A Denoising Method for Multi-Noise on Steel Surface Detection

Zhiwu Chen, Wenjing Wang, QingE Wu *, Yingbo Lu, Lintao Zhou and Hu Chen

School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China
* Correspondence: wqe969699@163.com

Abstract: In order to solve the problem that steel surface defects are easily covered or submerged by other objects or noise, this paper proposes an open–closed transformation algorithm which can eliminate or weaken multiple noises. In the case of a small number of samples, this paper establishes a super-resolution generative adversarial neural network to achieve the enhancement of sample data. For avoiding unrealistic image defects caused by cuts or brightness variations, an enhancement method is given which incorporates the original defective high-frequency information into classical image fusion methods, such as rotation and error slicing. Experimental results show that the accuracy of the proposed denoising method reaches over 90%, which is more than 2.6% of that of the most primitive classification network. To compare with existing denoising methods, the denoising method proposed in this paper not only has higher accuracy, faster denoising speed, and stronger anti-interference ability, but also has better adaptation to the environment. This research will provide a new solution method for the denoising of multi-noise phenomena in multiple different environments.

Keywords: denoising method; open–closed transformation; generative adversarial networks; data augmentation

1. Introduction

With the rapid development of the country’s industrial information technology, the use of steel has spread to all walks of life; from civilian household products to military special products, all need a great deal of steel, and in the field of machinery manufacturing, aerospace, and other areas, the quality of steel products needs ever higher requirements. According to the statistics of relevant researchers in recent years regarding large steel enterprises orders, the unsatisfactory performance of the quality of steel is mostly related to surface defects. Common steel defects include delamination, scarring, pulling, cracking, hairline, bubbles, impurities, pitting, etc. These surface defects will seriously affect the use of steel, but also reduce the corrosion resistance, oxidation resistance, wear resistance, and strength of steel products. Therefore, the accurate and rapid detection of surface defects on steel products is not only very important for large rolling mills and even the steel industry market, but also a key task for metallurgical enterprises. The presence of a variety of noises on the steel surface is an important obstacle to the accurate detection of defects. This paper deals with the multiple noises existing on the steel surface in order to lay a good foundation for the detection and classification of steel defects.

Along with the rapid progress of CCD (Charge-Coupled Device) technology [1], line array cameras have been used in the field of steel surface defects in large rolling mills [2,3]. If the production line is arranged with line array cameras, the collected image data can be monitored in real-time, and the quality control workers can observe the surface defects of steel products through the monitor. Bayati Farzaneh et al. [4–6] proposed the use of ultrasonic detectors for denoising methodologies, and Hou et al. [7–10] proposed the use of infrared technology for the tracking and detection of small targets. Shang et al. [11,12] also used bright-field and dark-field illumination techniques to monitor surface defects on steel plates. Mou Shuxian [13] used LED light sources to design a complete system for
detecting surface defects on rolled steel. Today, more large steel rolling enterprises still use this method to monitor the quality of steel surface defects, and although this method has improved the working environment of quality control workers to a certain extent, its drawbacks are also significant, with still very costly human resources, a method of detection whose results are far from robust, and damage to economic benefits for enterprises. With the progress of image processing technology and the increase in labor costs, the use of machine vision algorithms to replace manual vision has become a trend [14–17]. On this basis, machine vision-based steel surface defect detection technology is becoming increasingly mature, and larger steel rolling enterprises have thus reduced the defect rate and gradually improved their business efficiency.

With the ability to improve production efficiency while also largely relieving manual fatigue, powerful computer arithmetic and efficient vision processing algorithms can make the detection of surface defects more efficient and accurate. Traditional machine learning algorithms are generally composed of structures such as image processing, image noise reduction, feature extraction, classification, and detection, and each of these structures play a significant role in the final result. Zhang Ju et al. [18] propose a new method for image denoising based on a motion decomposition framework, but there are limited usage scenarios. A localized discharge signal denoising method based on improved variational modal decomposition was proposed by Yang Jingjie et al. [19], but the accuracy of the method could be improved. Xu Shaoping et al. [20] proposed a three-stage denoising method based on a priori deep image using generation and fusion strategies, but there is no wide application. Li Caoyuan et al. [21–24] used an image denoising technique based on nonlocal Bayesian singular value thresholding and the Stein unbiased risk estimator. Fu Xuenian et al. [25–28] proposed a fusion denoising method for laser blood flow speckle imaging based on homomorphic transform and three-dimensional transform domain synergistic filtering, which is widely used, but only in medical species. Sohlberg Antt et al. [29,30] apply deep learning and generative networks to process noise, but real-time performance needs to be improved. The detection algorithm is not robust and real-time detection is difficult to realize. Cai Jianxian et al. [31–33] proposed a method using convolutional neural networks to remove noise. However, this method requires a large amount of data and has some limitations.

The traditional method can only illuminate part of the surface of the steel plate, and the quality inspection personnel cannot see the whole steel plate, which makes it not only easy to cause leakage detection, but quality inspection personnel, for some time and due to the high frequency and high intensity flash irradiation, will be affected by physical and mental health issues to varying degrees, so there is a certain degree of safety hazard. Most of the methods based on deep learning are artificially designed for a specific set of functions for specific defects, are not universal, and not accurate enough to locate the defects. In most cases, this can only locate defects in the rough position, the detection method is inefficient, real-time is poor, and the method is mixed with personal subjective factors and lack of scientific standardization; therefore, there cannot be a production process for multiple surfaces of steel at the same time in order to carry out real-time detection. The research in this paper studies a denoising method for multi-noise in steel surface detection. These noises may include Gaussian noise, pretzel noise, streak noise, color noise, artifacts, blurring and distortion, pressure noise, and other interfering information, which usually needs to be processed and removed in order to accurately analyze the steel surface image to obtain clearer and more accurate image data.

2. Related Works

In this paper, the main focus is on open datasets and images collected from steel mills. The surface defect dataset constructed by Professor Kechen Song’s team in the School of Mechanical Engineering and Automation at Northeastern University was highly evaluated by Professor Elizabeth A. Holm (ASM Fellow and TMS Fellow, President of the Minerals, Metals, and Materials Society, and member of the Materials Advisory Board) of Carnegie
Mellon University. On a corporate level, the Steel Defect Detection Challenge was held using this dataset as the base data. The open data set consists of 1800 images, of which the size of a single defect image is $200 \times 200$ and the size is about 20 KB, including six types of defects, such as scratches, scars, impurities, pitting, etc. The open data collected in the steel plant are 18CrNiMo7-6+HH and 42CrMo, including two types of gear steel and alloy steel, central sparseness and a white bright band, and two kinds of defects. The open data collected from the steel mill are $5742 \times 3648$, size 1–2 M. In more steel production enterprises in China, the above mentioned common defects are mainly observed manually through the naked eye, if necessary, with the help of a magnifying glass or microscope at each location of steel for detailed inspection, so as to ensure the reliability of surface defects detection. Finally, according to the standards set by the industry, the detected surface defects are judged to be within a reasonable range, and if they are out of this range, they are defined as scrap and need to be reprocessed. The open data set and the actual measured data from steel companies are shown in Figures 1 and 2 below:

Figure 1. Open data set.
The research in this paper helps steel sorting and inspection processors to understand the laws of inspection targets. In addition, there is the influence of other interfering objects to give appropriate decisions. It is essential to find ways to suppress or eliminate the influence of undesirable factors on them all. It can save resources and costs, reduce unnecessary losses or injuries, etc. for the business units using the study, as well as providing an experimental basis for steel inspection. From the above, it can be seen that it is very important to use a steel surface defect detection method with a higher detection efficiency and accuracy in large steel rolling enterprises, especially in the production processing lines. Based on this, this paper researches the problem of multi-noise denoising on steel surfaces by combining computer vision and deep learning, and further refines and verifies the effectiveness of the proposed algorithm through experiments. Figure 3 illustrates the overall structure of this paper.

The rest of this paper will be discussed in the following sections. Section 3 proposes an open and closed transform algorithm to eliminate or attenuate multiple noises. Section 4 establishes a super-resolution generative adversarial neural network for sample image enhancement. Based on this, a data enhancement method is given to incorporate raw defective high frequency information into classical image enhancement methods such as
rotation and error slicing. Section 5 presents the experimental results and analysis and Section 6 is the conclusion.

Highlights:
The novelty of present work had been abstracted as follows:
- an open-closed transformation algorithm is proposed;
- a super-resolution generative adversarial neural network is established;
- a data enhancement method is given.

3. Open-Closed Transformation Denoising Method for Defects of Steel Surface

For the situation that small targets in complex backgrounds are easily obscured or submerged by other objects or noise, this paper proposes an open-closed transformation algorithm to eliminate or weaken the background and noise. A low signal-to-noise ratio and the measurement and acquisition of small target information are currently the main technical problems to be overcome by various advanced measurement systems. The target image occupies fewer pixels in the region and the background environment is so complex that the target is almost background covered by clutter and sometimes lost, thus making the detection of small targets very difficult; its solution is of great practical importance to improve the performance of detection systems.

Due to the complexity of the scene, the degree of stability can affect the effectiveness of target detection. For example, the target image can be affected by uneven illumination, changes in objects in the background, etc. To extract targets from images efficiently, it is necessary to propose an algorithm that attenuates or eliminates the effects of background noise. In this paper, it combines mathematical morphology to calculate local maxima and minima to reduce the amount of subsequent processing and minimize the false alarm rate. To select a possible target, a region is grown for each local maximum point as a way to achieve the minima point.

This algorithm is used to achieve a high-frequency component with a large area of zero background and a small area target, as well as random dot noise, for each frame after opening and closing the filtered image. Then, according to the correlation of the moving objects between adjacent frames, the multi-frame superposition is carried out. The small target, because of its mobility, in the superposition of the performance of the track is a very strong correlation point, and the high frequency point noise, because of its randomness, offset each other’s part of the non-offset part of the geometric area, showing a poor correlation between the random distributions of interference points. For example, using open or closed operations for an input image, the formulas are as follows, respectively.

\[
g = f - f \circ B \quad (1)
\]

\[
g = f \bullet B - f \quad (2)
\]

where, \( f \) is the image gray frame, \( B \) is the structure, \( \circ \) denotes the open operation, and \( \bullet \) denotes the closed operation.

Filtering is performed to detect larger changes in the image, which is equivalent to the high-frequency component, and to filter out flatter changes in the image gray value, which is equivalent to the low-frequency component. Using this transform to filter the input image, the low-frequency component is filtered out, which is equivalent to filtering out large areas of the background, leaving the high-frequency component part, including the small target.

Due to the differences between image acquisition systems, image data is also often eroded by noise in practical applications. And, various redundant features are mixed together, making it very difficult to conduct the subsequent interpretation of image data. To achieve the accuracy of image data, it becomes important to find a target recognition method that can measure both the intensity discontinuity and the accurate orientation of the image, and the above problems must be overcome by a new type of uncertainty processing.
In some cases, most of the acquired images are blurred. In addition to objective reasons, there are some subjective reasons for blurred images, such as segmented and soiled images. For the ambiguity in these images, based on the uncertainty classification of the signal and the impact analysis, the paper uses the threshold method to process the fuzzy and proposes a fuzzy signal processing method, which can improve the image quality with mean square error, peak signal-to-noise ratio, and mean absolute error indicators of $3.24 \times 10^{-6}$, $1.03 \times 10^5$, and $0.0014$, respectively, all of which are superior. The results are shown in Figure 4 below:

![Processed image](image1.png) ![Original image](image2.png)

**Figure 4.** Comparison of images processed with open-closed transformation algorithm.

Since the open-closed transform filter is related to the structure size, the size of the structure determines the effect of high-pass filtering. The smaller the structure size, the more thoroughly the low-frequency background is filtered, and the smaller the target size can be retained.

In order to accurately identify the target, it is not enough to perform single image processing; small target enhancement and interference suppression are required. Since the shape and size of the two images before and after the small target dataset are different, but they still belong to the same category and have a strong correlation, the images can be superimposed before and after. All images of the same category are combined to form a set of image sequences. In the last of the same categories of the superposition, the defective target points appear as strongly correlated points, but the noise may still swamp the small targets, so, a multi-image differential superposition algorithm is proposed. The algorithm selects image sequences containing small target points, odd frames, and even frames.

\[
f_z = \sum_{i=0}^{n} (f_{2i+1} - f_{2i})
\]  

(3)

where $f_i$ is the first frame in the image sequence and $f_z$ is the last superimposed frame.

The differential superposition of multiple images further enhances small target points which can further suppress large background areas, but may also enhance noise points. Therefore, it can be performed by the threshold of the overlapping frames. The method is as follows:

\[
f_z(i,j) = \begin{cases} 
  f_z(i,j), & f_z(i,j) \geq \theta \\
  0, & f_z(i,j) < \theta 
\end{cases}
\]  

(4)
where $\theta$ is the threshold value obtained from multiple experiments, then takes the value of:

$$
\partial = \frac{2}{3} \max_{i=1}^{N} \max_{j=1}^{M} \{f_z(i,j)\}
$$

where, $M, N$ are the sizes of the superimposed frame images.

After such processing, small target points are further enhanced and random noise points will be further suppressed or eliminated, laying a better foundation for the classification accuracy obtained from the subsequent image input classification convolutional neural network.

4. Super-Resolution Adversarial Data Processing

Since deep detection networks require a large number of data sets to make them more generalizable and robust, and since surface defect targets are high-frequency network targets according to the results of the sample study, we propose to use a super-resolution generative adversarial network for the super-resolution enhancement of image data, simultaneously augmenting the number of samples. It is considered a black box test; the main function of the black box is to preserve the high-frequency information of the original image, while obtaining a high-resolution image. The loss function used to train the network in traditional super-resolution models is usually the mean square error, so the signal-to-noise index obtained may be high, but after these operations, the image performance is unsatisfactory because the high-frequency details in the image are often easily lost, yet the defective features are high-frequency details, which are obviously not allowed [20]. The super-resolution generative adversarial network uses perceptual loss and adversarial loss to enhance the realism of the recovered images, and its cognitive loss includes content loss. The super-resolution adversarial data processing network is shown in Figure 5.

![Figure 5. The super-resolution adversarial data processing network.](image)

As can be seen from Figure 5, the main function of the generator network is to input low-resolution images and then generate high-resolution images through the network. In the first step, the low-resolution image input to the generator network is processed by convolution operations and activation functions; in the second step, several residual network structures are processed, and each structure consists of two convolution operations, as well as a normalization and an activation function through the extracted regression function layer, which is a process connected to the whole residual. In the third step, after the first two processes, the feature matrix of the image is sent to the next sampling segment. After two sampling segments, the length, width, and height dimensions of the original
image become four times the original size, completing the sharpness improvement. It can be seen that the first two steps of the generative network are mainly used for feature extraction, and the third step is used to improve the clarity of the original image.

Similarly, the network structure of the discriminator is relatively simple, consisting mainly of several constantly repeating convolutions, activation functions, and a normalization process. For the purpose of the discriminator network, it is mainly to judge the authenticity of the input and output images, while the input and output are different from the generative network, because the input and output are both one image, and the output is the discriminator result. When the discriminant result is closer to 1, the discriminator considers the image as real, and vice versa, with closer to 0 as fake. The network structure of the discriminator is basically the same as that of the most common convolutional networks, i.e., the input image is down-sampled after using repeated convolutional operations, the entire connection layer is linked after repeated convolutional operations, and the final discriminant result is obtained.

The general idea of the algorithm to achieve image super-resolution using a generative adversarial network includes training of the generator and the discriminator. In general, the discriminator is trained first, followed by the generator. When training the discriminator, the discriminator is expected to be able to discriminate the authenticity of the input image; so, the image information input to the network is the real image and the fake image, and the corresponding label information. A number of real high-resolution images are randomly selected in the network system, and using the real low-resolution images after scaling, they are passed to the generator to obtain a number of virtual high-resolution images; with the label information of the real image defined as 1 and the label information of the virtual image defined as 0, the real image and the virtual image are jointly used as the training set into the discriminator to be trained.

For the training of the generator, it is expected that it will obtain extremely realistic virtual images, so when training the generative network system, it is necessary to master the criteria for the discriminator to judge the authenticity of the input image. In the training network, the real low-resolution images provided are first processed by the training generator, and after the virtual real high-resolution images are obtained, the loss is calculated and obtained by comparing the judgment results obtained from the virtual real high-resolution images with those obtained at the beginning, which means that the training generator unit completes the training according to the judgment results of the discriminator. Subsequently, the real high-resolution image and the virtual high-resolution image are put into the VGG network system to obtain the characteristics of the two images, and another loss is calculated by comparing the characteristics of the two images. The formula for each loss of the super-resolution generative adversarial network is as follows:

$$l_{SRGAN} = l_c^{SR} + 10^{-3} l_{\text{generator}}^{SR}$$

$$l_c^{SR} = \text{mse}(\text{vgg}(G(LR)), \text{vgg}(HR))$$

$$l_{\text{generator}}^{SR} = \sum_{n=1}^{N} - \log D_{\beta}(G_{\beta}(I^{LR}))$$

where $LR$ represents the low-resolution image, $HR$ represents the real high-resolution image, and $D_{\beta}(G_{\beta}(I^{LR}))$ represents the chance that the image produced by the generator is a real high-resolution image by identifying it.

In the super-resolution task, the estimation usually includes three losses: the error loss represented by MSE, the cognitive loss, and the adversarial loss of the generative adversarial network; the loss of the generative adversarial network system is used to train the network, and the cognitive loss is used by the network for the super-resolution task as a way to evaluate the performance of the generative adversarial network [30]. Content loss and adversarial loss are used together as loss functions for the optimization of the network.
After augmenting the dataset using Super Resolution Generative Adversarial Network (SRGAN), the image data were then rotated and misplaced to further augment it, and some of the data results are shown in Figure 6.

**Figure 6.** Image amplification using different methods.

The first column is the original image, the second column is the super-resolution adversarial neural network processed image, which is scaled to the same size as the original image for typographic purposes, the third column is a series of images obtained by rotation, and the last column is a series of images obtained by cutting. After these operations, a new dataset of four times the original images are obtained; Table 1 shows the number of defective images in each category. The diversity of the information set is enhanced and the generalization function of the detection network system is improved. Combining the labeling information of the defects in the original dataset, the defects in the
newly constructed dataset are re-labeled with the help of open-source labeling software (Labelme v.3.0.), which is then fed into the subsequent classification and detection network to extract the feature maps at each level and finally complete the detection task, and the final weight information obtained from the network can be applied to the actual measurement data of the plant.

<table>
<thead>
<tr>
<th>Types of Defects</th>
<th>Number of Original Images</th>
<th>After Image Amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>crazing</td>
<td>300</td>
<td>1200</td>
</tr>
<tr>
<td>inclusion</td>
<td>300</td>
<td>1200</td>
</tr>
<tr>
<td>patches</td>
<td>300</td>
<td>1200</td>
</tr>
<tr>
<td>pitted surface</td>
<td>300</td>
<td>1200</td>
</tr>
<tr>
<td>rolled-in scale</td>
<td>300</td>
<td>1200</td>
</tr>
<tr>
<td>scratches</td>
<td>300</td>
<td>1200</td>
</tr>
</tbody>
</table>

As can be seen from Figure 6 and Table 1, the super-resolution generative adversarial network proposed in this paper processes the image samples. By obtaining a set of high-resolution images that retain the defect information, combined with the classical image amplification method, it makes the sample data amplify four times of the original image and does not change the features of defects in the original image, which makes the data set richer. The problem of the insufficient number of samples is solved, and the generalization ability of the sample data and the robustness of the algorithm are improved. A good dataset is provided for the input of the subsequent classification network as well as the target detection network.

Synthetic data were generated to expand the training dataset, thus reducing the need for real data. The application in image super-resolution allows generating high-resolution images from low-resolution images. This is particularly useful for less-sample learning or resource-limited scenarios, where a small amount of real data can be passed through the generator to produce more synthetic data, which can reduce data acquisition for high-resolution images in some scenarios. By removing noise from the image, this thus reduces transmission and storage overhead. In some applications, the generator can be used to generate some samples and cache them to avoid frequent re-generation of data, thus reducing computational overhead.

5. Comparison of the Proposed Denoising Method with Other Denoising Methods

The Pytorch framework was chosen for the experiments, the system environment was Windows 11, and the GPU acceleration was CUDA 11.5.1, CUDNN8.3.1_CUDA11.5. The Intel® Core (TM) i7-10750H CPU @ 2.60 GHz was used, and the graphics card used was an NVIDIA GeForce RTX 2060 6G for computing. The program script compilation environment was PyCharm, and the program language used was Python. In order to ensure that the model can effectively reduce noise for different types of noise and scenes, this paper selects the surface defect dataset constructed by Prof. kchen Song’s team at the School of Mechanical Engineering and Automation, Northeastern University, which was introduced in the first part of this paper. On the other hand, the collected images of steel surfaces under various conditions, including different time periods, lighting conditions, shooting angles, etc., totaled 7200 images in six different categories. The extended dataset were divided into training and test sets in the ratio of 7:3. That is, 840 samples for each class of defective samples were used for training, another 360 samples were used to validate the model’s classification performance metrics. The coverage of the training dataset was high enough to cover possible noise sources and variations to ensure the breadth of the training dataset.
5.1. Comparison of the Denoising Effect

For the noisy image, noise reduction also used the Sobel operator [6] and Prewitt operator [7].

The Sobel operator is a first-order differential operator. It calculates the gradient of each pixel using the gradient values of the neighboring regions of the pixel. The detection of edges is based on the phenomenon of the grayscale between the upper, lower, left, and right neighbors of an image pixel point, while the weighted difference reaches its extreme value at the edge. It is given by the following Equation:

$$S = \left( dx^2 + dy^2 \right)^{1/2}$$  \hspace{1cm} (9)

The Sobel operator is a $3 \times 3$ matrix used to detect horizontal and vertical edges in an image; $-1, 0, 1$ and $-2, 0, 2$ denote the numbers in the matrix, which represent the relative position of each pixel point in the image. Figure 7 shows the 2 convolution kernels $dx, dy$. The maximum value of the two convolutions is the output value at that point. The result is an edge magnitude image. The edges are picked up according to a certain threshold value. The algorithm has a smoothing effect on noise and can provide more accurate edge orientation information, but the accuracy of edge localization is not high enough.

![Figure 7: Sobel operator.](image)

The idea of the Prewitt operator edge detection is similar to that of the Sobel operator. The two convolution kernels are applied to each pixel point in the image and then the two convolution results are squared to get the final edge intensity value. If the edge intensity value exceeds the set threshold, then the pixel point is considered to be an edge point. The Prewitt is given by the following Equation:

$$S_p = \left( dx^2 + dy^2 \right)^{1/2}$$ \hspace{1cm} (10)

The two convolution kernels $dx, dy$ shown in Figure 8 form the Prewitt operator. Its differential operations are defined in an odd-sized template.

![Figure 8: Prewitt operator.](image)
The algorithm template is convolved with the image pixel gray values from left to right and from top to bottom in order. The operator has a smoothing effect on noise, but the localization accuracy is also not high enough.

According to the multi-noise denoising method proposed in this paper, a simulation is performed and then compared with other noise reduction algorithms. Results can be seen in Figure 8.

Figure 9a is the image with added noise, and Figure 9b is processed by the denoising method proposed in this paper. Compared with Figure 9c, d, it can be seen that the denoising method in this paper is superior to other methods in terms of noise reduction and almost all the noise is filtered out. It shows that the multi-noise denoising method is feasible and effective.

Figure 9. Comparison of denoising effect.

5.2. Comparison of Experimental Validation of Each Index after Denoising

The more common image denoising techniques are spatial image feature denoising and transform domain denoising; the former is mainly in the use of image space for denoising processing [19], while the latter is in the use of image change space for denoising processing [20].

The spatial pixel characteristic denoising method analyzes the relationship between the center pixel and other adjacent pixels within a sliding window of a certain width, thus changing the data of the intermediate image. It can also represent the window size of partial Gaussian and Laplacian filters. Representative filter methods include computational mean filter, operational median filter, Gaussian filter, bilateral filter, etc.

The transform domain denoising method separates the useful signal from the noise in the transform domain by means of mathematical transformation, thus filtering out the redundant noise. In denoising using the image transform domain, the most central point is what kind of transformation is performed to convert the image from the spatial domain to
the transform domain. The landmark research methods include Fourier transform, discrete cosine transform, wavelet transform, and multi-dimensional geometric analysis methods.

This paper focuses on the above two denoising methods. In terms of image denoising, it needs an evaluation index to measure whether the image quality has improved after denoising an image containing noise, so it introduces an objective evaluation index to compare the image before and after denoising. As a more comprehensive evaluation calculation, three main technical indices of the maximum similarity measurement of images, namely the Mean Square Error (MSE), Peak Signal Noise Ratio (PSNR), and Mean Absolute Error (MAE), are used in the experiments. Low values of MSE and MAE only represent that the processed image is close to the original image. If the original image is blurred and the processed image is a similarly blurred image, then the MSE and MAE can be low, but the usefulness of the processed image would be limited. When training the model, the MSE causes the model to focus more on those data points that differ more from the actual values, thus prompting the model to minimize those differences. The MAE causes the model to focus more on the overall error average across all data points, without zooming in too much or ignoring any part of it.

The objective quality evaluation of images generally involves calculating certain statistical characteristics and basic physical parameters of the evaluated image, and the most frequently applied is the similarity test of the image. The similarity of the image is measured by using the deviation of the data between the image processed by the algorithm and the original image to determine how well the algorithm is processed. If the deviation is small, statistically speaking, the similarity of the image is higher because the difference between the algorithm-processed image and the original image is smaller, and thus the quality evaluation index of the algorithm-processed image is higher, and vice versa. This kind of image evaluation is usually applied to the quality evaluation of binary and grayscale maps.

Among them, the Mean Square Error (MSE) is defined as follows:

$$S_e = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i,j) - K(i,j)\|^2$$  \hspace{1cm} (11)

$I(i,j)$ denotes the original image and $K(i,j)$ denotes the algorithm-processed image. The mean square error, an image quality evaluation index, is equivalent to an intermediate value in image evaluation and provides the initial basis for many subsequent evaluation indexes.

The Peak Signal Noise Ratio (PSNR) is defined as follows:

$$P_n = 10 \times \log_{10} \left( \frac{\text{MAX}_{i}^2}{\text{MSE}} \right) = 20 \times \log_{10} \left( \frac{\text{MAX}_{i}}{\text{MSE}} \right)$$  \hspace{1cm} (12)

$\text{MAX}_{i}$ indicates the value of the largest color value among all pixels in the image. The quality of the computed image is evaluated based on the peak signal-to-noise ratio, which is the logarithm of the mean square error between the original image and the processed image relative to $(2^n - 1)^2$. The higher the value of the peak S/N ratio, the lower the distortion of the image and the higher the quality of the image.

The Mean Absolute Error (MAE) is defined as follows:

$$M_e = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |I(i,j) - K(i,j)|$$  \hspace{1cm} (13)

The mean absolute error directly reflects the effect of image processing, and the larger the error value, the worse the image is processed by the algorithm.

The experimental validation pairs of each index are shown in Figures 10–12.
II

\[
\text{MAX} \times \text{MAXP} = \frac{\text{MSE}}{\text{MSE}} = \text{log}(2^{10}) \times \text{log}(2^{20})
\]

\[I_{\text{MAX}}\] indicates the value of the largest color value among all pixels in the image.

The quality of the computed image is evaluated based on the peak signal-to-noise ratio, which is the logarithm of the mean square error between the original image and the processed image relative to \(2^{(2^n - 1)}\). The higher the value of the peak S/N ratio, the lower the distortion of the image and the higher the quality of the image.

The Mean Absolute Error (MAE) is defined as follows:

\[
\text{MAE} = \sum_{i} \sum_{j} |e_{ij}| = \sum_{i} \sum_{j} |I_{\text{MI}} - I_{\text{Kij}}|
\]

The mean absolute error directly reflects the effect of image processing, and the larger the error value, the worse the image is processed by the algorithm.

The experimental validation pairs of each index are shown in Figures 10–12.

Figure 10. Comparison of mean square error of different denoising methods.

Figure 11. Comparison of peak signal-to-noise ratio of different denoising methods.

Figure 12. Comparison of the mean absolute error of different denoising methods.
It can be seen from Figures 10–12 that through the above different denoising methods for different kinds of defects, nine different sets of evaluation index curves are obtained. Through the curves, it can be learnt that the mean square error of the denoising method proposed in this paper is lower than the other two denoising methods for six different kinds of defects; the peak signal-to-noise ratio is lower than the other two groups for only one kind of defect, and the resistance is higher than the other two denoising methods. Additionally, the mean absolute error is lower than the other two denoising methods for all kinds of defects.

The correct average recognition rate in Table 2 is calculated for each algorithm, i.e., Equations (11)–(13), under a multi-noise simulation experimental environment, averaged over 50 simulations and then averaged over six time steps. The correct recognition rate in the table is averaged over space and time and, thus, is the overall average of the correct recognition rate. The computational speed is the computational time of the algorithm itself used by the algorithm to compute the six time steps in the simulation environment. The noise suppression capability is estimated based on the computational process of the various algorithms, the complexity of which is estimated. The adaptation environment in the table refers to the environment in which the algorithm applies to the target, while the parameter setting requirement refers to the algorithm containing some parameters set according to the actual environment.

<table>
<thead>
<tr>
<th>Several Detection Methods</th>
<th>Denoising Accuracy (%)</th>
<th>Denoising Speed</th>
<th>Noise Resistance</th>
<th>Adaptation to Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>92.70</td>
<td>Fastest</td>
<td>Strongest</td>
<td>Most complex</td>
</tr>
<tr>
<td>Transformation domain</td>
<td>88.23</td>
<td>Fast</td>
<td>Stronger</td>
<td>More complex</td>
</tr>
<tr>
<td>Space domain</td>
<td>85.65</td>
<td>Faster</td>
<td>Stronger</td>
<td>General</td>
</tr>
</tbody>
</table>

From Figures 9–11 and Table 2 it can be seen that compared with the existing denoising methods, the denoising method proposed in this paper not only has a higher accuracy, faster denoising speed, and stronger anti-interference ability, but also has a better adaptation to environment. The multi-noise noise reduction method proposed in this paper lays a good foundation for steel surface defect detection; in addition, it shows that the denoising method has good comprehensive performance.

6. Conclusions

In this paper, some improvements are made based on the traditional image denoising methods, then, the advantages and disadvantages of the image multi-noise noise reduction methods in the spatial domain and transform domain are carried out. Through the analysis of the sample data, it is known that the original defect sample data all contain noise; this paper proposes an open-closed transform algorithm for noise reduction in the original data, which can effectively retain the defect information, thus reducing the possibility of the defect target being overwhelmed by noise because of its small size. In order to solve the problem of the small number of samples in the dataset publicly used for detection, a super-resolution generative adversarial network is established to augment the image samples, while combining classical image data enhancement methods, such as rotation and misplacing, without affecting the high-frequency information of the defective target so that the sample data is augmented four times of the original image, which does not only change the characteristics of the defects in the original image, but also can avoid the image defects caused by cutting or brightness change. The proposed method improves the generalization ability of the sample data and the robustness of the algorithm. It is obtained through experiments that the denoising method proposed in this paper not only has a higher denoising accuracy and faster denoising speed, but also has a stronger anti-interference ability and better denoising effect, which lays the foundation for the input of the subsequent classification network and target detection network.
However, the multi-noise denoising method proposed in this paper has been experimented with only in the case of the existence of multiple noises on the steel surface, which has certain limitations. First, the sample data selected in this paper are only the hot-rolled steel defect data in the open dataset, which is too homogeneous and does not use other types of defect sample data. In addition, the method in this paper needs further enhancement for distortion or occlusion caused by camera acquisition. Future research will consider the bounded nature of the activation function in neural networks, as the output of the activation function is also in a limited range within a certain range of the input space, which helps to prevent problems such as vanishing or exploding gradients. In this paper, further research will be carried out to verify the generalization ability of the proposed network algorithm using more open and other defective datasets, and to expand more application scenarios in order to promote generalization and wide application.

**Author Contributions:** Conceptualization, Z.C.; methodology, Y.L.; software, H.C.; validation, L.Z.; writing—original draft preparation, W.W.; writing—review and editing, Q.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Key Science and Technology Program of Henan Province (222102210084); Key Science and Technology Project of Henan Province University (23A413007), respectively.

**Data Availability Statement:** All data generated or analyzed during this study are included in this published article. We do wish to share our raw data, and these data are original.

**Acknowledgments:** This work is supported by Shuyan Wu.

**Conflicts of Interest:** This manuscript has not been published and is not under consideration for publication elsewhere. We have no conflict of interest to disclose.

**References**

2. Barniv, Y. Dynamic programming solution for detecting dim moving targets. *IEEE Trans. AES* 1985, 21, 144–156. [CrossRef]
29. Sohlberg, A.; Kangasmaa, T.; Constable, C.; Tikkakoski, A. Comparison of deep learning-based denoising methods in cardiac SPECT. *EJNMMI Phys.* 2023, 10, 9. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.