

Article

Dual Attention-Based Industrial Surface Defect Detection with Consistency Loss

Xuyang Li ^{1,†}, Yu Zheng ^{2,†}, Bei Chen ¹ and Enrang Zheng ^{1,*}
¹ School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China; 200611021@sust.edu.cn (X.L.); chenbei@sust.edu.cn (B.C.)

² School of Cyber Engineering, Xidian University, Xi'an 710126, China; yuzheng.xidian@gmail.com

* Correspondence: zhenger@sust.edu.cn

† These authors contributed equally to this work.

Abstract: In industrial production, flaws and defects inevitably appear on surfaces, resulting in unqualified products. Therefore, surface defect detection plays a key role in ensuring industrial product quality and maintaining industrial production lines. However, surface defects on different products have different manifestations, so it is difficult to regard all defective products as being within one category that has common characteristics. Defective products are also often rare in industrial production, making it difficult to collect enough samples. Therefore, it is appropriate to view the surface defect detection problem as a semi-supervised anomaly detection problem. In this paper, we propose an anomaly detection method that is based on dual attention and consistency loss to accomplish the task of surface defect detection. At the reconstruction stage, we employed both channel attention and pixel attention so that the network could learn more robust normal image reconstruction, which could in turn help to separate images of defects from defect-free images. Moreover, we proposed a consistency loss function that could exploit the differences between the multiple modalities of the images to improve the performance of the anomaly detection. Our experimental results showed that the proposed method could achieve a superior performance compared to the existing anomaly detection-based methods using the Magnetic Tile and MVTec AD datasets.

Keywords: surface defect detection; anomaly detection; industrial security; attention mechanism



Citation: Li, X.; Zheng, Y.; Chen, B.; Zheng, E. Dual Attention-Based Industrial Surface Defect Detection with Consistency Loss. *Sensors* **2022**, *22*, 5141. <https://doi.org/10.3390/s22145141>

Academic Editor: Paolo Bellavista

Received: 30 May 2022

Accepted: 4 July 2022

Published: 8 July 2022

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



Copyright: © 2022 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/>).

1. Introduction

Over recent years, surface defect detection has attracted attention in various fields, such as transportation [1–3], agriculture [4,5] and biomedicine [6,7], but surface defect detection has been especially extensively studied within manufacturing [8–12]. The process of industrial production is often accompanied by quality problems among the manufactured products and not all products can be monitored for quality through appearance observation. Therefore, surface defect detection plays an important role in industrial production. However, the surface defect detection of industrial products suffers from two main problems. First, a lack of defect instances: defective samples are usually rare among industrial products, while normal samples are common. Thus, it is difficult to collect enough defective samples and in extreme cases, only normal samples can be obtained. Second, the diverse types of defects, as shown in Figure 1: there are various types of defects among industrial products and the appearance of defects is not necessarily uniform on the same product. As a result, it is difficult to treat all defective products as one valid category. Under these circumstances, it is more appropriate to view the surface defect detection problem as a semi-supervised anomaly detection problem.

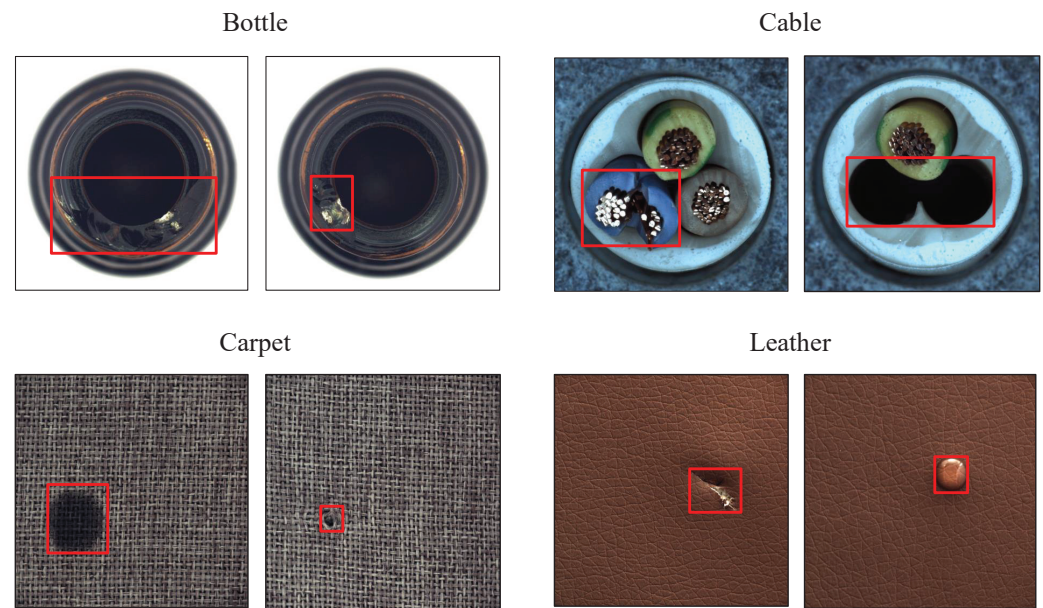


Figure 1. Industrial product samples with surface defects: each sub-figure represents a different industrial product (four in total) and each picture represents a different defect. The area in the red box contains the surface defect of each product.

Anomaly detection has been widely used in various fields, including cyber security [13,14], communications security [15–17], IoT [18–20], video surveillance [21,22], etc. In general, anomaly detection refers to finding special instances that differ from given normal instances. Conventional computer vision-based anomaly detection mainly adopts image processing methods [23,24] and machine learning methods that are based on hand-crafted feature extraction [25,26]. However, image processing methods are non-learning methods that do not utilize existing data. The performance of machine learning methods mainly depends on the quality of the hand-crafted features and few features are specially designed for anomaly detection, so it is difficult to obtain satisfactory results. Recently, deep learning-based anomaly detection methods have received extensive attention. There are many existing methods for industrial product anomaly detection [8,27,28]. Recent studies have shown that image reconstruction-based methods can be effective in addressing the problem of the lack of defective samples. However, most image reconstruction-based methods are only trained on easily accessible defect-free images. Schlegl et al. [29] proposed an anomaly detection method that was based on a vanilla generative adversarial network (GAN), which captured the manifold of normal images and reconstructed the pseudo-images that were closest to the distribution of the normal images. Soukup et al. [30] proposed an autoencoder (AE)-based network that mapped images on to latent spaces through an encoder network and created reconstructed images that were similar to the input images using a decoder network. In the test phases, the above two methods used the differences between the query image and the reconstructed image to detect surface defects.

Although both the GAN-based method and AE-based method could solve the problem of the small number of defective samples, the GAN training requires expensive computational resources and the reconstructed images from the AE differ greatly from the input images, resulting in low defect detection accuracy. To overcome these problems, this paper addresses the challenge from three perspectives. First, in order to quickly match the latent vectors that were closest to the normal image distribution, we reconstructed the defect-free images using an encoder–decoder network, thereby avoiding the process of updating the input vectors to capture the normal image manifold. Second, to further enhance the representation ability of the extracted features, a novel attention mechanism that utilizes the parallel fusion of channel attention and pixel attention was added to the encoder network

to enhance the detailed information attention of the network. Third, the latent vector that was extracted by the network only contained the features of the normal samples and it was not clear whether the features of abnormal samples were also learned. Therefore, this study proposed a new consistency loss function that was based on the pixel consistency, structural consistency and gradient consistency of the images to further improve the ability of the network to reconstruct normal samples, inhibit abnormal reconstruction and improve the accuracy of the detection of defective samples.

To summarize, the contributions of this paper are threefold:

- An encoder–decoder generative adversarial network is proposed that directly maps image spaces on to latent spaces;
- A novel dual attention block is proposed within the encoder network;
- A consistency loss function is proposed to enhance the ability of the network to reconstruct defect-free images.

This paper is organized as follows. Section 2 presents related work on surface defect detection and anomaly detection. Section 3 introduces our proposed network structure and the training strategy for our dual attention-based industrial surface defect detection method with consistency loss. Section 4 describes the dataset that we used, the training details and the experimental results. In Section 5, we draw conclusions through experiments.

2. Related Work

At present, surface defects on industrial products seriously affect the quality and efficiency of production and a number of industrial enterprises have introduced products that are related to the detection of surface defects. Cognex's deep learning defect detection tool can learn to find a variety of unacceptable product defects throughout the manufacturing process. This tool inspects the screen, band and back of a smartphone before it is packaged. It is used to detect any combination of dents, scratches and discolorations anywhere on the smartphone. Zeiss proposed SurfMax, which obtains three different modes of captured images (grayscale images, gloss images and slope images) based on deflection measurements using a high-resolution Zeiss optical sensor. It completely captures the relevant surface features and is then combined with machine learning methods to carry out surface defect detection in automotive, aerospace, medical and consumer electronics manufacturing. Creaform designed 3D scanners for the non-destructive inspection of gas pipelines and aerospace surfaces. Therefore, surface defect detection has become a research hot spot for some companies at present. In this section, we summarize the related work within surface defect detection and anomaly detection.

2.1. Surface Defect Detection

According to the extracted features and detection algorithms, the traditional surface defect detection methods within image processing can be divided into three categories: the statistical method [31], frequency spectrum method [32] and model method [33]. The traditional methods are no longer applicable due to their high human costs and their inability to represent high-dimensional data features. The rapid development of deep learning within the field of computer vision, especially the strong feature extraction ability of deep networks, has opened up new possibilities for industrial surface defect detection [34,35].

Industrial surface defect detection can improve the qualified rate and overall quality of products and it is used in a variety of tasks. Therefore, many defect detection algorithms have been proposed [36–39]. Due to the variety of defective samples and the difficulty in collecting them, most of the current surface defect detection methods are based on unsupervised or semi-supervised image reconstruction methods that rely on reconstruction errors or other measurement methods (such as latent vector errors, etc.) to detect defects. The ultimate goal of the AE-based method is to enable the encoder to learn the good low-dimensional features of a normal input image and to reconstruct the input image. Youkachen et al. [40] used a convolutional autoencoder (CAE) to reconstruct an image and complete the surface defect segmentation of a hot rolled strip. Their final

surface defect segmentation results were obtained from the reconstruction error, following a sharpening treatment. Bergmann et al. [41] believe that the l^p distance measure between pixels could lead to large residuals in the reconstruction of image edges, so they added the structural similarity (SSIM) measure to the loss function. Their results showed that the detection performance was significantly improved compared to the per-pixel reconstruction error metric.

2.2. Anomaly Detection

Anomaly detection, also known as outlier detection or novel detection [42], refers to the process in which detected data deviate significantly from normal data. In surface defect detection, defects in images can be regarded as abnormal instances, so we can apply anomaly detection to find defect images. The experiments of a large number of researchers have shown that using anomaly detection methods to detect defects is effective. Nakanishi et al. [43] considered the insufficient reconstruction accuracy of many of the AE-based methods. Natural images are mostly low frequency, so they introduced a weighted frequency domain loss (WFDL) from the perspective of the frequency domain to improve the reconstruction of high-frequency components, which made the reconstructed images clearer and improved the accuracy of the anomaly detection. Recently, many researchers have completed anomaly detection tasks using the GAN-based method [27,29,44]. The ultimate goal of the GAN-based method is to enable the generator to learn the intrinsic laws of normal samples and create reconstructed images that are similar to the normal images using the learned knowledge. In order to reduce computing resources, Akcay et al. proposed the GANomaly [28] network to reconstruct images by encoding and decoding the input image without the need to iteratively search for the latent vectors. They defined the anomaly score by encoding the latent vectors of input images and reconstructed images. Inspired by the skip connection structure of U-Net [45], Akcay et al. proposed skip-GANomaly [46], which has a stronger image reconstruction power than GANomaly. The anomaly score emphasizes the differences between the reconstructed and input images, but this method still has the problem of inaccurate detection. Tang et al. proposed a dual autoencoder GAN (DAGAN) [47], which combined the ideas of BEGAN [48] and skip-GANomaly. The generator and discriminator were composed of two autoencoders to improve the image reconstruction ability and training stability. Carrara et al. proposed CBiGAN [49], which introduced a consistency constraint regularization term into the encoder and decoder of BiGAN [50] to improve the quality and accuracy of the image reconstruction.

3. Proposed Method

In this section, we first introduce the proposed framework for the detection of industrial surface defects (as shown in Figure 2). Then, we describe in detail the dual attention module structure that we embedded into the generative network, as well as the discriminative network structure. Next, the training strategy that we employed to train our model using normal images is introduced. Finally, we define the method that we used to calculate the anomaly scores for our defect detection.

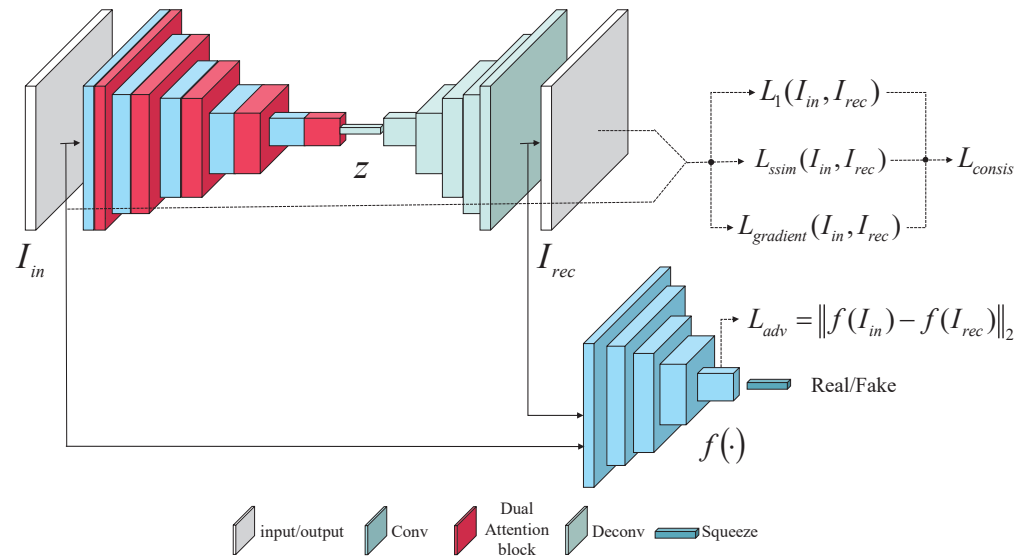


Figure 2. The network architecture of the proposed method.

3.1. Network Architecture

3.1.1. Generative Network

As shown in Figure 2, the proposed generative network was based on the autoencoder structure, which is mainly composed of an encoder G_E and a decoder G_D . The encoder network consisted of a convolutional layer, a dual attention block and a batch normalization layer. The decoder network was composed of a deconvolutional layer and a batch normalization layer. The goal of the generative network was to reconstruct the image that was closest to the defect-free input image. The input image first entered the encoder network, which acted as the feature extraction process by mapping the image on to the latent space. The encoding process could be represented as:

$$z = f_e(I_{in}) \quad (1)$$

where z represents the feature vector in the latent space, f_e represents the encoding process and I_{in} is the input image.

The latent vectors were then decoded by the network and reconstructed in the image space. The decoding procedure could be expressed as:

$$I_{rec} = f_d(z) \quad (2)$$

where I_{rec} is the reconstructed image and f_d represents the decoding process.

3.1.2. Dual Attention Block

In order to improve the quality of the network reconstruction of normal images, inspired by the methods that were proposed by Zhao et al. [51] and Dai et al. [52], we combined a pixel attention module (PAM) and a multi-scale channel attention module (MS-CAM) in the encoder network to form a dual attention block, which was connected to the convolutional layer. The dual attention block is shown in Figure 3.

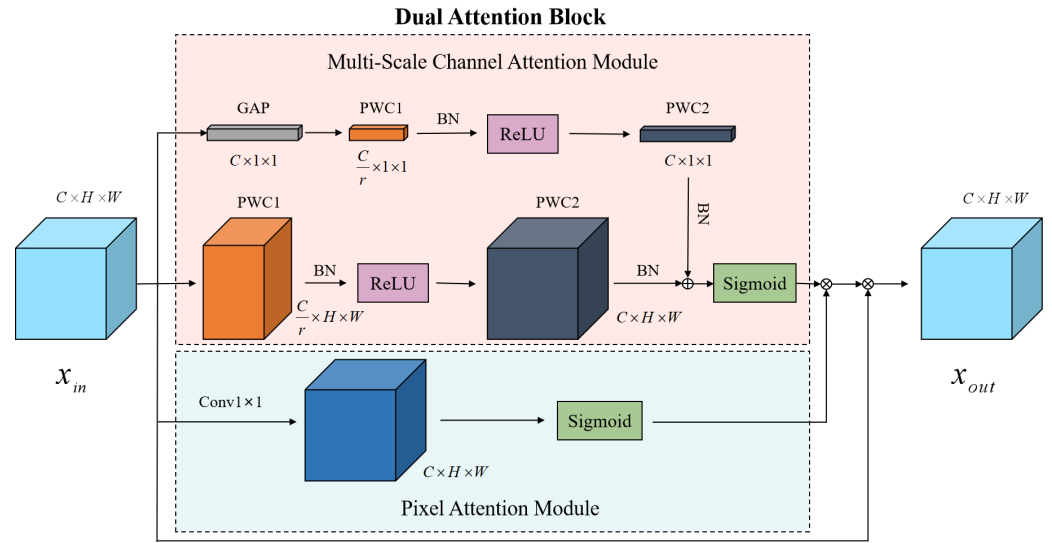


Figure 3. The dual attention block: the parallel fusion of multi-scale channel attention and pixel attention. $\oplus$ denotes the broadcasting addition and $\otimes$ denotes the element-wise multiplication.

We fused multi-scale channel attention and pixel attention in parallel. By using MS-CAM to enhance the network's attention to image channel information, varying the size of the spatial pooling allowed for channel attention at multiple scales. First of all, the channel attention of global features performed a global averaging pooling (GAP) operation on the feature map x_{in} to obtain $x_1 = \text{GAP}(x_{in})$ and then used a kernel size of $\frac{C}{r} \times C \times 1 \times 1$ for point-wise convolution (PWC1, where $r > 1$) to extract the features $x_2 = \text{PWC1}(x_1)$. After processing with the batch normalization (BN) layer and ReLU activation function, $x_3 = \text{ReLU}(\text{BN}(x_2))$ was obtained using a kernel size of $C \times \frac{C}{r} \times 1 \times 1$ for the point-wise convolution (PWC2) operation and the feature map x_4 was obtained from the BN layer $x_4 = \text{BN}(\text{PWC2}(x_3))$. The channel attention of local features also used PWC1 with a kernel size of $\frac{C}{r} \times C \times 1 \times 1$ and PWC2 with a kernel size of $C \times \frac{C}{r} \times 1 \times 1$ to extract features that were different from the channel attention for the global features. No global average pooling operation of the feature map was performed and the feature map x_5 was obtained from the channel attention of the local features. The feature map x_4 was broadcast into $C \times H \times W$ dimensions and then added pixel by pixel to the feature map x_5 to obtain a more comprehensive focus on the feature information. The Sigmoid activation function was used to obtain the attention map x_6 , $x_6 = \delta(x_4 \oplus x_5)$ (δ denotes the Sigmoid activation function and $\oplus$ denotes the broadcasting addition). Then, the pixel attention module paid more attention to the information of each pixel within the image so it could generate a 3D ($C \times H \times W$) attention feature matrix, which used a 1×1 convolution kernel to perform the convolution operations on the feature map of the previous layer and used the convolution results to obtain the attention map x_7 using the Sigmoid activation function $x_7 = \delta(\text{Conv}_1(x_{in}))$. Finally, the attention map was obtained using parallel MS-CAM and PAM and the results were multiplied pixel by pixel to create the final attention feature map x_{out} , $x_{out} = x_{in} \otimes (x_6 \otimes x_7)$ ($\otimes$ denotes the element-wise multiplication).

3.1.3. Discriminative Network

The discriminative network consisted of a convolutional layer and a batch normalization layer. The network received the real input image and the corresponding reconstructed image and then output a scalar value. After the Sigmoid function operation, the scalar value range was limited to between 0 and 1. The discriminator output a large scalar value (close to 1) for the real input image and a small scalar value (close to 0) for the reconstructed image. As the reconstructed image became more and more realistic after reaching the

Nash equilibrium [53], the reconstructed image became realistic enough to deceive the discriminator. The output of the discriminative network could be represented as:

$$D_s = f_{dis}(I_{query}) \quad (3)$$

where D_s represents the output of the discriminative network, f_{dis} represents the discriminant process and I_{query} is the query image. The query image could be either the real input image or the corresponding reconstructed image.

3.2. Training Strategy

In the training phase, the encoder performed feature extraction on the input defect-free image and mapped the image on to the latent space. The decoder then reconstructed the extracted latent feature vectors into a pseudo-image. The discriminator distinguished between the input image and the pseudo-image and output a discriminant score, which eventually caused the pseudo-image that was reconstructed by the decoder to become infinitely closer to the input image.

The training process of the entire network could be described as follows:

- First, the generative network weights and the discriminative network weights were initialized, then the generative network weights were fixed and the discriminative network weights were updated. The discriminative loss adopted the binary classification cross-entropy loss within the classical GAN;
- After the discriminative network weights were updated, the discriminative network weights were fixed and the generative network weights were updated. Adversarial loss and consistency loss were introduced when updating the generative network weights.

The adversarial loss reduced the GAN's training instability through feature matching. The L^2 distance in the middle-layer feature representations of the input and reconstructed images was employed as the loss function of the discriminator, which was expressed as follows:

$$L_{adv} = \|f(I_{in}) - f(I_{rec})\|_2 \quad (4)$$

To enhance the retention of the pixel and detailed information in the input image, the introduced consistency loss considered not only the pixels, but also the structural consistency and gradient consistency between the input image and the reconstructed image. The pixel consistency exploited the differences between the pixels in the input image and the reconstructed image to improve the image reconstruction ability. The structural consistency used SSIM to compare the real input image to the reconstructed image in terms of brightness, contrast and structure. The gradient of the image could reflect the frequency of image changes and improve the reconstruction quality of the high-frequency parts of the image. The consistency loss could be defined as:

$$L_{consis} = L_1(I_{in}, I_{rec}) + L_{ssim}(I_{in}, I_{rec}) + L_{gradient}(I_{in}, I_{rec}) \quad (5)$$

where I_{in} represents the input image, I_{rec} represents the reconstructed image and $L_1(I_{in}, I_{rec}) = \|I_{in} - I_{rec}\|_1$, $\|\cdot\|_1$ represents the L^1 norm. A larger value of SSIM indicated a higher similarity between the two images, so it could be used as $L_{ssim}(I_{in}, I_{rec}) = 1 - SSIM(I_{in}, I_{rec})$ to compute $L_{gradient}(I_{in}, I_{rec}) = \|\nabla I_{in} - \nabla I_{rec}\|_1$, where ∇ represents the gradient operations.

The adversarial loss and consistency loss were combined to update the total loss of the generative network parameters, which was indicated as:

$$L_{total} = \alpha_1 L_{adv} + \alpha_2 L_{consis} \quad (6)$$

where α_1 and α_2 indicate the weight coefficients of the adversarial loss and the consistency loss, respectively.

In the testing phase, since the network could only reconstruct defect-free images, normal images were reconstructed from unseen defect images. Therefore, the inputs and outputs of defect images were quite different, especially around the defective areas, so the anomaly score could be obtained through using the discriminative network.

3.3. Anomaly Score

Assuming that the trained generative network was good enough to reconstruct defect-free images, we used the absolute value of the pixel-by-pixel difference between the query image and the reconstructed image as the anomaly score. Given the different thresholds for the different datasets, the anomaly score was determined as an anomaly when it was greater than the relevant threshold. The anomaly score was defined as:

$$S^{(i)} = |I_{query}^{(i)} - I_{rec}^{(i)}| \quad (7)$$

where $I_{query}^{(i)}$ and $I_{rec}^{(i)}$ are the i_{th} query image and the reconstructed image, respectively, and $|\cdot|$ is the absolute value operation.

Using Equation (7), we were able to calculate the anomaly score for each query image. The anomaly scores of all of the query images formed an anomaly score vector of S , which was restricted to $[0,1]$ by feature scaling. The final anomaly score could be expressed as:

$$S_i = \frac{S^{(i)} - S_{\min}}{S_{\max} - S_{\min}} \quad (8)$$

where $S_{\max}$ and $S_{\min}$ represent the maximum and minimum values of the vector S , respectively.

4. Experiments

In this section, we evaluate the proposed method in terms of the surface defect detection problem. We first present the datasets that were used, followed by a discussion of some of the training details and evaluation metrics that were used in the experiments. Finally, we compare our method to several existing defect detection algorithms. Using the MVTec AD dataset [54], we compared the AnoGAN [29], GANomaly [28], skip-GANomaly [46], DAgAN [47] and CBiGAN [49] algorithms. Using the Magnetic Tile dataset [55], we compared the GANomaly and Adgan [27] algorithms.

4.1. Datasets

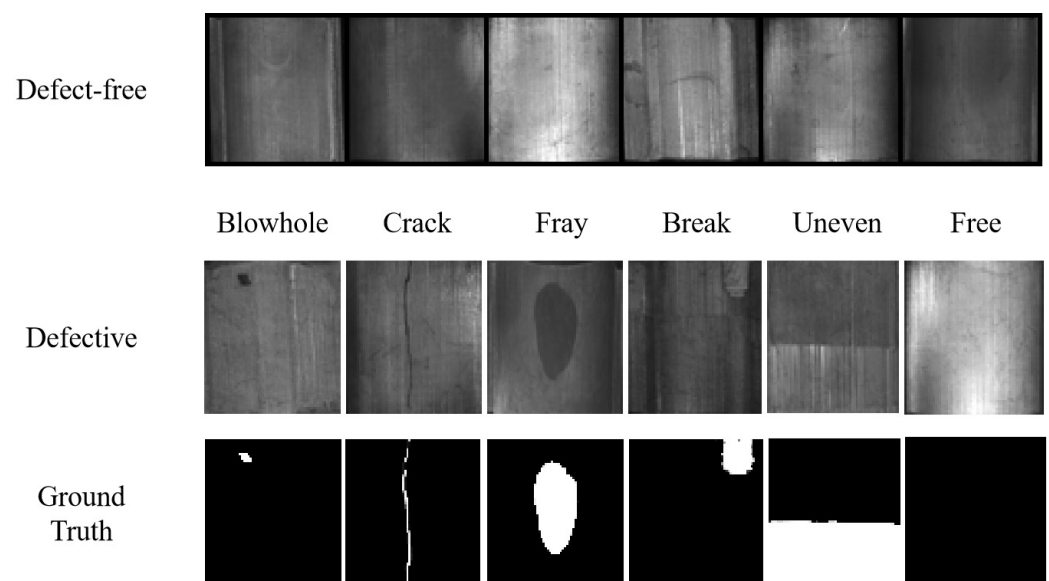
This experiment used the MVTec AD dataset [54] and the Magnetic Tile dataset [55] for the defect detection.

MVTec AD is a real-world dataset of industrial surface defects with 5354 high-resolution images. The dataset contains 15 different industrial product surfaces, each of which is divided into a training set and a testing set. The training set only contains defect-free images, while the testing set contains both defect-free images and 70 types of defect images. The details of the MVTec AD dataset are shown in Table 1.

Table 1. The MVTec AD dataset: N represents a defect-free sample and P represents a defect sample.

Category	Training Set (N)	Testing Set (N)	Testing Set (P)	Resolution
Bottle	209	20	63	900 × 900
Cable	224	58	92	1024 × 1024
Capsule	219	23	109	1000 × 1000
Carpet	280	28	89	1024 × 1024
Grid	264	21	57	1024 × 1024
Hazelnut	391	40	70	1024 × 1024
Leather	245	32	92	1024 × 1024
Metal Nut	220	22	93	700 × 700
Pill	267	26	141	800 × 800
Screw	320	41	119	1024 × 1024
Tile	230	33	84	840 × 840
Toothbrush	60	12	30	1024 × 1024
Transistor	213	60	40	1024 × 1024
Wood	247	19	60	1024 × 1024
Zipper	240	32	119	1024 × 1024

The Magnetic Tile dataset has 1344 grayscale images under multiple illumination conditions, including 952 defect-free images. We randomly selected 80% as the training set and the remaining defect-free images and 392 defect images were merged together as the testing set, which included six defect types: blowhole, crack, fray, break, uneven and free. All of the images had pixel-level labels, as shown in Figure 4.

**Figure 4.** The defect-free and defect samples in the Magnetic Tile dataset.

4.2. Training Details

To enhance the robustness of the generative network for defect image reconstruction, we used Random Erasing [56] data enhancement processing for the training set, with the Random Erasing probability set to 0.3. The data enhancement is shown in Figure 5. In addition, considering the different resolutions of the images in each dataset, we resized the input images to 256×256 . In particular, images from the MVTec AD dataset employed 3-channel images as the input, while the Magnetic Tile dataset employed single-channel images.

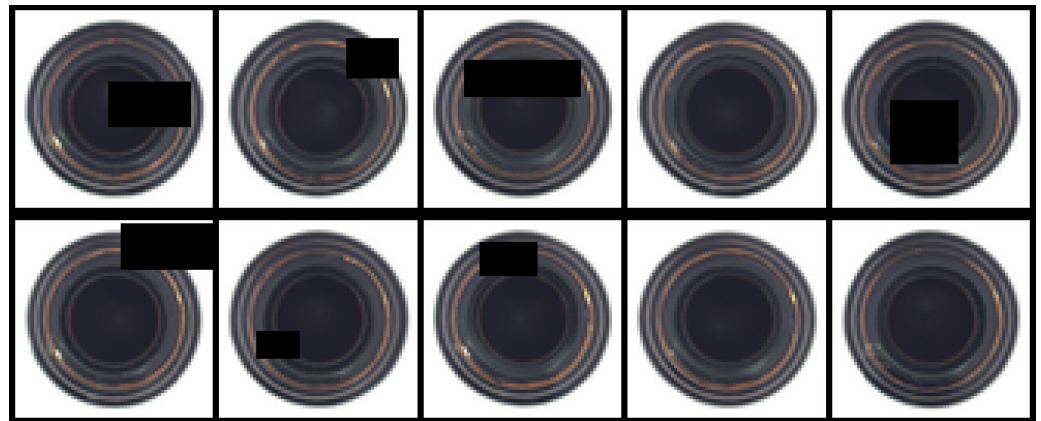


Figure 5. The Random Erasing data enhancement processing.

For all of the experiments in this paper, we employed five convolutional layers and used a dual attention block after each layer as the encoder network. The latent vectors were reconstructed after five deconvolutional layers. For the training of the generator and discriminator networks, we set the batch size to 64, used Adam [57] as the optimizer with a learning rate of 1×10^{-4} and set the momentum parameters as $\beta_1 = 0.5, \beta_2 = 0.999$. The weights for the total loss L_{total} were set to $\alpha_1 = 1$ and $\alpha_2 = 40$. All of the experiments in this paper used Pytorch 1.8.0, CUDA 11.1 and CUDNN 8.0.5. All of the experiments were performed on a computer with an Intel Core i9-10900K CPU, 64GB RAM and NVIDIA GeForce RTX 3090 GPU.

4.3. Evaluation Indicators

To evaluate the performance of the proposed method for defect detection, the AUC [58] value was utilized (the area under the curve of the receiver operating characteristics (ROC)), which had a true positive rate on the horizontal axis and a false positive rate on the vertical axis.

4.4. Experimental Results

In this subsection, we compare several popular reconstruction-based defect detection methods to verify the superiority of the proposed method for the surface defect detection problem.

First, we compared the performance of the proposed method to several other defect detection methods: AnoGAN [29], GANomaly [28], skip-GANomaly [46], DAGAN [47] and CBiGAN [49]. Among them, AnoGAN generates pseudo-images that are similar to the probability distribution of the normal samples using random noise and the anomaly score consists of the difference between the pixel space of the input image and that of the generated image and the difference between the feature maps of the last layer of the discriminator network. GANomaly reconstructs images using an encoder–decoder–encoder process and defines the anomaly score by encoding the differences between the input image and the generated image to obtain a latent vector. Inspired by the skip connection structure, skip-GANomaly improves the structure of GANomaly to obtain a stronger image reconstruction ability and the anomaly score emphasizes the differences between the reconstructed image and the input image. The generator and discriminator networks of DAGAN are composed of two autoencoders, which improves the network’s ability to reconstruct images and its training stability. CBiGAN introduces a consistency-constrained regularization term within the encoder and decoder, resulting in an improved reconstruction accuracy. The comparison results are shown in Table 2. From the results, it can be observed that our method achieved the best performance using the MVtec AD dataset. Although skip-GANomaly and DAGAN demonstrated a strong reconstruction ability, they do not have attention mechanisms added into their networks, which resulted in

a lack of attention to detail in the images. Our model showed a more comprehensive feature extraction ability due to the addition of the dual attention block, which provided a further supplement to the detailed features. As can be seen from Table 2, the defect detection performance of our method was greatly improved compared to the other methods. Our results for the cable and pill categories were 7% and 8% higher than CBiGAN, respectively. For the capsule category, our result was 13% higher than GANomaly. For the carpet, leather, toothbrush and zipper categories, our results were 1%, 1%, 5% and 13% higher than DAGAN, respectively. For the transistor category, our result was 7% higher than skip-GANomaly. From the mean experimental results of the 15 categories, our method outperformed the existing methods by 3.3% and achieved the best results.

Table 2. The area under the receiver operating characteristic curve of the MVTec AD dataset. The results in bold were the best AUC results during the tests and the underlined results were the suboptimal AUC results.

Category	AnoGAN	GANomaly	Skip-GANomaly	DAGAN	CBiGAN	Ours
Bottle	0.80	0.79	0.93	0.98	0.87	<u>0.94</u>
Cable	0.47	0.71	0.67	0.66	<u>0.81</u>	0.88
Capsule	0.44	<u>0.72</u>	0.71	0.68	0.56	0.85
Carpet	0.33	0.82	0.79	<u>0.90</u>	0.55	0.91
Grid	0.87	0.74	0.65	<u>0.86</u>	0.99	<u>0.94</u>
Hazelnut	0.25	0.87	0.90	1.00	0.77	<u>0.95</u>
Leather	0.45	0.80	0.90	<u>0.94</u>	0.83	0.95
Metal Nut	0.28	0.69	<u>0.79</u>	0.81	0.63	0.69
Pill	0.71	0.67	0.75	0.76	<u>0.81</u>	0.89
Screw	0.10	1.00	1.00	1.00	<u>0.58</u>	1.00
Tile	0.40	0.72	0.85	0.96	<u>0.91</u>	0.80
Toothbrush	0.43	0.70	0.68	<u>0.95</u>	0.94	1.00
Transistor	0.69	0.80	<u>0.81</u>	0.79	0.77	0.88
Wood	0.56	0.92	0.91	0.97	<u>0.95</u>	0.94
Zipper	0.71	0.74	0.66	<u>0.78</u>	0.53	0.91
Mean	0.499	0.779	0.800	<u>0.869</u>	0.766	0.902

To highlight the superiority of the proposed method, we plotted an AUC line chart for each category of the MVTec AD dataset (as shown in Figure 6). It can be seen more intuitively from the figure that our proposed method showed a more robust performance for industrial defect detection than the other GAN-based methods. On the other hand, the line chart fluctuations for AnoGAN were large because it needed to iteratively search for the appropriate latent vectors, resulting in unstable training.

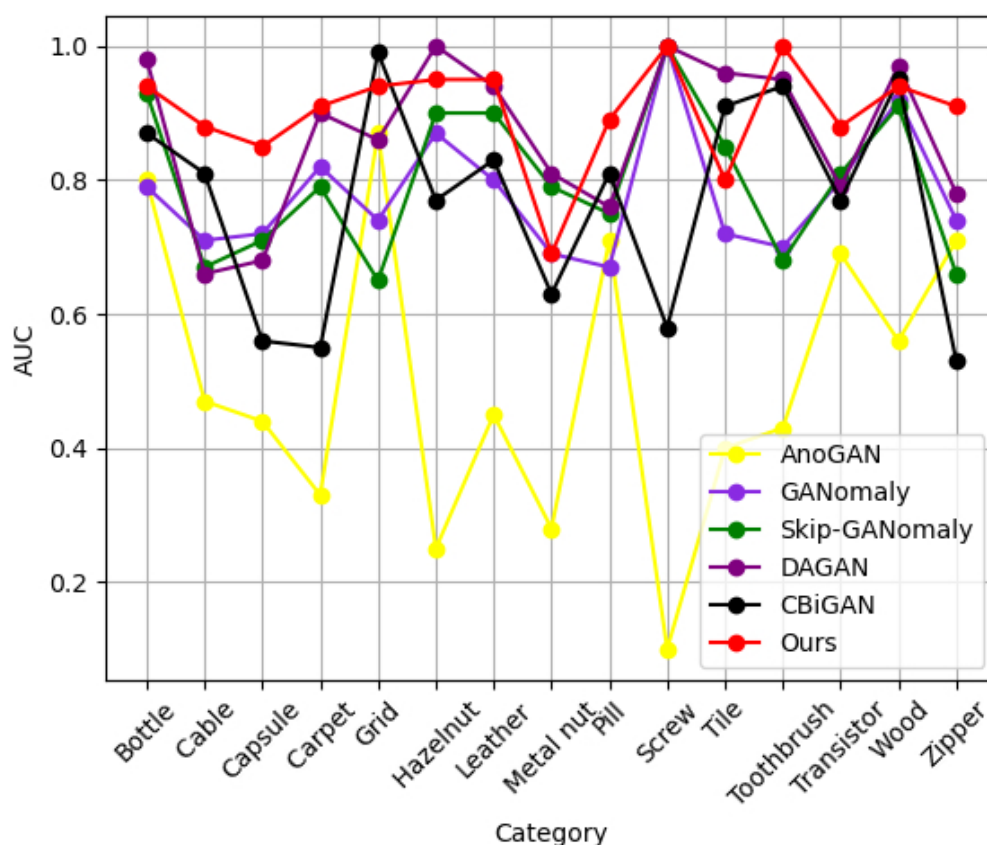


Figure 6. The AUC of our proposed method and five other GAN-based methods, which were tested using the MVTec AD dataset.

Then, the proposed method was further verified using the Magnetic Tile dataset. Compared to GANomaly and Adgan [27] (Adgan uses a scalable encoder–decoder–encoder architecture), fine-grained reconstructed images of normal classes could be obtained by extracting and exploiting the multi-scale features of normal samples. The comparison results are shown in Table 3. The images from the Magnetic Tiles dataset are under different illumination conditions that have a great impact on defect detection, so the two methods showed poor defect detection performances. Our proposed consistency loss enhanced the sensitivity of our model under different illumination conditions and improved the defect detection ability. From the experimental results, it can be observed that our model could also be trained stably using grayscale images under varying illumination conditions, with an 8% and 38% improvement over GANomaly and Adgan, respectively. GANomaly and Adgan have similar encoder–decoder–encoder structures and have good reconstruction abilities under simple conditions, but for complex scenes, their defect detection performances were poor. We comprehensively considered the characteristics of image pixels, structure and gradient so that our model could maintain a good reconstruction ability for complex scenes and obtain an excellent defect detection ability.

Table 3. The area under the receiver operating characteristic curve of the Magnetic Tile dataset. Bold number represents the optimal result.

Method	GANomaly	Adgan	Ours
AUC	0.76	0.46	0.84

The results of the sample study using the MVTec AD and Magnetic Tile datasets are shown in Figure 7. This figure shows that our proposed method could reconstruct

defect images as defect-free images. Using the residual map of the defect image and the reconstructed image, the defect could be found easily. By comparing the heat maps and ground truths, it can be observed that our model could accurately detect the location of defects.

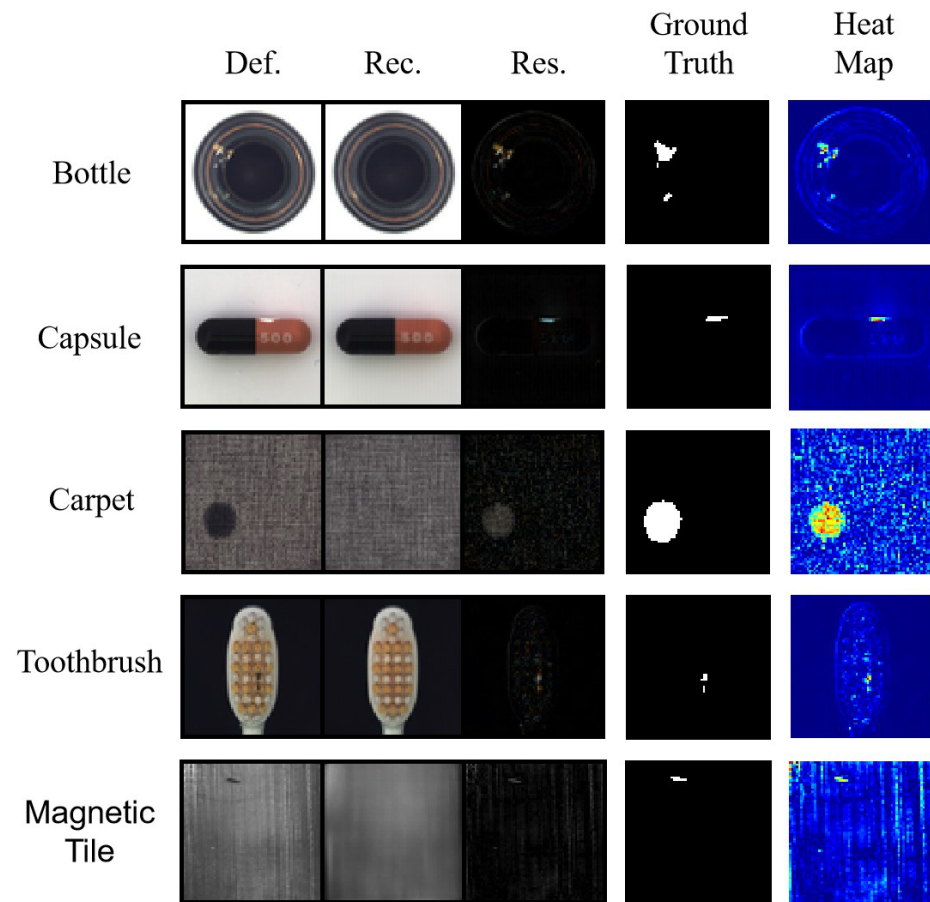


Figure 7. The first four rows of images show the test results from the partial MVTec AD dataset and the fifth row shows the test results from the Magnetic Tile dataset. Def. represents the defect image, Rec. represents the reconstructed image and Res. represents the residual image.

4.5. Ablation Studies

In this subsection, we present the results from our ablation studies, in which we performed a group of experiments to verify the effectiveness of the individual strategies within our proposed model, mainly from two perspectives: the effectiveness of the dual attention block and the effectiveness of the consistency loss.

4.5.1. Effectiveness of the Dual Attention Block

We fixed our proposed consistency loss as the loss function of the generative network by changing the attention in the encoder network and constructing four different structures. First, the dual attention block was removed from the encoder network to evaluate the defect detection performance without the attention mechanism, which was named **Struc1**. Multi-scale channel attention was introduced into the encoder network to evaluate the effects of channel attention on the defect detection performance, which was named **Struc2**. Then, the channel attention in the encoder network was replaced with pixel attention to evaluate the impacts of pixel attention on the defect detection, which was named **Struc3**. Finally, our proposed method was named **Struc4**. As can be seen from Tables 4 and 5, the different structures achieved different results by adjusting the attention mechanism, although **Struc2** and **Struc3** had higher AUC values for some categories in the MVTec AD dataset, the encoder network for the parallel fusion of multi-scale channel attention and

pixel attention could effectively extract the key information so **Struc4** achieved a mean AUC value of 90.2%. Using the Magnetic Tile dataset, we counted the running speed that was needed to train and test each image for the different structures. Although **Struc4** took a little longer than the other structures, it achieved the highest result of 84%. As shown in Figure 8, we used the heat map of the metal nut category to test the ability of the four structures to detect defects in the same image. By comparison, **Struc4** had less noise in the defect heat map and could detect defects more accurately.

Table 4. The test results from the different attention structures using the MVTec AD dataset. Bold number represents the optimal result.

Category	Struc1	Struc2	Struc3	Struc4
Bottle	0.95	0.94	0.96	0.95
Cable	0.90	0.87	0.87	0.88
Capsule	0.79	0.85	0.85	0.85
Carpet	0.83	0.91	0.88	0.91
Grid	0.87	0.87	0.92	0.94
Hazelnut	0.94	0.97	0.92	0.95
Leather	0.89	0.89	0.95	0.95
Metal Nut	0.63	0.64	0.62	0.69
Pill	0.86	0.87	0.89	0.89
Screw	1.00	1.00	1.00	1.00
Tile	0.71	0.70	0.71	0.80
Toothbrush	1.00	1.00	0.99	1.00
Transistor	0.86	0.87	0.86	0.88
Wood	0.94	0.93	0.94	0.94
Zipper	0.91	0.88	0.89	0.91
Mean	0.872	0.879	0.884	0.902

Table 5. The AUC results from the different attention structures, which were tested using the Magnetic Tile dataset, and the running speeds (in seconds), which were measured for each image during training and testing. Bold number represents the optimal result.

Method	Struc1	Struc2	Struc3	Struc4
AUC	0.75	0.82	0.79	0.84
Training Speed (s)	0.0931	0.1550	0.1026	0.1557
Testing Speed (s)	0.0276	0.0487	0.0413	0.0511

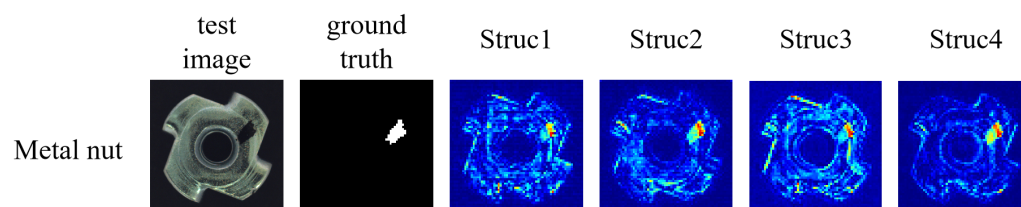


Figure 8. The heat maps of an image of a defective metal nut using the different attention structures.

4.5.2. Effectiveness of the Consistency Loss

Without changing the dual attention network structure, we considered the loss function in the network in three ways. First, the loss function in the generative network used the pixel consistency loss function L_1 to evaluate its impact on the image reconstruction ability. Second, based on the pixel consistency loss function, the structural consistency loss function was added and the two functions were used as the generative network loss function to evaluate the generation effects, namely $L_1 + L_{ssim}$. Third, we combined pixel consistency, structural consistency and gradient consistency to constitute the consistency loss function, namely $L_1 + L_{ssim} + L_{gradient}$. The experimental results are shown in Tables 6

and 7. The results show that the proposed consistency loss achieved the best performance using the two datasets.

Table 6. The AUC values from the different loss functions using the MVTec AD dataset. Bold number represents the optimal result.

Category	L_1	$L_1 + L_{ssim}$	$L_1 + L_{ssim} + L_{gradient}$
Bottle	0.95	0.94	0.94
Cable	0.82	0.86	0.88
Capsule	0.85	0.85	0.85
Carpet	0.88	0.92	0.91
Grid	0.87	0.86	0.94
Hazelnut	0.96	0.94	0.95
Leather	0.91	0.88	0.95
Metal Nut	0.62	0.62	0.69
Pill	0.86	0.90	0.89
Screw	1.00	1.00	1.00
Tile	0.73	0.72	0.80
Toothbrush	1.00	1.00	1.00
Transistor	0.87	0.87	0.88
Wood	0.94	0.93	0.94
Zipper	0.89	0.90	0.91
Mean	0.876	0.879	0.902

Table 7. The AUC values from the different loss functions using the Magnetic Tile dataset. Bold number represents the optimal result.

Method	L_1	$L_1 + L_{ssim}$	$L_1 + L_{ssim} + L_{gradient}$
AUC	0.82	0.83	0.84

5. Conclusions

We studied the problem of detecting surface defects during the production process of industrial products, i.e., the wide variety of surface defects among different products and the difficulty in collecting defective samples. Therefore, we proposed a semi-supervised anomaly detection method that was based on dual attention and consistency loss to accomplish this task. We used an encoder–decoder structure for the generative network and introduced a dual attention module into the encoder network, which combined multi-scale channel attention and pixel attention. The parallel fusion of the two kinds of attention mechanism improved the performance of key feature extraction and reconstructed higher quality defect-free images. In addition, the consistency loss made use of the differences between the pixels, structures and gradients of the defect images and defect-free images to further improve the performance of the defect detection. Comprehensive experiments using the MVTec AD and Magnetic Tile datasets showed that the proposed method could achieve a superior performance over the existing methods.

Author Contributions: All authors participated in some part of the work for this article. Investigation, X.L.; methodology, Y.Z.; software, X.L.; supervision, E.Z.; writing—original draft preparation, X.L. and Y.Z.; writing—review and editing, Y.Z., B.C. and E.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (grant number: 61906143), in part by the China Post-Doctoral Science Foundation (grant number: 2021T140529 and 2018M643584), in part by the Fundamental Research Fund for the Central Universities (grant number: XJS211504) and in part by the Natural Science Basic Research Plan in Shaanxi Province of China (grant number: 2020JQ-305).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

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

References

1. Kang, G.; Gao, S.; Yu, L.; Zhang, D. Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 2679–2690. [\[CrossRef\]](#)
2. Yu, H.; Li, Q.; Tan, Y.; Gan, J.; Wang, J.; Geng, Y.a.; Jia, L. A coarse-to-fine model for rail surface defect detection. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 656–666. [\[CrossRef\]](#)
3. Chen, X.; Lv, J.; Fang, Y.; Du, S. Online Detection of Surface Defects Based on Improved YOLOV3. *Sensors* **2022**, *22*, 817. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wang, C.; Xiao, Z. Lychee surface defect detection based on deep convolutional neural networks with gan-based data augmentation. *Agronomy* **2021**, *11*, 1500. [\[CrossRef\]](#)
5. Huang, T.; Zheng, B.; Zhang, J.; Yi, C.; Jiang, Y.; Shui, Q.; Jian, H. Mango Surface Defect Detection Based on HALCON. In Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference, Chongqing, China, 12–14 March 2021; pp. 2627–2631.
6. Ding, F.; Yang, G.; Ding, D.; Cheng, G. Retinal Nerve Fiber Layer Defect Detection with Position Guidance. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; pp. 745–754.
7. Panda, R.; Puhan, N.B.; Rao, A.; Mandal, B.; Padhy, D.; Panda, G. Deep convolutional neural network-based patch classification for retinal nerve fiber layer defect detection in early glaucoma. *J. Med. Imaging* **2018**, *5*, 044003. [\[CrossRef\]](#)
8. Rudolph, M.; Wandt, B.; Rosenhahn, B. Same same but different: Semi-supervised defect detection with normalizing flows. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2021; pp. 1907–1916.
9. Su, B.; Chen, H.; Zhou, Z. BAF-Detector: An Efficient CNN-Based Detector for Photovoltaic Cell Defect Detection. *IEEE Trans. Ind. Electron.* **2021**, *69*, 3161–3171. [\[CrossRef\]](#)
10. Xu, X.; Chen, J.; Zhang, H.; Ng, W.W. D4Net: De-deformation defect detection network for non-rigid products with large patterns. *Inf. Sci.* **2021**, *547*, 763–776. [\[CrossRef\]](#)
11. Alvarenga, T.A.; Carvalho, A.L.; Honorio, L.M.; Cerqueira, A.S.; Filho, L.M.; Nobrega, R.A. Detection and classification system for rail surface defects based on Eddy current. *Sensors* **2021**, *21*, 7937. [\[CrossRef\]](#)
12. Ho, C.C.; Chou, W.C.; Su, E. Deep Convolutional Neural Network Optimization for Defect Detection in Fabric Inspection. *Sensors* **2021**, *21*, 7074. [\[CrossRef\]](#)
13. Bakhshi, T.; Ghita, B. Anomaly Detection in Encrypted Internet Traffic Using Hybrid Deep Learning. *Secur. Commun. Netw.* **2021**, *2021*, 5363750. [\[CrossRef\]](#)
14. Qu, Y.; Uddin, M.P.; Gan, C.; Xiang, Y.; Gao, L.; Yearwood, J. Blockchain-Enabled Federated Learning: A Survey. In Proceedings of the 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI), Beijing, China, 15 July–15 August 2022.
15. Feng, B.; Zhou, H.; Li, G.; Zhang, Y.; Sood, K.; Yu, S. Enabling Machine Learning with Service Function Chaining for Security Enhancement at 5G Edges. *IEEE Netw.* **2021**, *35*, 196–201. [\[CrossRef\]](#)
16. Li, G.; Zhou, H.; Feng, B.; Zhang, Y.; Yu, S. Efficient provision of service function chains in overlay networks using reinforcement learning. *IEEE Trans. Cloud Comput.* **2019**, *10*, 383–395. [\[CrossRef\]](#)
17. Feng, B.; Li, G.; Li, G.; Zhang, Y.; Zhou, H.; Yu, S. Enabling efficient service function chains at terrestrial-satellite hybrid cloud networks. *IEEE Netw.* **2019**, *33*, 94–99. [\[CrossRef\]](#)
18. Vu, L.; Nguyen, Q.U.; Nguyen, D.N.; Hoang, D.T.; Dutkiewicz, E. Learning latent representation for iot anomaly detection. *IEEE Trans. Cybern.* **2020**, *52*, 3769–3782. [\[CrossRef\]](#)
19. Bhatia, R.; Benno, S.; Esteban, J.; Lakshman, T.; Grogan, J. Unsupervised machine learning for network-centric anomaly detection in IoT. In Proceedings of the 3rd Acm Conext Workshop on Big Data, Machine Learning and Artificial Intelligence for Data Communication Networks, Orlando, FL, USA, 9 December 2019; pp. 42–48.
20. Feng, B.; Tian, A.; Yu, S.; Li, J.; Zhou, H.; Zhang, H. Efficient Cache Consistency Management for Transient IoT Data in Content-Centric Networking. *IEEE Internet Things J.* **2022**. [\[CrossRef\]](#)
21. Liu, W.; Luo, W.; Lian, D.; Gao, S. Future frame prediction for anomaly detection—a new baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6536–6545.
22. Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6479–6488.

23. Chalapathy, R.; Menon, A.K.; Chawla, S. Anomaly detection using one-class neural networks. *arXiv* **2018**, arXiv:1802.06360.
24. Yang, C.; Liu, P.; Yin, G.; Jiang, H.; Li, X. Defect detection in magnetic tile images based on stationary wavelet transform. *Ndt E Int.* **2016**, *83*, 78–87. [[CrossRef](#)]
25. Ahn, E.; Kumar, A.; Feng, D.; Fulham, M.; Kim, J. Unsupervised deep transfer feature learning for medical image classification. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8–11 April 2019; pp. 1915–1918.
26. Li, H.; Boulanger, P. Structural Anomalies Detection from Electrocardiogram (ECG) with Spectrogram and Handcrafted Features. *Sensors* **2022**, *22*, 2467. [[CrossRef](#)]
27. Cheng, H.; Liu, H.; Gao, F.; Chen, Z. Adgan: A scalable gan-based architecture for image anomaly detection. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, Chongqing, China, 12–14 June 2020; pp. 987–993.
28. Akcay, S.; Atapour-Abarghouei, A.; Breckon, T.P. Ganomaly: Semi-supervised anomaly detection via adversarial training. In Proceedings of the Asian Conference on Computer Vision, Perth, WA, Australia, 2–6 December 2018; pp. 622–637.
29. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In Proceedings of the International Conference on Information Processing in Medical Imaging, Boone, NC, USA, 25–30 June 2017; pp. 146–157.
30. Soukup, D.; Pinetz, T. Reliably Decoding Autoencoders' Latent Spaces for One-Class Learning Image Inspection Scenarios. In Proceedings of the OAGM Workshop, Hall/Tyrol, Austria, 15–16 May 2018.
31. Zhang, H.; Jin, X.; Wu, Q.J.; Wang, Y.; He, Z.; Yang, Y. Automatic visual detection system of railway surface defects with curvature filter and improved Gaussian mixture model. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 1593–1608. [[CrossRef](#)]
32. Li, J.; Levine, M.D.; An, X.; Xu, X.; He, H. Visual saliency based on scale-space analysis in the frequency domain. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 996–1010. [[CrossRef](#)]
33. Chao, S.M.; Tsai, D.M. Anisotropic diffusion with generalized diffusion coefficient function for defect detection in low-contrast surface images. *Pattern Recognit.* **2010**, *43*, 1917–1931. [[CrossRef](#)]
34. Zenati, H.; Romain, M.; Foo, C.S.; Lecouat, B.; Chandrasekhar, V. Adversarially learned anomaly detection. In Proceedings of the 2018 IEEE International Conference on Data Mining, Singapore, 17–20 November 2018; pp. 727–736.
35. Wang, J.; Yi, G.; Zhang, S.; Wang, Y. An Unsupervised Generative Adversarial Network-Based Method for Defect Inspection of Texture Surfaces. *Appl. Sci.* **2021**, *11*, 283. [[CrossRef](#)]
36. Choi, J.; Kim, C. Unsupervised detection of surface defects: A two-step approach. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 1037–1040.
37. Kim, M.S.; Park, T.; Park, P. Classification of steel surface defect using convolutional neural network with few images. In Proceedings of the 2019 12th Asian Control Conference, Kitakyushu, Japan, 9–12 June 2019; pp. 1398–1401.
38. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 4–19 June 2020; pp. 4183–4192.
39. Wang, Y.; Zhang, Y.; Zheng, L.; Yin, L.; Chen, J.; Lu, J. Unsupervised Learning with Generative Adversarial Network for Automatic Tire Defect Detection from X-ray Images. *Sensors* **2021**, *21*, 6773. [[CrossRef](#)] [[PubMed](#)]
40. Youkachan, S.; Ruchanurucks, M.; Phatrapomnant, T.; Kaneko, H. Defect segmentation of hot-rolled steel strip surface by using convolutional auto-encoder and conventional image processing. In Proceedings of the 2019 10th International Conference of Information and Communication Technology for Embedded Systems, Bangkok, Thailand, 25–27 March 2019; pp. 1–5.
41. Bergmann, P.; Löwe, S.; Fauser, M.; Sattlegger, D.; Steger, C. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv* **2018**, arXiv:1807.02011.
42. Pang, G.; Shen, C.; Cao, L.; Hengel, A.V.D. Deep learning for anomaly detection: A review. *ACM Comput. Surv.* **2021**, *54*, 1–38. [[CrossRef](#)]
43. Nakanishi, M.; Sato, K.; Terada, H. Anomaly Detection By Autoencoder Based On Weighted Frequency Domain Loss. *arXiv* **2021**, arXiv:2105.10214.
44. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Langs, G.; Schmidt-Erfurth, U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal.* **2019**, *54*, 30–44. [[CrossRef](#)]
45. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
46. Akçay, S.; Atapour-Abarghouei, A.; Breckon, T.P. Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In Proceedings of the 2019 International Joint Conference on Neural Networks, Budapest, Hungary, 14–19 July 2019; pp. 1–8.
47. Tang, T.W.; Kuo, W.H.; Lan, J.H.; Ding, C.F.; Hsu, H.; Young, H.T. Anomaly detection neural network with dual auto-encoders GAN and its industrial inspection applications. *Sensors* **2020**, *20*, 3336. [[CrossRef](#)]
48. Berthelot, D.; Schumm, T.; Metz, L. Began: Boundary equilibrium generative adversarial networks. *arXiv* **2017**, arXiv:1703.10717.
49. Carrara, F.; Amato, G.; Brombin, L.; Falchi, F.; Gennaro, C. Combining gans and autoencoders for efficient anomaly detection. In Proceedings of the 2020 25th International Conference on Pattern Recognition, Milan, Italy, 10–15 January 2021; pp. 3939–3946.
50. Donahue, J.; Krähenbühl, P.; Darrell, T. Adversarial feature learning. *arXiv* **2016**, arXiv:1605.09782.

51. Zhao, H.; Kong, X.; He, J.; Qiao, Y.; Dong, C. Efficient image super-resolution using pixel attention. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 56–72.
52. Dai, Y.; Gieseke, F.; Oehmcke, S.; Wu, Y.; Barnard, K. Attentional feature fusion. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA 3–8 January 2021; pp. 3560–3569.
53. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13 December 2014.
54. Bergmann, P.; Batzner, K.; Fauser, M.; Sattlegger, D.; Steger, C. The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *Int. J. Comput. Vis.* **2021**, *129*, 1038–1059. [[CrossRef](#)]
55. Huang, Y.; Qiu, C.; Yuan, K. Surface defect saliency of magnetic tile. *Vis. Comput.* **2020**, *36*, 85–96. [[CrossRef](#)]
56. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random erasing data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 13001–13008.
57. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
58. Ling, C.X.; Huang, J.; Zhang, H. AUC: A statistically consistent and more discriminating measure than accuracy. In Proceedings of the International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 9–15 August 2003; pp. 519–524.