

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.



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

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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

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 °C 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.

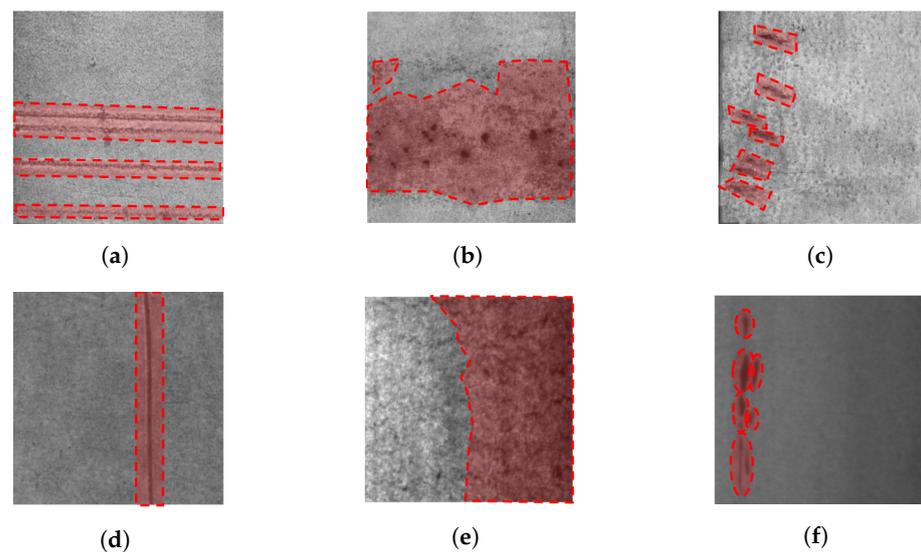


Figure 1. Samples of four types of steel surface defects. (a) finishing roll printing; (b) oxide; (c) patches; (d) scratches; (e) crazing; and (f) inclusions.

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.

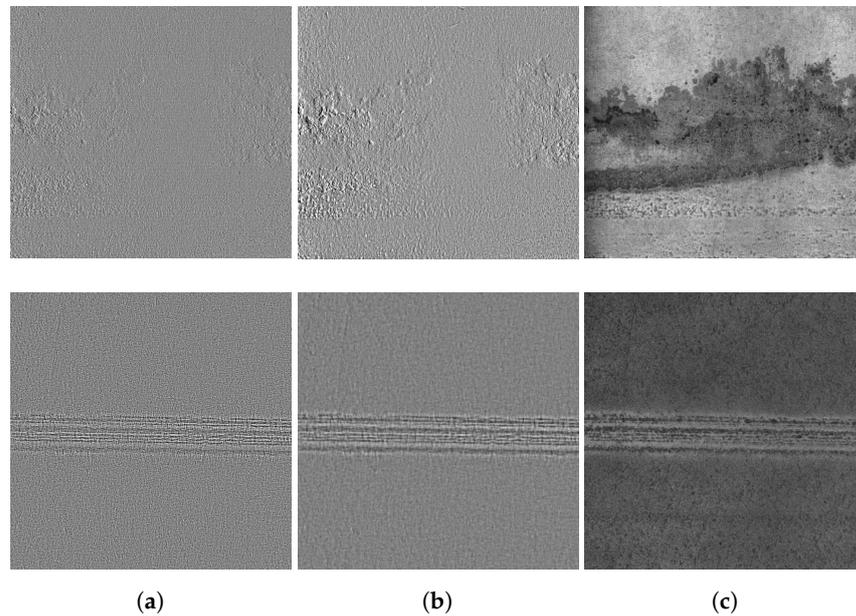


Figure 3. Different image types obtained by the photometric-stereo system of two different steel sheet samples (one sample in each row); (a) curvature image; (b) range image; and (c) texture image.

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.

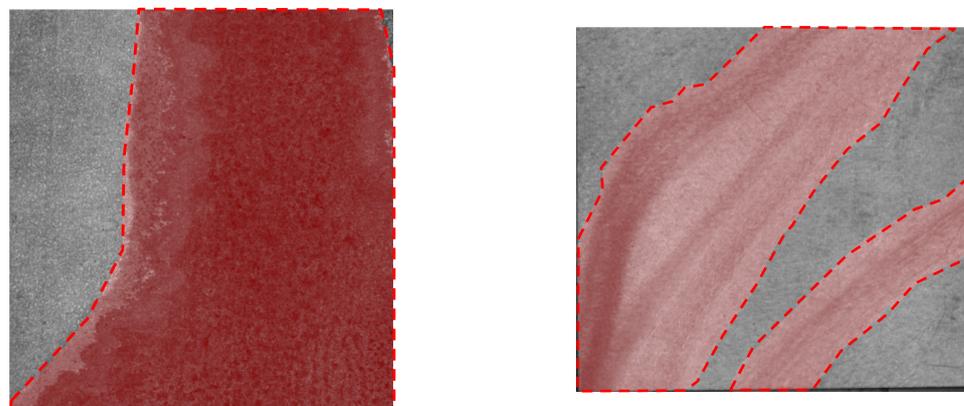


Figure 4. Texture channel of RGB photometric-stereo images with corresponding ground truth masks in red and dotted lines.

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.

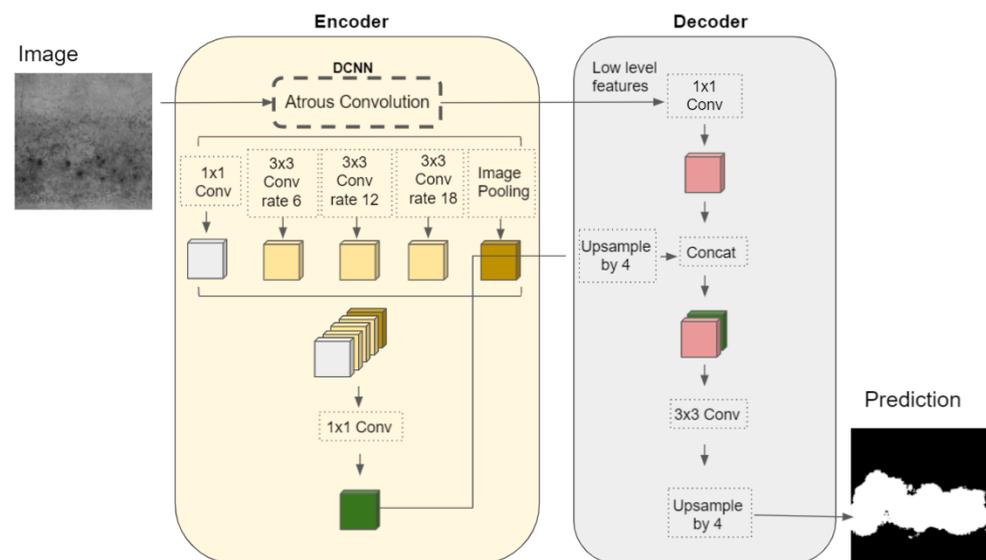


Figure 5. Architecture diagram of the DeepLabV3+ network [26].

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.

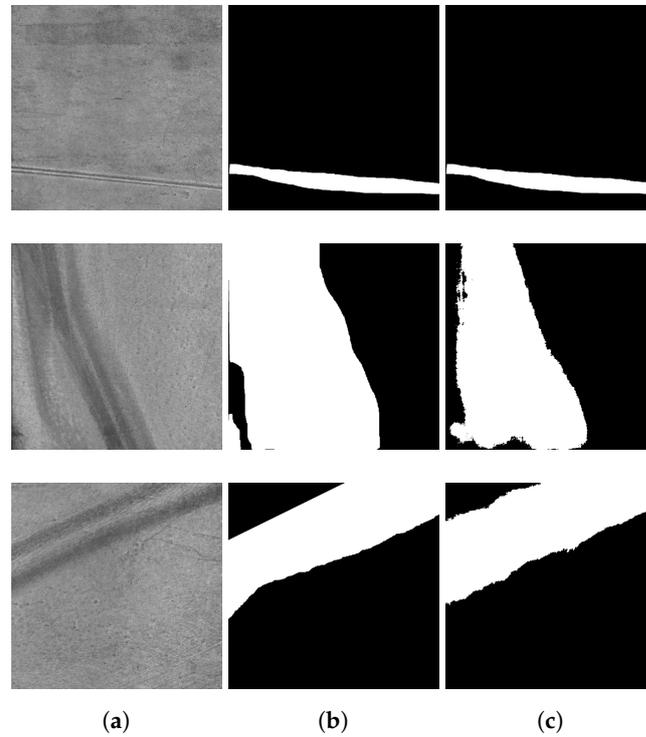


Figure 6. Some visual results of defect detection by the DeepLabv3+ trained on the conventional augmented dataset. (a) photometric-stereo texture channel images; (b) ground truth defect masks; and (c) predicted defect masks.

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.

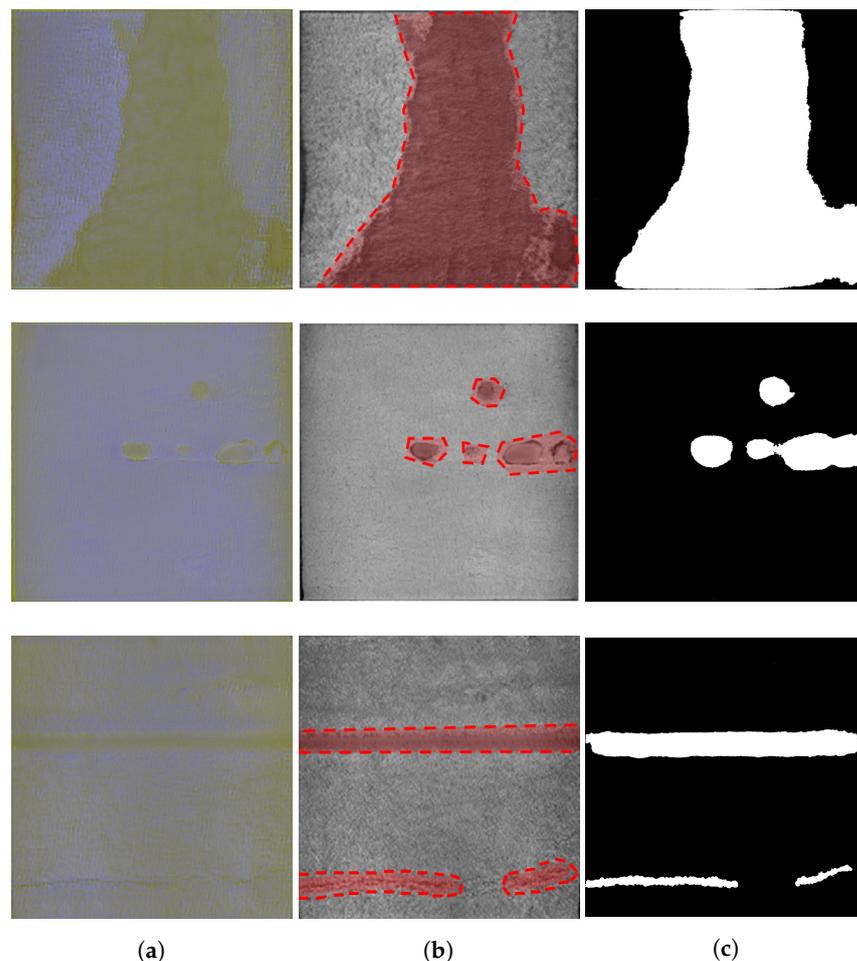


Figure 7. (a) Synthetic RGB images; (b) texture channel of Synthetic images with the defects in red (c) ground truth masks.

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.

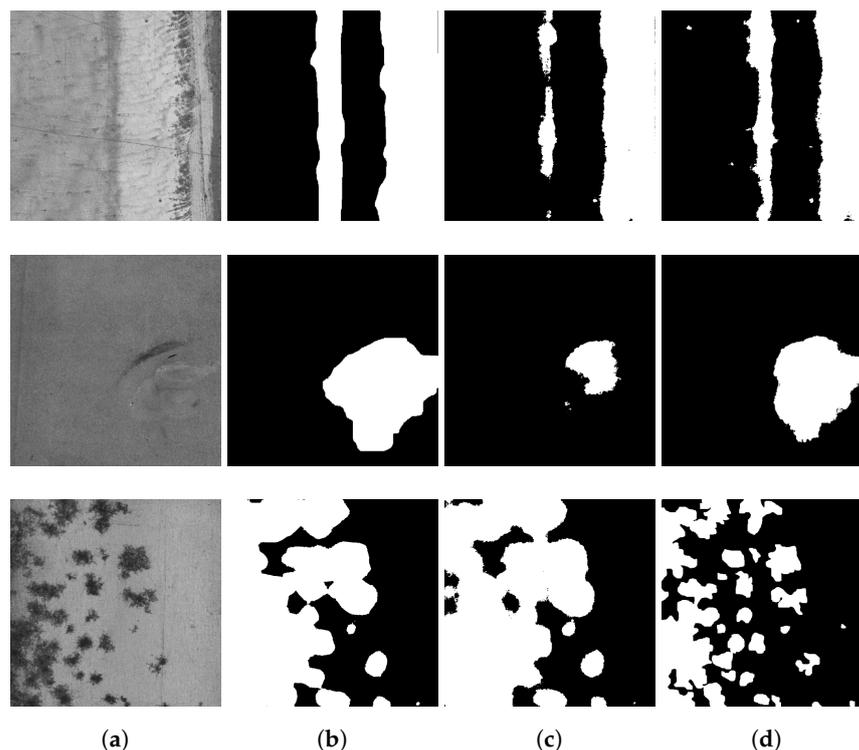


Figure 8. Detection of defects using two instances of the DeepLabv3+ semantic segmentation neural network. (a) photometric-stereo texture channel images; (b) ground truth masks; (c) predictions with the original RGB photometric-stereo database; and (d) predictions adding GAN-generated RGB photometric-stereo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

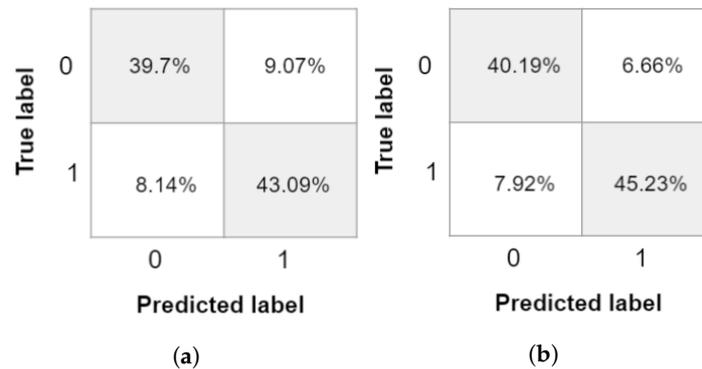


Figure 9. Confusion matrices of the test set, where label 0 represents background and label 1 means defect. (a) Confusion matrix of DeepLabV3+ trained only with real images and (b) confusion matrix of DeepLabV3+ trained with the original dataset augmented with synthetic images generated by the GAN with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

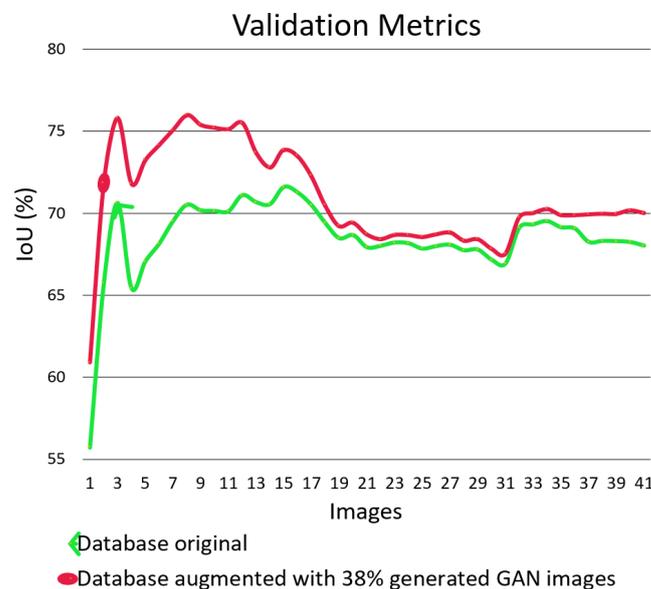


Figure 10. Mean of IoU (%) result with original dataset augmented with a conventional method and the original augmented with synthetic images generated by the GAN with a 38% ratio.

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.

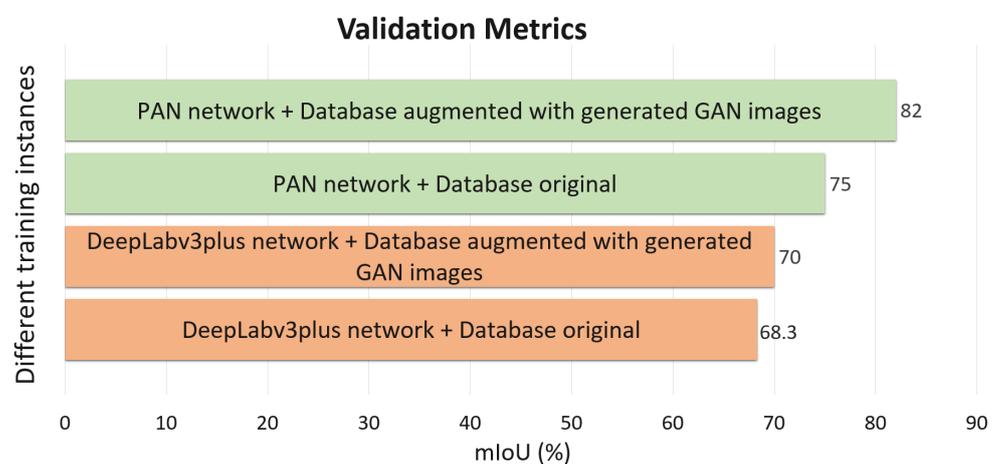


Figure 11. Mean of IoU (%) during evaluation with the original photometric-stereo RGB augmented database and the original one plus synthetic images generated by the GAN with a 38% ratio.

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.

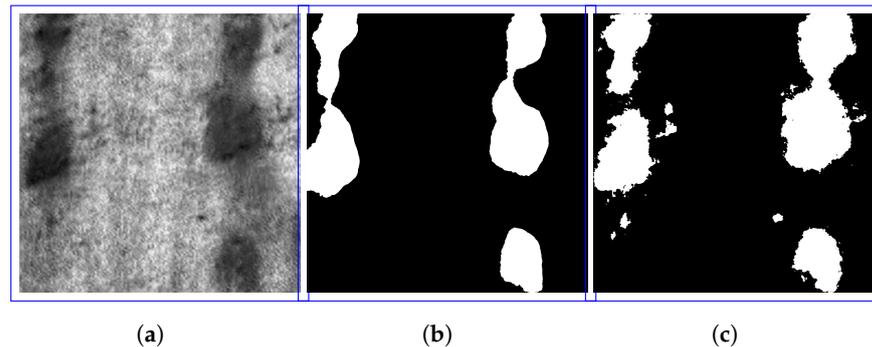


Figure 12. Results of defect segmentation by the PAN trained on the synthetically augmented NEU dataset. (a) StyleGAN2 generated synthetic image; (b) ground truth defect mask; and (c) predicted defect mask.

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