to get an idea of the transmission dynamics that could occur in each country [1] [2].

The origin of the outbreak is unknown. The first cases were detected in December 2019. The clinical characteristics of COVID-19 include respiratory symptoms, fever, cough, dyspnea, and viral pneumonia [3] [4]. The main problem of these symptoms is that there are virus-infected asymptomatic patients.

* Corresponding author.
E-mail address: fsaiz@vicomtech.org

The test to detect the COVID-19 is based on taking samples from the respiratory tract. It is carried out by a health care professional at home, generally when the case study is asymptomatic or symptoms are mild, or in a health center or hospital, if the patient is admitted for a serious condition. Carrying out as many tests as possible has shown to be the key tool to stop the virus in countries like Germany or South Korea. Spain was not able to carry out so many tests, therefore it is important to research and develop alternative methods to perform these tests in a quick and effective way.

AI and radiomics applied to X-Ray and Computed Tomography (CT) are useful tools in the detection and follow-up of the disease [5] [6]. As stated in [7], conspicuous ground grass opacity lesions in the peripheral and posterior lungs on CT images are indicative of COVID-19 pneumonia. Therefore, CT can play an important role in the diagnosis of COVID-19 as an advanced imaging evidence once findings in chest radiographs are indicative of coronavirus. AI algorithms and radiomics features derived from Chest X-rays would be of huge help to undertake massive screening programs that could take place in any country with access to X-ray equipment and aid in the diagnosis of COVID-19 [8] [9].

The current situation evokes the necessity to implement an automatic detection system as an alternative diagnosis option to prevent COVID-19 spreading among people. There are different studies that apply machine learning for this task, such as Size Aware Random Forest method (iSARF) that was proposed by [10], in which subjects were categorized into groups with different ranges of infected lesion sizes. Then a random forest-based classifier was trained with each group. Experimental results show that their proposed method yielded an accuracy of 0.879 under five-fold cross-validation, a sensitivity of 0.907 and a specificity of 0.833.

Deep learning techniques are also used in order to achieve better results than using more traditional machine learning approaches. One of the most used approach in image classification is the use of convolutional neural networks (CNNs). This type of models are used in different studies for COVID-19 detection in medical images like in [11], in their study the authors propose a CNN model trained with a randomly selection of image regions of interest (ROIs), achieving a 85.2% of accuracy, 0.83 of specificity and 0.67 of sensitivity. Other example of the results that can be obtained using CNNs is presented by [12]. They propose the COVID-Net CNN network obtaining 92.4% of accuracy, 80% of sensibility and 88.9% of specificity.

A method called COVIDX-Net is presented by [13], COVIDX-Net includes seven different architectures of deep convolutional neural network models, such as a modified version of Visual Geometry Group Network (VGG19) and the second version of Google MobileNet. Each deep neural network model is able to analyze the normalized intensities of the X-ray image to classify the patient status either negative or positive COVID-19 case. Their experiments evaluation achieves f1-scores of 0.89 and 0.91 for healthy and COVID-19 detection respectively.

The results shown in mentioned works demonstrate that deep learning techniques are useful for the virus detection, and that improve the obtained metrics using more traditional machine learning approaches [11] [12] [13].

The main contribution of our paper is focused on the improvement of the detection accuracy of COVID-19, by proposing a new dataset that combines COVID-19 and pneumonia images to make more stable predictions and by applying image processing that allows image-standardization and also improves model learning.

II. Method

We propose to use a deep convolutional neural network specialized for object detection along with a new dataset composed of COVID-19 and pneumonia images. Both are publicly available on GitHub [14] and Kaggle [15] respectively. The chest X-ray or CT images that are available in GitHub belong to COVID-19 cases. It was created by assembling medical images from public available websites and publications. This dataset contains 204 COVID-19 X-ray images. On the other hand, the Kaggle dataset was created for a pneumonia detection challenge. The images have bounding boxes around diseased areas of the lung. Samples without bounding boxes are negative and contain no definitive evidence of pneumonia. Samples with bounding boxes indicate evidence of pneumonia.

We propose a new dataset by merging COVID-19 and pneumonia images to obtain a wider and diverse one. The fact of having pneumonia images in the training dataset supposes an extra advantage, due to normal pneumonia and COVID-19 have similar appearance in chest X-ray images. This dataset merge allows to get a robuster model that is able to better distinguish between those diseases. Another advantage of this merge is the fact of enlarging the train dataset, because the COVID-19 images are not abundant at the time of writing this paper. This merge does not enlarges COVID-19 image set but improves detection quality because of the similarity between pneumonia and COVID-19. Train with pneumonia images gives an extra knowledge to the model in order to not confuse COVID-19 with pneumonia, being more effective and stable in disease detection.

We split the images in train and test sets, dividing all the data in a balanced way, meaning that all samples of each class in the training sets are well-balanced, in order to avoid biased results. For this purpose, even though we have a large number of pneumonia and normal images, we compose a dataset of 1500 images.

We compose the dataset as follows: we select 104 COVID-19 images, 205 health lung images and 204 pneumonia images for training, and 100 COVID-19 images, 444 health lung images and 443 pneumonia images for testing. In the Kaggle dataset there are more samples of pneumonia and normal images, but we select only 205 for training with the purpose of having a balanced dataset. In the test stage, we add more pneumonia and normal images to demonstrate the robustness of the model giving no false positives in the detection of COVID-19. Summing up, we use a train set composed of 513 images and a test set of 887 images in total.

In the next step, training images are labeled using a XML annotation file based on Pascal VOC format [16]. Each sample in the dataset is an image with ground truth bounding boxes for each object in the images as shown in the Fig. 1.

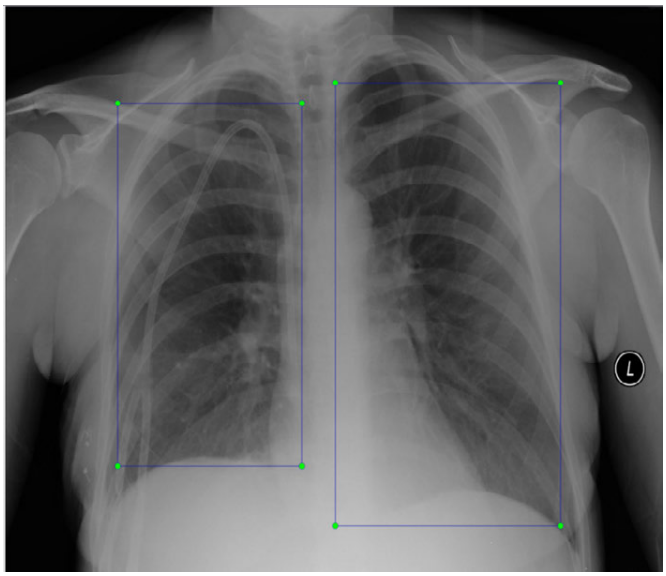


Fig. 1. Created bounding boxes of normal regions in a chest X-Ray image.

III. Model Architecture

The selection of the used architecture is based on the good results obtained with CNNs in the state-of-the-art works for COVID-19 image classification, and the good results obtained in other similar tasks with this kind of architecture [11] [12] [13]. We used the same network architecture as proposed in [17], based on Single Shot Multibox Detector (SSD). This architecture is optimized for detecting objects in images using a single deep neural network. This approach discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.

Experimental results on different remarkable datasets confirm that SSD has comparable accuracy to methods that utilize more than one architecture for detecting objects being much faster, while providing a unified framework for both training and inference. Compared to other single stage methods, SSD has much better accuracy, even with a smaller input image size [17].

We use VGG-16 [18] as the base network for performing feature extraction in this architecture. This model is also based on Fast R-CNN. During training, we have multiple boxes with different sizes

and different aspect ratios across the whole image. SSD finds the box that has more Intersection-Over-Union (IoU) compared with the ground truth. A detailed description of the layers architecture of SDD network is shown in [17].

Our main goal is to obtain a more robust model to various input object sizes and shapes. Therefore, during the SSD training a data augmentation step is performed. This process is composed by the following operations applied to every image in the dataset:

- Use the entire original input image.
- Sample a patch so that the minimum overlap with the objects is 0.1, 0.3, 0.5, 0.7, or 0.9. The size of each sampled patch is a percentage between [0.1, 1] of the original image size.
- Randomly sample a patch.

During the inference, SDD uses 8732 boxes for a better coverage of object location. After the inference step, a set of boxes representing the detected objects are given along with the respective label and score, as shown in Fig. 2.

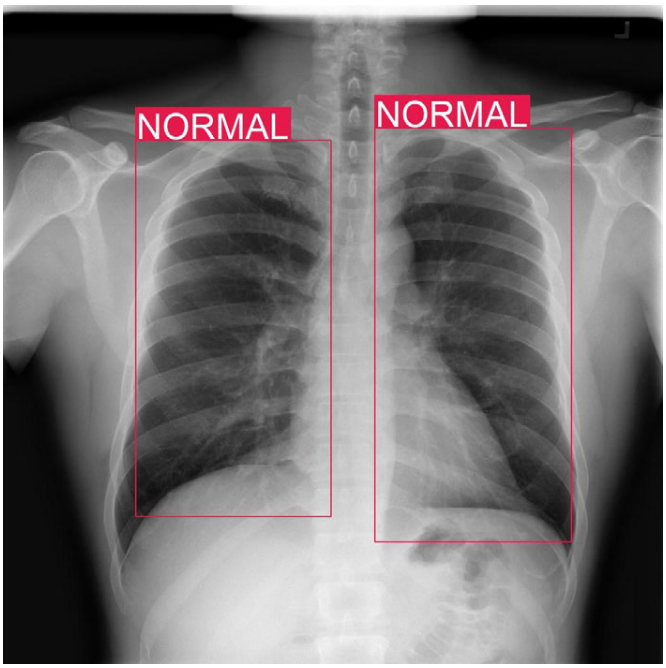


Fig. 2. Detection boxes of SSD model inference output detecting normal lungs.

One of the advantages of this model is the possibility of precise object localization, which is not the case in previous works [11] [12] [13].

IV. Experiments and Results

The first experiment made in this work, is the contrast adjustment of each image in the dataset. This adjustment is necessary because the exposure time in X-Ray images can be different between acquisitions. All the images of the dataset are from different hospitals around the world, so the image acquisition settings and conditions are different in each place. In X-Ray images, an adjustment in the voltage spike results in a change in the contrast of the radiography. Exposure time, which refers to the time interval during which x-rays are produced, is also a factor that affects the contrast of the obtained image [19].

In order to get image similarity between the dataset, Contrast Limited Adaptive Histogram Equalization (CLAHE) [20] is applied. This is a transformation that aims to obtain a histogram with an even distribution for an image. That is, there is the same number of pixels for each level of gray in the histogram of a monochrome image. As cited in [20], in X-ray imaging, when continuous exposure is used to obtain an image sequence or video, usually low-level exposure is administered until the region of interest is identified, so reducing the radiation applied to the patients. As a drawback, images with low signal-to-noise ratio are obtained. In this case, and many other similar situations, it is desirable to improve image quality by using some type of image enhancement such as histogram equalization algorithms. An example of the application of this image operation is shown in Fig. 3.

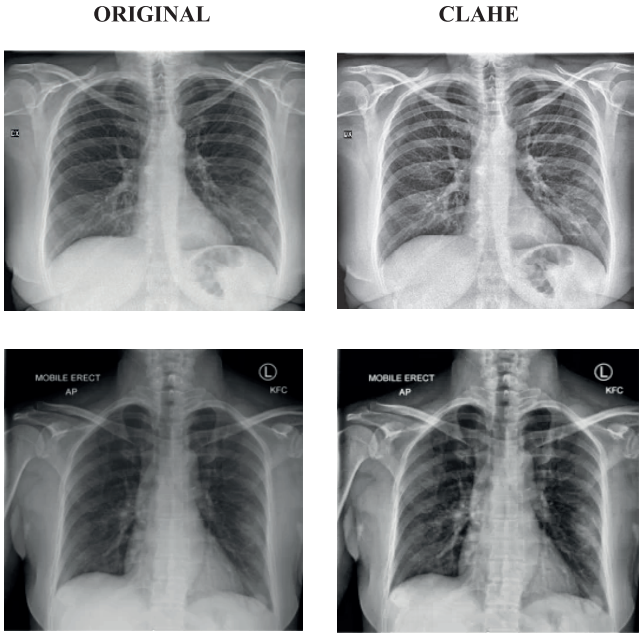


Fig. 3. Comparison between original image and CLAHE applied images.

The differences obtained in the detection applying or not CLAHE in the train and test datasets are shown in TABLE I. As the table shows, the fact of applying this pre-processing increases notably the detection accuracy in health and infected lungs.

We load VGG-16 weights trained on ImageNet. There are many works that evaluate the accuracy improvement using transfer learning, specially in small datasets [21] [22]. We apply these weights because though lower layers learn features that are not necessarily specific to this dataset, this action improves the detection accuracy and the sensibility and specificity metrics. The key idea is to take advantage of a trained model for similar images and adapt a base-learner to a new task for which only a few labeled samples are available.

The last experiment evaluates the detection accuracy obtained in the detection of COVID-19. For this purpose, we set apart two sets of images, one with non-COVID-19 and the other one with COVID-19 positives. We obtain different metrics that can be seen in TABLE II. Another metric that we take in care is the inference time that we obtain running the model on a GPU, achieving 0,13s per image.

TABLE I. Obtained Results in Image Classification Applying or Not CLAHE in the Dataset.

Image class	CLAHE	Total images	True detection	Accuracy
Normal	No	887	827	93.24%
Normal	Yes	887	842	**94.92%**
COVID-19	No	100	83	83.00%
COVID-19	Yes	100	92	**92.00%**

TABLE II. Obtained Metrics Values

Metric	Operation	Value
Sensibility	842 / (842+45)	0.9492
Specificity	92 / (92+8)	0.9200

V. Conclusion

This study demonstrates the useful application to detect COVID-19 in chest X-ray images based on image pre-processing and the proposed object detection model.

The proposed merged dataset using pneumonia images allows getting a more robust model that is able to distinguish between COVID-19 and pneumonia diseases. With the histogram equalization operation, we can get a normalized dataset that helps to model training step. It also improves the normal image detection and minimizes the false positives rate.

With our proposed method, we achieve a 94.92% of sensibility and 92.00% of specificity in COVID-19 detection. The detection accuracy obtained using this architecture and the proposed dataset improves the results described in [11] [12] [13]. These results demonstrate that object detection models trained with more images of similar diseases and applying transfer learning, combined with CLAHE algorithm for image normalization, could be successful in medical decision-making processes related with COVID-19 virus diagnosis.

References

[1] M. Dur-e-Ahmad and M. Imran, "Transmission Dynamics Model of Coronavirus COVID-19 for the Outbreak in Most Affected Countries of the World," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. In Press, no. In Press, pp. 1-4, 2020.

[2] S. J. Fong, N. D. G. Li, R. Gonzalez-Crespo and E. Herrera-Viedma, "Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 6, no. 1, pp. 132-140, 2020.

[3] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong and others, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *New England Journal of Medicine,* 2020.

[4] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang and Z. Peng, "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China," *JAMA,* vol. 323, pp. 1061-1069, 3 2020.

[5] S. Chauvie, A. De Maggi, I. Baralis, F. Dalmasso, P. Berchialla, R. Priotto, P. Violino, F. Mazza, G. Melloni and M. Grosso, "Artificial intelligence and radiomics enhance the positive predictive value of digital chest tomosynthesis for lung cancer detection within SOS clinical trial," *European Radiology,* p. 1–7, 2020.

[6] G. Chassagnon, M. Vakalopoulou, N. Paragios and M.-P. Revel, "Artificial intelligence applications for thoracic imaging," *European Journal of Radiology,* vol. 123, p. 108774, 2020.

[7] F. Song, N. Shi, F. Shan, Z. Zhang, J. Shen, H. Lu, Y. Ling, Y. Jiang and Y. Shi, "Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia," *Radiology,* vol. 295, pp. 210-217, 2020.

[8] J. C. L. Rodrigues, S. S. Hare, A. Edey, A. Devaraj, J. Jacob, A. Johnstone, R. McStay, A. Nair and G. Robinson, "An update on COVID-19 for the radiologist-A British society of Thoracic Imaging statement," *Clinical Radiology,* 2020.

[9] J. Wu, J. Liu, X. Zhao, C. Liu, W. Wang, D. Wang, W. Xu, C. Zhang, J. Yu, B. Jiang and others, "Clinical characteristics of imported cases of COVID-19 in Jiangsu province a multicenter descriptive study," *Clinical Infectious Diseases,* 2020.

[10] F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui and D. Shen, *Large-Scale Screening of COVID-19 from Community Acquired Pneumonia using Infection Size-Aware Classification, arXiv preprint arXiv:2003.09860,* 2020.

[11] S. Wang, J. M. Bo Kang, X. Zeng and M. Xiao, "A deep learning algorithm using CT images to screen for Corona Virus Disease COVID-19)," *medRxiv,* 2020.

[12] L. Wang and A. Wong *COVID-Net A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images, arXiv preprint arXiv:2003.09871* ,2020.

[13] E. E.-D. Hemdan, M. A. Shouman and M. E. Karar, *COVIDX-Net A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images, arXiv preprint arXiv: arXiv:2003.11055,* 2020.

[14] J. P. Cohen, P. Morrison and L. Dao, *COVID-19 Image Data Collection, arXiv preprint arXiv: arXiv, arXiv:2003.11597,* 2020.

[15] *RSNA Pneumonia Detection Challenge. Kaggle. [online]* Available at: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge, Accessed 29 April 2020.

[16] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision,* vol. 88, p. 303–338, 2010.

[17] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg, "SSD Single Shot MultiBox Detector," *Lecture Notes in Computer Science,* p. 21–37, 2016.

[18] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR,* vol. abs1409.1556, 2014.

[19] H. Jansen, Radiología dental. Principios y técnicas., Mc Graw Hill, 2002.

[20] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology,* vol. 38, p. 35–44, 2004.

[21] H.-W. Ng, V. D. Nguyen, V. Vonikakis and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015.

[22] Q. Sun, Y. Liu, T.-S. Chua and B. Schiele, "Meta-transfer learning for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

Fátima A. Saiz

Fátima Aurora Saiz Álvaro studied Engineering in Geomatics and Topography the University of the Basque Country during 2012-2017, being the first of its promotion. She obtained the title with her final degree project "Applications of programming languages to artificial intelligence, complex geophysical calculations and augmented reality" valued with honors, joining in the Talentia program of the Bizkaia Provincial Council. She completed the Master in Visual Analytics and Big Data of the Faculty of Engineering of the International University of La Rioja during 2017-2018. She obtained the title with her work "Study of architectures for the extraction and exploitation of superficial defects data using Deep Learning techniques". Since October 2017 she has been part of the Vicomtech staff as a researcher in the Department of Industry and Advanced Manufacturing, working in the field of cognitive vision and Deep Learning.

Íñigo Barandiaran

Iñigo Barandiaran received his PhD in Computer Science from the University of the Basque Country in the fields of computer vision and pattern recognition. He worked as a researcher in the group of "Artificial intelligence and computer science" at the University of Basque Country. Since 2003 he is a researcher at Vicomtech, participating and leading various R&D national and international projects, in various sectors such as biomedicine and industry. He has several scientific publications in conferences and journals in areas such as image analysis, or pattern recognition. He is currently the director of the Industry and Advanced Manufacturing department at Vicomtech.

sensors

MDPI

Article

Photometric Stereo-Based Defect Detection System for Steel Components Manufacturing Using a Deep Segmentation Network

Fátima A. Saiz [1,2,*], Iñigo Barandiaran [1], Ander Arbelaiz [1] and Manuel Graña [2]

[1] Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Mikeletegi 57, 20009 Donostia-San Sebastián, Spain; ibarandiaran@vicomtech.org (I.B.); aarbelaiz@vicomtech.org (A.A.)
[2] Computational Intelligence Group, Computer Science Faculty, University of the Basque Country, UPV/EHU, 20018 Donostia-San Sebastián, Spain; manuel.grana@ehu.es
* Correspondence: fsaiz@vicomtech.org

Abstract: This paper presents an automatic system for the quality control of metallic components using a photometric stereo-based sensor and a customized semantic segmentation network. This system is designed based on interoperable modules, and allows capturing the knowledge of the operators to apply it later in automatic defect detection. A salient contribution is the compact representation of the surface information achieved by combining photometric stereo images into a RGB image that is fed to a convolutional segmentation network trained for surface defect detection. We demonstrate the advantage of this compact surface imaging representation over the use of each photometric imaging source of information in isolation. An empirical analysis of the performance of the segmentation network on imaging samples of materials with diverse surface reflectance properties is carried out, achieving Dice performance index values above 0.83 in all cases. The results support the potential of photometric stereo in conjunction with our semantic segmentation network.

Keywords: photometric stereo; quality control; deep learning; image processing; semantic segmentation

Citation: Saiz, F.A.; Barandiaran, I.; Arbelaiz, A.; Graña, M. Photometric Stereo-Based Defect Detection System for Steel Components Manufacturing Using a Deep Segmentation Network. *Sensors* **2022**, *22*, 882. https://doi.org/10.3390/s22030882

Academic Editors: Hugo Landaluce and Alberto Tellaeche

Received: 30 November 2021
Accepted: 21 January 2022
Published: 24 January 2022

1. Introduction

Quality control is a centerpiece of any manufacturing industry, regardless of the industrial sector. This complex and demanding process must be carried out with a high degree of precision and rigour. Moreover, the quality of manufactured components directly impacts the positioning and profit of companies in their industrial sector. Component quality inspection requires checking many aspects of the products such as its dimensions, colour or surface characteristics. Many of these inspections are usually carried out by qualified operators, especially those aspects related to surface or cosmetic defects. The main problems with this manual inspection methodology are subjectivity, monotony and being prone to human error [1]. These problems, and recent advances in computer vision and machine learning, are encouraging the trend towards the integration of automated inspection, which eliminates subjectivity and analyzes all components quickly and effectively.

Surface inspection systems are especially difficult to automatize when surfaces are highly reflective or specular because of strong image variations due to reflection. These variations tends to generate very bright image regions and very deep shadows, impeding the detection of small surface defects. This paper describes our design for a photometric stereo image acquisition technique along with a convolutional segmentation network that achieves accurate detect surface detection in highly reflective surfaces.

1.1. Related Works

The first neural networks appeared thanks to the ideas about unsupervised learning published by [2,3]. Subsequently, the supervised and unsupervised learning concepts were

published [4] and the first neural networks were born, which were no more than variants of linear regressors. Deep neural networks were introduced by [5], who published the first learning algorithm for supervised deep feedforward multilayer perceptrons. Since that time, the development of deep learning has continued advancing, accompanied by technological progresses. These advances and the explosion of data have allowed the development of multiple applications in many different sectors, such us medicine [6], security systems [7] or robotics [8].

In the last decade there have been multiple advances in the field of machine learning. Especially, these advances come from the development of deep learning-related techniques. Deep learning architectures have become popular thanks to the increase of computing power and the availability of huge amounts of data [9,10].

Deep Learning algorithms are also increasingly applied in machine vision systems for industrial quality control. Due to the wide variety of surfaces that can be found in the manufacturing environment the way to acquire component images in each application must be customized to obtain the best image quality. Material surfaces may be specular, so choosing the right lighting schema for the acquisition of optimal images can be a difficult task. There are many works that try to eliminate the brightness produced by incident light over shiny surfaces that usually saturates image sensor regions, thus limiting any image processing in those areas [11]. For example, [12] proposes a method for specular reflection removal in a single image at the level of the individual pixel. The chromaticity of diffuse reflection is approximately estimated by employing the concept of modified specular-free image and the specular component is adjusted according to the criterion of smooth color transition along the boundary of diffuse and specular regions. Experimental results indicate that the proposed method is promising when compared with other state-of-the-art techniques, in both separation accuracy and running speed.

However, the best approach for obtaining useful images of specular surfaces is to have a reliable acquisition system that avoids surface reflections. Photometric stereo systems have been very suitable for acquiring these kind of images [13,14]. This technique estimates the object surface normals by imaging it under different lighting conditions [15]. Photometric stereo images can be used as input to a deep learning model for better defect classification. For example, a commonly used approach for classification tasks is convolutional neural networks (CNN). There are different industrial applications that combine photometric stereo acquisition and CNNs applied to defect detection. For example [16] applies this method in order to find rail defects. By means of differently colored light-sources illuminating the rail surfaces from different and constant directions, cavities are made visible in a photometric dark-field setup. Then, they experimented with classical CNNs trained in purely supervised way and also explored the impact of regularization methods, such as unsupervised layer-wise pre-training and training data-set augmentation.

In the detection of defects in the manufacturing industry, it is often also interesting to know the geometry of the defect in order to estimate its size and classify it based on criteria established by the client. For this purpose, defect detection is usually carried out by means of pixel-level segmentation. For example, [17] uses a UNet architecture for the localisation and segmentation of defects in metallic components, achieving a Dice value of 0.9167 [18]. Alternatively, [19] also propose a precise pixel-level segmentation of surface defects, using segmentation networks with different modules that allow a fast and accurate segmentation.

1.2. Main Contributions

Related work shows the ability of convolutional neural networks for automatic defect detection tasks but also the importance of the image quality, which is highly related with having an adequate acquisition set up according to the characteristics of the surface to be inspected. Specially in specular surfaces, topographic information is very relevant to defect characterization, thus, being a valuable source of information for its detection. Additionally, processing time is critical in order to cope with the inspection of components

in high-production-rate scenarios. Therefore, the challenges that our proposed method must overcome are the following:

1. Obtain high resolution images of non-Lambertian surfaces without light reflections.
2. Use both topographic and spectral (color) information for surface analysis and defect detection.
3. Obtain a robust neural model able to detect all defects without obtaining a high false rejection rate.
4. Obtain a neural model with low processing time.

In order to overcome these challenges, we propose a system based on diffuse illumination over non-Lambertian surfaces with a white light source for extracting topographic surface information and spectral (color) response at the same time. We propose combining these sources of information into a compact RGB image format for feeding a segmentation CNN for surface defect detection. We demonstrate the benefit of this combination compared with the use of each source of information individually. We integrated the convolutional segmentation network and the photometric stereo optical sensor in an automatic inspection system for product quality assurance in a manufacturing process of metallic pieces with highly specular surface finishing and high production rates.

The structure of this paper is as follows: Section 2 describes our approach for automatic metallic component inspection based on photometric stereo images and a convolutional neural network. Section 3 describes the results obtained during the evaluation of the proposed approach. Section 4 describes the system implementation and an analysis of its performance. Subsequently, a discussion is presented in Section 5. Finally, conclusions are depicted in Section 6.

2. Materials and Methods

The manufactured components under analysis in this work are machined nickel-plated components that may show different types of small defects on their surfaces. These components have a dimension of 50 mm $\times$ 7 mm and a thickness of 2 mm. Some samples of these defects are shown in Figure 1.

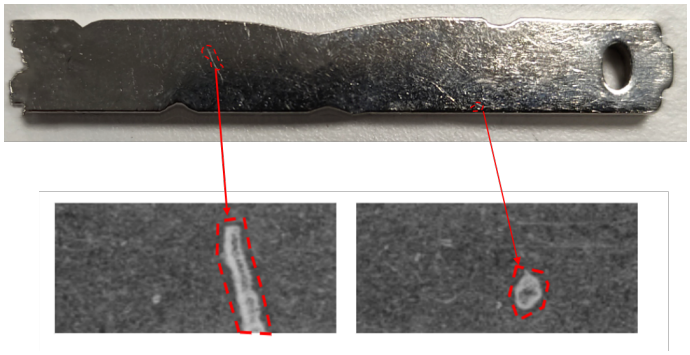


Figure 1. (**Top**) Conventional (non-photometric stereo) image of a manufactured component. (**Bottom**) Samples of component defects acquired with photometric stereo imaging.

The nature of the defects is caused by the customer's manufacturing process. This process performs some cuttings and machining on the component causing surface defects. The sizes and types of defects to be detected are established by the customer's quality control experts.

These components, due to nickel-plated process and some coatings, have very high specular indexes resulting in very shiny components. In order to obtain high resolution and high quality images of components surfaces, we propose to use an approach based on photometric stereo.

2.1. Photometric Stereo Image Acquisition

Reflectance-based shape recovery of non-planar surfaces from several reflectance images obtained under different irradiance sources is a classic task in computer vision [15]. This type of approach determines the absolute depth of the surfaces by reconstructing the shape of the object under changing illumination conditions such as color or orientation, among others. This problem is called shape from shading when just one irradiance image is used for reconstruction process [20]. Photometric stereo methods firstly recover surface orientations and can be combined with an integration method to calculate a height or depth map. Even without a subsequent integration step the surface orientations can be used, for example, to determine the curvature parameters of object surfaces [21]. To acquire images for photometric stereo the object is consecutively illuminated by several light sources. In our approach we use four different light sources from four different orientations around the component to be inspected.

In this way, we get five different photometric images for each acquisition:

- Curvature images: Provides the contour lines of the surface topography.
- Texture images: Provides color or spectral response.
- Gradient image X: Signal variation in x direction.
- Gradient image Y: Signal variation in y direction.
- Range images: Computed as the image gradient magnitude. It highlights information about the changes in the intensity of the image.

To achieve these images, a system of equations introduced by Woodham [15] formalizes the solution of the problem assuming Lambertian reflectance, i.e., that the distant light sources are known point sources. Given a known vector $\vec{I}$ of i observed intensities, the known matrix of normalised light directions $[\vec{L}\,] = (L^1, L^2, L^3)^T$ and the reflectivity ρ, the unknown surface normal $\vec{n}$ can be obtained by inverting the following lineal equation:

$$\vec{I} = \rho[\vec{L}\,]\vec{n} \tag{1}$$

If the three illumination vectors L^k do not lie in the same plane, the matrix $[\vec{L}\,]$ is non-singular and can be inverted, giving the following equation:

$$[\vec{L}\,]^{-1}\vec{I} = \rho\vec{n} \tag{2}$$

As $\vec{n}$ has a unit length, we can estimate the surface normal and the albedo. The problem comes when we have more than three input images; in this case the illuminations matrix $[\vec{L}\,]$ would no longer be square, and therefore could not be inverted [22].

When the light sources are more than three, this problem is solved by least squares, using the Moore–Penrose pseudo-inverse [23] in Equation (1), thus obtaining the following solution:

$$[\vec{L}\,]^T\vec{I} = [\vec{L}\,]^T[\vec{L}\,]\rho\vec{n} \tag{3}$$

$$\rho\vec{n} = ([\vec{L}\,]^T[\vec{L}\,])^{-1}[\vec{L}\,]^T\vec{I} \tag{4}$$

where $([\vec{L}\,]^T[\vec{L}\,])^{-1}[\vec{L}\,]^T$ is the Moore–Penrose pseudo-inverse. After that, ρ and $\vec{n}$ can be solved as before.

2.2. Defect Segmentation Model

The proposed model is a deep network, trained with a pixel-wise loss, that considers each pixel as an individual training sample that achieves an accurate defect segmentation, adapted to its shape. With this approach we also reduce the size of the training dataset from approaches based on sliding windows. Another goal that the network has to achieve is to be able to learn the contours and engravings of each component, which are different for each product reference.

The architecture is shown in Figure 2. The proposed network consists of six sequential layers, each composed by a combination of convolutional, normalization and max pooling operations, with an image input size of 512 × 2500. In total, the network has 19 convolutional layers for feature extraction and 2 max-pooling layers for down-sampling feature maps, combined with batch normalization and ReLU layers, in order to improve the speed, performance and stability of the network. The kernel sizes are different for each layer. In the last layers, the size is fixed and not reduced in order to have a more accurate segmentation. With this approach, we have the speed of a simple architecture with the segmentation resolution of a complex one.

This architecture is optimized for the segmentation of small defects that appears in the components.

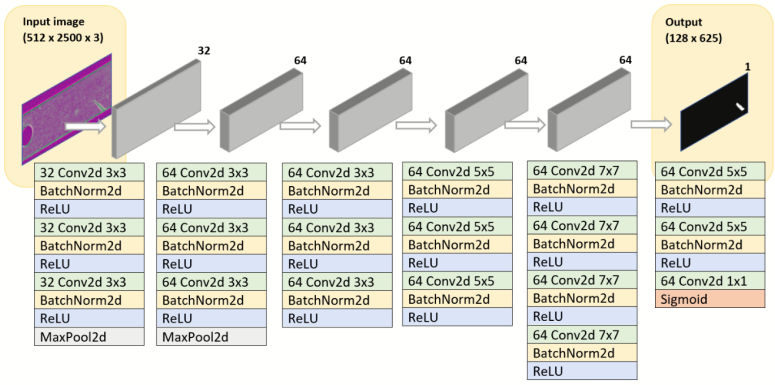


Figure 2. Defect segmentation network architecture.

3. Results

3.1. Data Set Generation

The images acquired in this application are images of metal components with different coatings: nickel (NI), nickel–silver (NS) and Ni7 nickel–silver (Ni7). Depending on the coating's material, the light reflected to the camera may vary, resulting in different light intensity and color detected by the sensor. Therefore, the acquisition of each material looks different and must be treated individually to obtain optimal results, thus resulting in three different sub-datasets. An example of the differences in the surface appearance is shown in Figure 3.

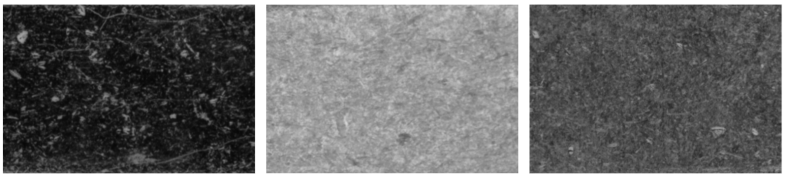


Figure 3. Zoom of a component region showing different texture image appearance caused by different coatings: nickel (NI), nickel–silver (NS) and Ni7 nickel–silver (Ni7).

Our data set contains the samples shown in Table 1. The images were acquired and annotated by the quality experts of the manufacturing company.

Table 1. Number of acquired training images for each coating and each defect type.

Number of Samples in the Training Dataset			
Coating Type	**Non-Defective Samples**	**Bump Marks and Scratches Samples**	**Total Images**
nickel (NI)	427	464	891
nickel–silver (NS)	489	470	959
Ni7 nickel–silver (Ni7)	465	434	899

In order to obtain a model robust to small changes that may occur during the acquisition step, we applied a data augmentation process. These type of techniques are very powerful to artificially create variations of existing images or samples by applying geometric and photometric transformations [24]. These transformations have to be adapted to the specific use case.

In our use case, we apply photometric transformations to simulate small changes on the surface coating that components may have. A change in the surface coating supposes a different response in the incident light and, therefore, differences in the acquired image brightness. Although we separate the classes into different subgroups, variations of this type may occur within each class.

We also apply geometric transformations such as horizontal and vertical flipping and image rotations of some degrees. In our case, the component is always placed horizontally, so we apply only ±2 degrees in-plane rotation in order to resemble the real manufacturing conditions.

To evaluate the trained models, new samples of the three materials were captured as described in Table 2. These testing samples were intended to be composed with equal proportions of defective and non-defective samples.

Table 2. Number of test images for each coating.

Number of Samples in the Testing Dataset	
Coating Type	**Samples**
nickel (NI)	297
nickel–silver (NS)	417
Ni7 nickel–silver (Ni7)	320

3.2. Performance Criteria

In the field of surface defect detection, the following statistics are often used to evaluate the results obtained:

- True Positives (*TP*): the defect is detected as a defect.
- True Negatives (*TN*): the background is detected as background.
- False Positives (*FP*): the background is mistakenly detected as a defect.
- False Negatives (*FN*): the defect is mistakenly detected as background.

With the values of the statistics described above, the following values commonly used to evaluate the quality of the segmentation can be calculated [25]:

- Mean Dice Value: is an spatial overlap-based metric. It focuses on measuring the similarity between two samples through the union and intersection of sets of predicted and ground truth pixels. This index takes values between 0 and 1, and is better when it is closer to 1, since this means having more surface in common between the ground truth and the result of the segmentation. This coefficient is calculated by the Equation (5), where A and B are the ground truth mask and the predicted mask, respectively.

$$Dice(A, B) = \frac{2\|A \bigcap B\|}{\|A\| + \|B\|} \quad (5)$$

- Sensitivity: is the ability of the model in not marking a negative sample as positive. It is measured by the formula of Equation (6).

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{6}$$

- Specificity: is the ability to find all positive samples. It is measured by the formula of Equation (7).

$$Specificity = \frac{TN}{(TN + FP)} \tag{7}$$

- Pixel accuracy: the percent of pixels in the image that are classified correctly. It is calculated by the Equation (8).

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{8}$$

In our use case, the criterion for positive detection is a Dice index greater than 0.5. The results shown in the following tables were obtained using this criteria.

3.3. Combination of Photometrtic Stereo Images for Defect Detection

One of our goals is to exploit the full potential of photometric stereo images. We propose creating RGB images by embedding different photometric stereo images in each color channel.

In order to compose the best combination of images, a study of their individual usefulness for defect prediction was performed. For this purpose, we trained the customized segmentation model in each image independently, obtaining the segmentation accuracy in each case, using the same NI dataset. Based on the results, the three channels that obtained the best results were texture, range and curvature as shown in Table 3.

Table 3. Defect segmentation results for each photometric stereo image using our customized segmentation network.

Segmentation Results for Each Photometric Stereo Image in NI Dataset	
Image Type	**Accuracy**
texture	0.9259
range	0.9482
curvature	0.9326
gradient X	0.7865
gradient Y	0.8103

We think that these three images contain better information for defect representation than gradient X and gradient Y. Given these results, we propose combining texture, range and curvature images into an RGB image for training the segmentation network, as shown in Figure 4.

3.4. Comparison of Segmentation Results between RGB and Texture-Only Images

The defect segmentation network of each material was trained using two different types of datasets: a dataset with a single channel images containing, only texture information, and a dataset with RGB images composed of all the photometric stereo information, i.e., with texture, range and curvature images combined. The single-channel images can resemble the images acquired by a conventional non-photometric diffuse illumination dome. With this experiment, we evaluated the potential benefits of using photometric-based imaging, comparing it with a more conventional non-photometric based approach.

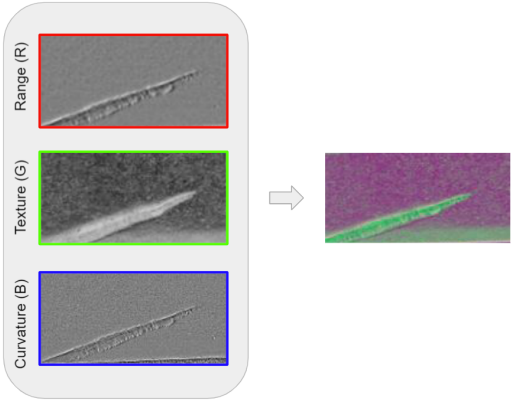


Figure 4. RGB image with the selected layers of photometric stereo acquisition.

As shown the results in Tables 4–6, respectively, all materials demonstrate an improvement in all metrics of the use of the RGB representation over the use of texture-only images. With the use of image combination, the false positive rate is reduced to approximately half for all materials. In addition, an increase in the rate of true positives is observed. These values are especially relevant in the industrial manufacturing environment. With respect to the Dice value, an increase of 9%, 6% and 4% is observed for NI, NS and NI7 respectively.

Table 4. Nickel material segmentation results.

Value	NI (RGB)	NI (Texture Only)
Dice mean	0.8957	0.8027
sensitivity	0.98	0.9477
specificity	0.931	0.9027
accuracy	0.956	0.9259
TP	150	145
TN	134	130
FP	10	14
FN	3	8

Table 5. Nickel Silver material segmentation results.

Value	NS (RGB)	NS (Texture Only)
Dice mean	0.829	0.767
sensitivity	0.954	0.9081
specificity	0.883	0.819
accuracy	0.916	0.8609
TP	186	178
TN	196	181
FP	26	40
FN	9	18

3.5. Defect Segmentation Performance for Each Material Using RGB Images

We evaluated the performance of the defect segmentation network using the RGB images composed by texture, curvature and range in each material. As shown in Table 7, the models reach a Dice value higher than 0.82, reaching a value of up to 0.94 in NI7. Regarding the performance of the algorithm in the inspection line, the results obtained on the test set reveal the good performance of the model. The achieved sensitivity and specificity ensure the defect detection without increasing the false rejection rate.

Table 6. Ni7 material segmentation results.

Value	NI7 (RGB)	NI7 (Texture Only)
Dice Mean	0.9739	0.9619
sensitivity	0.9855	0.9710
specificity	0.9615	0.9505
accuracy	0.9690	0.9593
TP	136	134
TN	175	173
FP	7	9
FN	2	4

Table 7. Material Segmentation results.

Value	NI	NS	NI7
Dice Mean	0.8957	0.829	0.9739
sensitivity	0.98	0.954	0.9855
specificity	0.931	0.883	0.9615
accuracy	0.956	0.916	0.9690
TP	150	186	136
TN	134	196	175
FP	10	26	7
FN	3	9	2

3.6. Comparison against Two Benchmark Segmentation Networks

To test the performance of our proposed network, we compared our approach with two well-known segmentation networks: DFANet [26] and UNet [27].

DFANet is an efficient segmentation architecture designed to be agile and to run with the minimum necessary resources. It is based on multi-scale feature propagation, thus reducing the number of parameters. Its design is intended to achieve a good trade off between speed and performance in segmentation.

We also use UNet, which is a network designed to be trained end-to-end with few image samples. In addition, thanks to its design for the medical field, it allows obtaining a very accurate detection result.

The results show that our customized segmentation network achieves better results in defect segmentation of the components, as shown in Table 8. DFANet obtains a higher number of false positives because its structure detects impurities on the non-defective surfaces that are not considered real defects. UNet obtains better results than DFANet, but fails in the detection of the defects, with only dimensional affection. If a defect does not show relevance in the texture image, this network is not able to detect it correctly, thus causing false negatives.

Table 8. Defect segmentation results using DFANet and UNet on NI dataset.

Segmentation Results Using NI Dataset	
Neural Network	**Accuracy**
DFANet	0.8732
UNet	0.9113
Our network	0.9560

4. System Implementation

Once the feasibility of the defect detection algorithms and the suitability of the acquisition system had been verified, the whole system architecture was designed and developed. Overall, the system was composed of several interoperable modules such as, image acquisition, image annotation and image processing in a distributed manner. Figure 5 shows an overview/scheme of the relation between the developed modules.

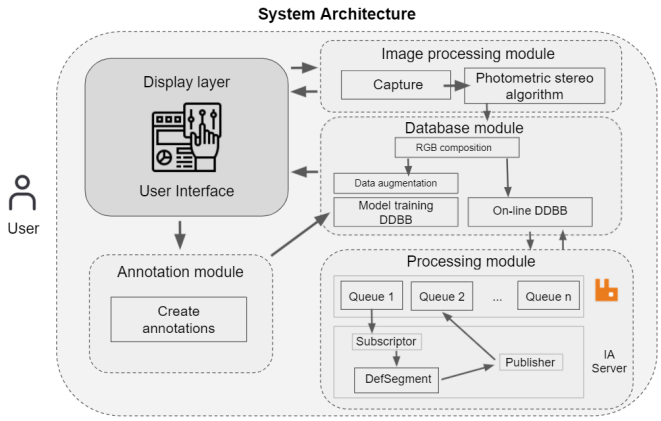


Figure 5. System architecture for defect detection in steel components.

4.1. Acquisition Set Up

For image acquisition, we combine a photometric stereo diffuse dome and a high resolution linear sensor with a telecentric lens. The field of view was chosen based on the maximum dimension of the component to be analyzed, in our case, 200 mm. To allow the inspection of small defects with enough detail (<1 mm), each line capture of the components is acquired at 4 K. Therefore, we obtain a very high-resolution capture of the component surface. This set up is shown in Figure 6.

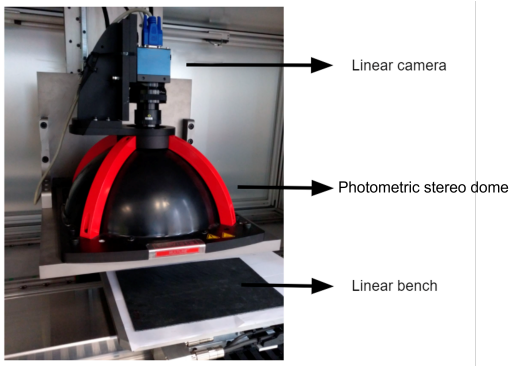


Figure 6. Acquisition set up to capture photometric stereo images.

4.2. Annotation Module

This module provides an annotation tool for capturing the knowledge of quality inspection experts. As previously described in Section 2.1, we acquire five input images during the photometric stereo surface reconstruction. In order to take advantage of this information we have developed a viewer that allows visualizing the different image layers synchronously. All types of defects are not always visible in the five photometric stereo images. For example, defects that are dimensionally affected, such as scratches, are more visible in the range and curvature images. In contrast, defects without topographic involvement, such as oxide, are more visible in the texture images. The main advantage of our customized annotation tool is that the user can visualize and interact with one, two or three channels of the same image independently. A shown in Figure 7, it allows the quality

expert to discern both type of defects at the same time in a side by side view of different photometric stereo images of the same component.

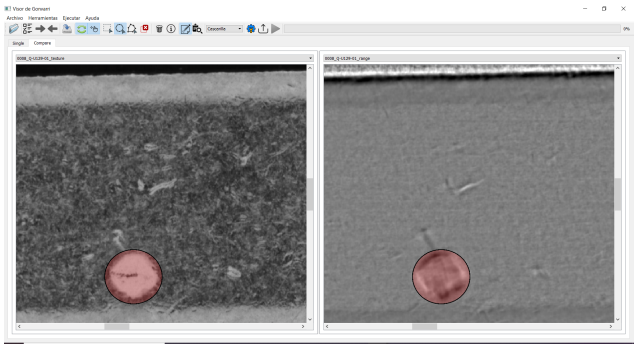


Figure 7. Customized annotation tool showing a side by side view of texture channel (**left**) and range channel (**right**) of an interest region of the component.

4.3. System Performance

Besides the accuracy of the segmentation network, another fundamental point for the integration of any application in the industry is the processing time of the inspection analysis. It is mandatory that the time spent in this analysis to be under a certain cycle time, in order to allow the continuous production of the components. In this regard, a comparison of the processing execution time between an industrial computer without a GPU and an external server with a GPU is made. The external server has an NVIDIA RTX 2070, allowing performing the inference step faster than a CPU.

Our system uses a message broker as a middleware to orchestrate and communicate all the processes involved in a distributed manner. This tool facilitates the definition of exchanges and communication channels between the different modules and processes with the use of message queues. In this case, the acquisition and inspection processes are performed in different computers with dedicated hardware for each task. Thus, we can leverage the use of specialized hardware resources such as GPUs for the inference computation for our network and easily scale when required.

Performance evaluation is measured for the following processes: inference and persistence (storing the result in the database). The time breakdown for each process is shown in Table 9. In summary, the total processing time executed locally in a single PC with a CPU is 707ms. In contrast, executed in a distributed manner on the external server with dedicated GPU, the processing time is 138ms. If the application is executed locally, the processing time does not reach the required production rate, so it is necessary to distribute the processing task into an external server. This action implies a five-fold reduction of the execution time.

Table 9. Processing times.

Action	Time CPU (ms)	Time GPU (ms)
inference time	471	126
DDBB storage	236	12
total time (ms)	707	138

5. Discussion

As demonstrated by experimentation and results shown in previous sections, the use of a photometric stereo sensor is adequate for the application of defect detection on reflective surfaces. The proposed method of stacking multiple sources of data from the photometric stereo sensor into a high-resolution, multi-channel (RGB) image supposes an improvement in the accuracy and performance achieved with the presented convolutional neural network. A typical problem with this type of application is the detection of very

small defective regions with respect to the dimension of the component. This makes it difficult to avoid false positives, as the dataset is unbalanced between the amount of normal and defect pixels. Therefore, the extra information provided by the multiple data layers from the photometric stereo allows the model to extract features related with defective regions in a more optimal way.

Another challenge to overcome in machine vision applications developed for industrial manufacturing processes is the speed at which they must operate to satisfy high production rates. Defect segmentation applications are often not fast enough to be integrated into an industrial production line. However, our proposed system is able to process images in the required cycle time. We demonstrate that, if the inference is performed on an industrial computer (CPU), the required time is not achieved. Hence, a distributed computation for parallelizing processes is mandatory. This is also a benefit in terms of organization and scalability, since the same system can inspect components in several production lines and more hardware resources can be easily added to cope with the required compute demand.

In comparison to other inspection methods proposed in the state of the art, our method allows the inspection of reflective surfaces in the cycle time required by the manufacturing industry. Other methods propose multiple image acquisitions from different angles to avoid surface reflections, thus affecting the acquisition time and also the required space for the installation of the inspection station in the production line. Compared with the methods that employ deep learning, our method is novel in using multiple combined sources of information, allowing a more detailed and robust description of the component surface. This feature gives greater stability in defect detection compared with traditional image acquisition-based systems. In addition, a specific segmentation network is created, allowing the exploitation of this information in a fast and accurate way. The proposed acquisition system can be also used on non-planar geometries, such as cylindrical components, just by modifying the automation of the component's displacement. The proposed method has disadvantageous in terms of hardware costs and the level of automation required for image acquisition. A photometric stereo dome is electronically more complex than conventional illumination, supposing a cost penalty. Furthermore, automation is needed to move the component in a constant and controlled way, which must be perfectly synchronised with the trigger of the camera and the dome.

6. Conclusions

This work proposes an automatic system for the quality inspection of metallic components using a photometric stereo based sensor and a customised segmentation model. We propose combining the photometric stereo information that better resemble the defects. For this purpose, a comparison analysis was carried out for each photometric stereo image, which allows choosing the most suitable image combination.

The photometric stereo image combination provides additional information to the segmentation network increasing the effectiveness and accuracy of the defect detection. This approach is confirmed in different materials showing a decrease in false rejection rate and an increase in the true detection rate.

A performance comparison with different well-known segmentation architectures was carried out to demonstrate the suitability of our proposed segmentation network.

Finally, some performance tests were carried out in terms of execution time. These tests demonstrate the need of accelerated hardware and parallel processing capabilities, such as, making use of distributed computing and GPU devices to fulfill the required speed in the production line.

Author Contributions: Conceptualization, F.A.S., I.B., A.A. and M.G.; methodology, F.A.S.; software, F.A.S. and A.A.; validation, F.A.S. and A.A.; investigation, F.A.S., I.B. and A.A.; resources, I.B.; writing—original draft preparation, F.A.S., I.B., A.A. and M.G.; writing—review and editing, F.A.S., I.B., A.A. and M.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kopardekar, P.; Mital, A.; Anand, S. Manual, hybrid and automated inspection literature and current research. *Integr. Manuf. Syst.* **1993**, *4*, *18–29*. [CrossRef]
2. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]
3. Hebb, D.O. *The Organisation of Behaviour: A Neuropsychological Theory*; Science Editions: New York, NY, USA, 1949.
4. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386. [CrossRef] [PubMed]
5. Ivakhnenko, A. *Cybernetic Predicting Devices*; Technical Report; Purdue University: Indiana, IN, USA, 1966.
6. Piccialli, F.; Di Somma, V.; Giampaolo, F.; Cuomo, S.; Fortino, G. A survey on deep learning in medicine: Why, how and when? *Inf. Fusion* **2021**, *66*, 111–137. [CrossRef]
7. Minaee, S.; Abdolrashidi, A.; Su, H.; Bennamoun, M.; Zhang, D. Biometrics Recognition Using Deep Learning: A Survey. *arXiv* **2021**, arXiv:1912.00271.
8. Mouha, R.A. Deep Learning for Robotics. *J. Data Anal. Inf. Process.* **2021**, *9*, 63. [CrossRef]
9. Chilimbi, T.; Suzue, Y.; Apacible, J.; Kalyanaraman, K. Project Adam: Building an Efficient and Scalable Deep Learning Training System. In Proceedings of the 11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14), Broomfield, CO, USA, 6–8 October 2014; USENIX Association: Broomfield, CO, USA, 2014; pp. 571–582.
10. Gjoreski, H.; Bizjak, J.; Gjoreski, M.; Gams, M. Comparing deep and classical machine learning methods for human activity recognition using wrist accelerometer. In Proceedings of the IJCAI 2016 Workshop on Deep Learning for Artificial Intelligence, New York, NY, USA, 9–15 July 2016; Volume 10.
11. Artusi, A.; Banterle, F.; Chetverikov, D. *A Survey of Specularity Removal Methods*; Computer Graphics Forum; Wiley Online Library: Hoboken, NJ, USA, 2011; Volume 30, pp. 2208–2230.
12. Shen, H.L.; Cai, Q.Y. Simple and efficient method for specularity removal in a image. *Appl. Opt.* **2009**, *48*, 2711–2719. [CrossRef] [PubMed]
13. Huang, S.; Xu, K.; Li, M.; Wu, M. Improved Visual Inspection through 3D Image Reconstruction of Defects Based on the Photometric Stereo Technique. *Sensors* **2019**, *19*, 4970. [CrossRef] [PubMed]
14. Fang, X.; Luo, Q.; Zhou, B.; Li, C.; Tian, L. Research progress of automated visual surface defect detection for industrial metal planar materials. *Sensors* **2020**, *20*, 5136. [CrossRef] [PubMed]
15. Woodham, R.J. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Image Understanding Systems and Industrial Applications I*; International Society for Optics and Photonics: Washington, DC, USA, 1979; Volume 155, pp. 136–143.
16. Soukup, D.; Huber-Mörk, R. Convolutional neural networks for steel surface defect detection from photometric stereo images. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 8–10 December 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 668–677.
17. Aslam, Y.; Santhi, N.; Ramasamy, N.; Ramar, K. Localization and segmentation of metal cracks using deep learning. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 4205–4213. [CrossRef]
18. Zou, K.H.; Warfield, S.K.; Bharatha, A.; Tempany, C.M.; Kaus, M.R.; Haker, S.J.; Wells, W.M., III; Jolesz, F.A.; Kikinis, R. Statistical validation of image segmentation quality based on a spatial overlap index1: Scientific reports. *Acad. Radiol.* **2004**, *11*, 178–189. [CrossRef]
19. Wu, X.; Qiu, L.; Gu, X.; Long, Z. Deep Learning-Based Generic Automatic Surface Defect Inspection (ASDI) With Pixelwise Segmentation. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–10. [CrossRef]
20. Horn, B.K.; Brooks, M.J. *Shape from Shading*; MIT Press: Cambridge, MA, USA, 1989.
21. Klette, R.; Kozera, R.; Schlüns, K. Shape from shading and photometric stereo methods. *Handb. Comput. Vis. Appl.* **1998**, *2*, 531–590.
22. Barsky, S.; Petrou, M. The 4-Source Photometric Stereo Technique for Three-Dimensional Surfaces in the Presence of Highlights and Shadows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1239–1252. [CrossRef]
23. Penrose, R. A generalized inverse for matrices. In *Mathematical Proceedings of the Cambridge Philosophical Society*; Cambridge University Press: Cambridge, UK, 1955; Volume 51, pp. 406–413.
24. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 1–48. [CrossRef]
25. Wang, Z.; Wang, E.; Zhu, Y. Image segmentation evaluation: A survey of methods. *Artif. Intell. Rev.* **2020**, *53*, 5637–5674. [CrossRef]

26. Li, H.; Xiong, P.; Fan, H.; Sun, J. DFANet: Deep Feature Aggregation for Real-Time Semantic Segmentation. *arXiv* **2019**, arXiv:1904.02216.
27. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.