applied sciences

MDPI

Article

An Improved Feature Pyramid Network and Metric Learning Approach for Rail Surface Defect Detection

Zhendong He, Shiju Ge , Yan He *, Jie Liu and Xiaoyu An *

College of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China
* Correspondence: heyan@zzuli.edu.cn (Y.H.); anxyu@zzuli.edu.cn (X.A.)

Abstract: When deep learning methods are used to detect rail surface defects, the training accuracy declines due to small defects and an insufficient number of samples. This paper investigates the problem of rail surface defect detection by using an improved feature pyramid network (FPN) and the metric learning approach. Firstly, the FPN is improved by adding deformable convolution and convolutional block attention modules to improve the accuracy of detecting defects of different scales, and it is pretrained on the MS COCO dataset. Secondly, a new model is established to extract network features based on the transfer learning model and learned network parameters. Thirdly, a multimodal network structure is constructed, and the distance between each modal representative and the embedded feature vector is calculated to classify the defects. Finally, experiments are carried out on the miniImageNet dataset and the rail surface defect dataset. The results show that the mAP (five-way five-shot) of our method is 73.42% on the miniImageNet dataset and 63.29% on the rail defect dataset. Our experiments show the effectiveness of the proposed method, and the results of the rail surface defect detection are satisfactory. As there are few sample classification studies of rail surface defects, this work provides a different approach and lays a foundation for further research.

Keywords: feature pyramid network; metric learning; rail surface; defect detection

check for updates

Citation: He, Z.; Ge, S.; He, Y.; Liu, J.; An, X. An Improved Feature Pyramid Network and Metric Learning Approach for Rail Surface Defect Detection . *Appl. Sci.* **2023**, *13*, 6047. https://doi.org/10.3390/app13106047

Academic Editor: Diogo Ribeiro

Received: 12 April 2023
Revised: 25 April 2023
Accepted: 26 April 2023
Published: 15 May 2023

1. Introduction

The speed and frequency of train operations are greatly accelerated by the implementation of the railroad speedup strategy, in which a rail surface produces defects under different degrees of wear and tear [1,2]. Defects on the rail surface can lead to coupled vibrations when the train is operating at a high speed, which aggravates the wear and tear of its components and causes accidents [3,4]. Thus, rail surface defect detection becomes even more important. Thus far, there have been many methods for detecting defects on a rail surface. Usually, mechanical detection is performed by manual work and vision, but such detection is time-consuming and subjective and offers low accuracy [5]. Automated detection methods are also used, such as ultrasonic detection [6,7], eddy current detection [8,9], magnetic flux leakage detection [10,11], and so on [12–18], but these methods are easily influenced by the hardware of the equipment. The detection methods of deep learning [19–23] focus on image features, and compared to the abovementioned methods, the methods of deep learning are quicker and more accurate with regard to detecting rail defects. Therefore, it is essential and significant to use deep learning to detect defects on a rail surface.

As techniques for deep learning develop rapidly, and computational power increases, deep convolutional neural networks (DCNNs) are gradually being used to extract and identify features. In order to apply DCNNs to target recognition, target detection, and other fields, researchers have proposed many approaches. Ming et al. [24] proposed a method to detect the surface defects of rails using 3D-range line-scan cameras combined with deep learning to effectively eliminate the false alarms caused by light, stains, and water stains. Elhanashi et al. [25] exploited different pretrained deep learning models, such as the residual network (ResNet)-50, ResNet-101, VGG-19, and U-Net architectures to extract features from chest X-ray images and studied the use of three architectures for classification methods. Kang et al. [26] introduced a detection system to detect surface defects with complex types and only a few samples via a DCNN, which included the use of the Faster RCNN to locate the defects and the use of a deep multitask neural network to obtain classification scores and anomaly scores. Shang et al. [27] proposed a detection model for classifying defects by traditional image processing and a DCNN, which included the traditional image processing method for extracting the track part during the first stage and the fine-tuned CNN, which was used to classify the image during the second stage. Liu et al. [28] established a DCNN-based detection model and presented a new sample expansion method to solve the problem of too few samples, which included using the sample generation method to solve the problem of sample imbalance and using a DCNN to detect defects. Gibert et al. [29] applied a DCNN to the automatic detection of fastener states, which included combining multiple detectors in a multitask learning framework to improve the detection performance, but it was complex and required many training samples. Feng et al. [30] improved the You Only Look Once (YOLO) and feature pyramid networks (FPNs) to detect rail defects using the backbone network and the detection layer of MobileNet, which satisfied the requirements of defect localization and real-time processing. Yang et al. [31] presented a method for detecting and localizing the defects of a rail surface; two traditional image processing methods were used to extract the rail images, differential box-counting and the GrabCut algorithm were combined for defect segmentation, and YOLOv2 was used to precisely locate and detect the defects. Ni et al. [32] presented an attention network to achieve the defect detection of a rail surface by crossing over the consistency of the joint-guided centroid estimates, which solved the problems of complex background interference and data imbalance. Liu et al. [33] extracted multiscale features via an FPN; then, the FPN was trained by a lightweight network to detect the defects of a rail surface, which reduced the model complexity and increased the real-time performance. The aforementioned detection methods improved the detection performance by improving the network structure or loss function, but they did not deeply explore the problems of multiscale defect detection and small defect samples in complex environments.

Although many detection methods based on deep learning have been applied to detect rail defects, there are still many difficulties. The main difficulties are as follows: (1) the defect samples are too small to meet the traditional convolutional neural network training; and (2) the defect scales are varied, and the identification rate is not high, especially for tiny defects.

For difficulty 1, researchers have conducted many studies on generative adversarial networks (GAN), transfer learning, and meta learning. Zhang et al. [34] considered how to use data expansion for few-sample learning and established a data enhancement method based on feature reconstruction and morphing information. Weiss et al. [35] summarized transfer learning and discussed how to use transfer learning in the case of few-sample learning. Snell et al. [36] introduced the prototype network of few-shot learning, and the network preserved a metric space that could be classified by computing the separations of the prototype representatives in each class. Gao et al. [37] presented a prototype network for noise low-sample relation classification with mixed attention, and the network was used to solve the problems of susceptibility to noise instances and sparse features in few-shot learning. Lv et al. [38] summarized a few-sample learning method that combined a CNN and an attention module to extract image features, calculated the similarity of images by a relational network, and predicted categories according to similarities. The GAN model was uncontrollable and difficult to train. Transfer learning does not correlate well with source tasks and target tasks; the effect of meta learning is poor when the distance between the test and the training task is far.

For difficulty 2, researchers have studied defect detection via a multiscale feature fusion. Xu et al. [39] introduced a bidirectional attention FPN to solve the issue of defect feature disappearance as the network deepened. Li et al. [40] combined the Faster RCNN and the FPN by increasing the shallow refining features to detect small targets better. Based on ResNet-101 and the FPN, Li et al. [41] presented a method to detect a printed circuit board defect using an extended FPN module. Dong et al. [42] combined a global contextual attention network and the FPN to detect complex defects on surfaces. Yang et al. [43] considered low detection precision for small- and medium-sized objects in a single-shot-detector (SSD) network and proposed a detection method of a pipeline flux leakage image. Wu et al. [44] designed an extended convolution module by multiscale convolutional kernels to accommodate defects of different sizes and to enhance the ability to extract features from the network. These methods detected small target defects via the FPN, but they did not study defects with geometric deformations.

Although many methods have been proposed to detect defects by the abovementioned references, most methods only focused on one of the difficulties; this paper presents an improved feature pyramid network and metric learning approach for rail surface defect detection. The contributions of the paper are as follows:

- Considering the detection difficulties of different rail defect sizes and the few samples, an improved feature pyramid network and metric learning approach are proposed to detect rail surface defects. Compared with the existing methods, our method is more effective at classifying defects.
- An improved FPN module is proposed to overcome the multiscale defects and enhance the defect weight of the training network. The improved FPN more accurately detects small defects.
- A metric learning method is proposed to classify rail defects by calculating the distance between multimodal networks and feature vectors. This method solves the problem of having few samples.

The rest of this article is organized as follows. The related works are introduced in Section 2. The proposed method is introduced in Section 3. The experimental setup and results are reported in Section 4. The conclusion is drawn in Section 5.

2. Related Works

Representative-based metric learning (RepMet) is a new distance metric learning (DML) method, which is useful for few-shot detection; see [45], and the structure of the RepMet model is given by Figure 1. In this paper, the metric learning method of RepMet is used to identify defects.

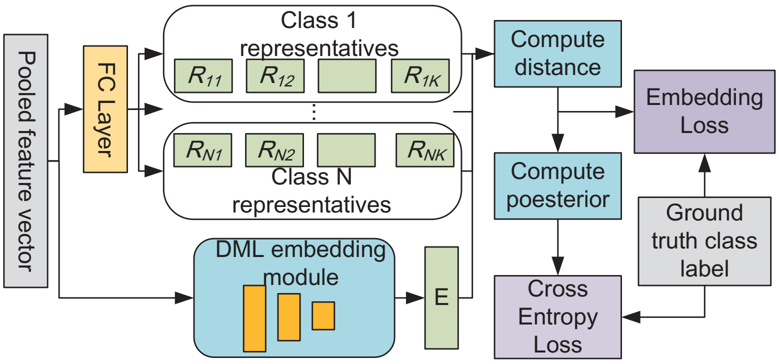

Figure 1. The structure of the RepMet model.

In Figure 1, R_{ij} is the center of the j-th mode, i is the i-th class, $1 \leq i \leq N$, N is the total number of classes, j is the j-th mode, $1 \leq j \leq K$, K is the fixed number of modules, and E represents the embedded feature vectors. As shown in Figure 1, the pooled feature vectors are converted into E through the DML embedded module, where the DML embedded module consists of several fully connected (FC) layers with batch normalization (BN) and a rectified linear unit (ReLU). The input feature transforms the pooled feature vectors into representatives of the individual classes through an FC layer. R_{ij} is obtained from the multimode mixed distribution to distinguish the mixture distribution learned in the embedding space. To classify and identify objects, the distance from E to R_{ij} is calculated and converted into the probability and the background probability of the region of interest (ROI) in each class.

Considering the detection difficulties of few samples and different scale sizes, an improved FPN was established by adding deformable convolution (DC) and an attention module, and a new method was proposed by combining FPN and Faster RCNN to locate defects. Then, the MS COCO dataset was pretrained by the improved FPN, and the trained parameters and model of the FPN were transferred to the model of defect detection, and the rail defect features were extracted and localized after finetuning. Finally, the multimodal network structure and feature vectors of DML were used to calculate the probability of classification and recognition. Figure 2 describes the structure for detecting defects on the surface of rails with few samples.

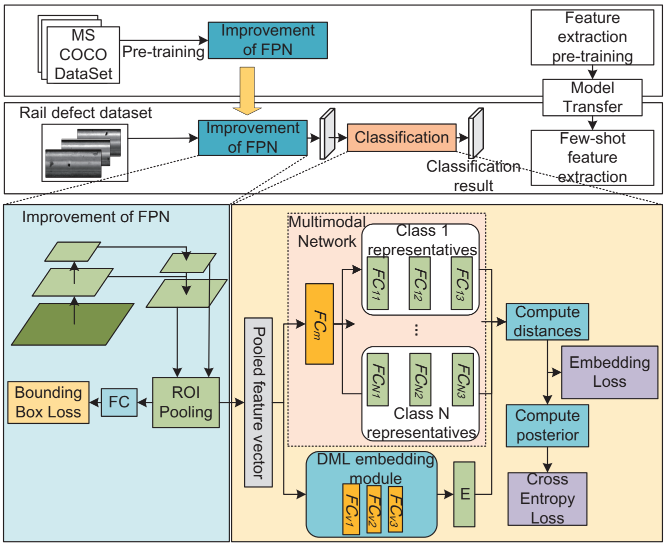

Figure 2. The structure for detecting defects on the surface of rails with few samples.

3. Main Results

3.1. Defect Feature Extraction and Location

In this subsection, we first describe the detection of the defect of the ROI and then the pooling of the ROI and the corresponding feature map of the ROI to extract defect features.

In the process of detecting defects, it is difficult to extract defects because the sizes of the defects vary. FPN is a useful tool to detect defects with different sizes, especially for small defects. Based on ResNet, the FPN extracts feature maps from different convolutional layers and then superimposes and fuses the feature maps of the previous layer and the feature maps of the current layer by twofold upsampling, which realizes the fusion of information in shallow and deep feature maps. Therefore, the FPN is improved to improve the accuracy of the model's detection of small defects.

Adding attention mechanisms to deep neural networks not only makes the network pay more attention to specific inputs but also increases the attention of important features and reduces the influence of unimportant features through weight allocation. The convolutional block attention module (CBAM) is a lightweight module that works with CNN for end-to-end training. In order to increase the weight of defective features in the network training, the features of the defect are extracted by adding a CBAM, and the CBAM structure is shown in Figure 3; see [46]. Figure 3 shows that the CBAM includes two separate submodules, i.e., the channel attention module (CAM) and the spatial attention module (SAM). The output of the CAM can be obtained using the max-pooling and avg-pooling of the shared network and the output of the SAM derived from the pool, which is transmitted between convolution layers via the channel axis. The CAM and SAM focus on the important features of the images in the channel dimension and the spatial dimension, respectively, and the two modules are connected in series. In order to obtain the attention map, the channel information of the feature map is summarized twice and connected and convolved through the standard convolution layer. The CBAM can be described by

$$\begin{aligned} F' &= M_c(F) \otimes F, \\ F'' &= M_s(F') \otimes F', \end{aligned} \tag{1}$$

where $F \in R^{C\times H\times W}$ is the input feature map, $M_c \in R^{C\times 1\times 1}$ is the 1D channel attention map, $M_s \in R^{1\times H\times W}$ is the 2D spatial attention map, $F''$is the final refined output, $\otimes$ denotes the element-wise multiplication, $R^{C\times H\times W}$ is the feature map, and C, H, and W are the number of channels and the height and width of the feature map, respectively.

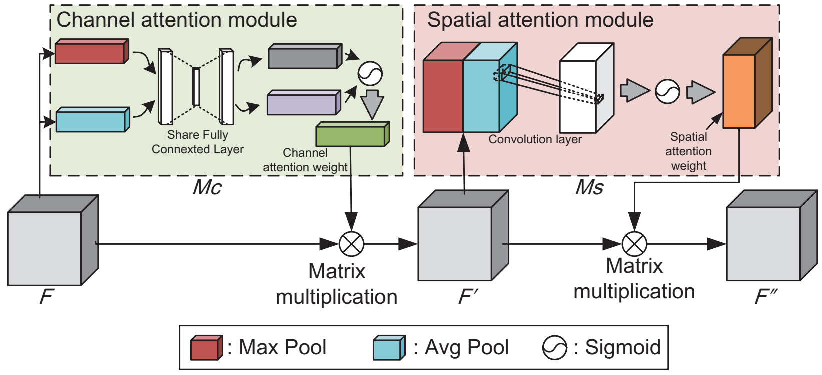

Figure 3. CBAM structure.

Next, the ROI extraction in the Faster RCNN was improved to extract the defective features by adding the FPN. Meanwhile, ResNet50 is a basic network, and the traditional convolution operation used in the basic ResNet50 can be defined as

$$y(P_0) = \sum_{P_n \in R} w(P_n) \times x(P_0 + P_n), \tag{2}$$

where $y(P_0)$ is the convolution value of point P_0 in the image, P_n is all positions in R, R is the feature space, x is the input, and w is the weight of features. From Equation (2), it can be seen that the traditional convolution has the same receptive field at any position of the image and can extract an image with a fixed size. The generalization ability of traditional convolution is limited when the scales of defects change. For the varying scale of rail defects, the DC is introduced into the ResNet50, and the DC can be defined as

$$\mathrm{y}(P_0) = \sum_{P_n \in R} w(P_n) \times x(P_0 + P_n + \Delta P_n), \tag{3}$$

where ΔP_n is the offset variable. When the offset is added to the DC, the magnitude and location of the DC kernel are adjusted based on the current object content after learning. In order to fit the shape and size of other objects, the convolution kernel changes the sampling points in variable locations according to the image context.

Next, based on ResNet50, an improved FPN was designed to obtain the feature maps of rail surfaces with different scales. The specific improvements were as follows: the DC was added to detect the rail defect features with variable dimensions; the CBAM was added to increase the weight of the defect features and reduce the amount of calculation. Region proposal networks (RPN) of Faster RCNN were used to extract the ROI of each feature map with different sizes. Usually, the sizes of the anchors in the feature map are 16 × 16, 32 × 32, 64 × 64, or 128 × 128, but the scales of the defects may not be square, and the scales of anchors were set to 1:2, 1:1 and 2:1, respectively. Figure 4 shows the structure of the improved FPN, and the structure can be described as follows: the DC operation was applied to the image, and the defect features were focused by CBAM; then, the feature maps were successively convolved 1×1, upsampled, and convolved 3×3 to obtain feature maps (denoted by P2, P3, P4, P5) with different sizes; finally, the ROI was obtained from the RPN and pooled to extract the defects.

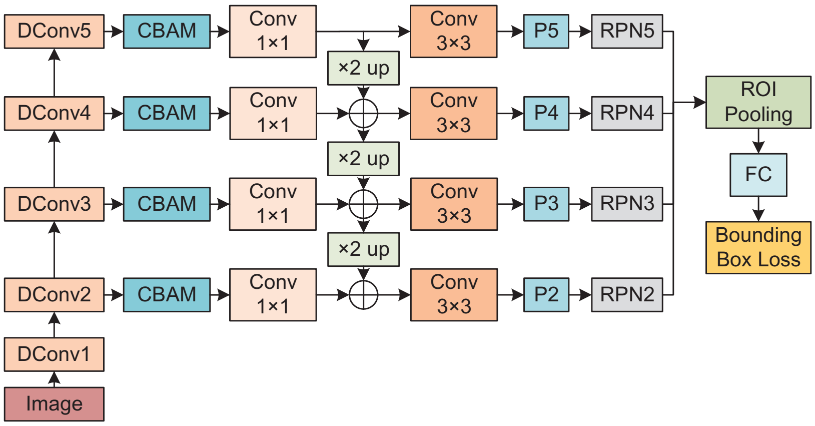

Figure 4. The structure of an improved FPN.

In dealing with the RPN, it is necessary to keep the prediction frame of the defect detection consistent with the true frame. Therefore, the frames need to be optimized by border regression functions. The position of the box is determined by its center coordinate, width, and height. Let $(x, y, w, h)^T$, $(x^*, y^*, w^*, h^*)^T$, and $(x_a, y_a, w_a, h_a)^T$ denote the center coordinates, width, and height of the predicted boundary, the true boundary, and anchor, respectively, where T denotes the transpose of a matrix. In order to calculate the deviations among the predicted frame, true frame, and anchors, the formula of the offset is established by

$$\begin{cases} t_x = (x - x_a)/w_a, & t_y = (y - y_a)/h_a, \quad t_w = \log(w/w_a), \quad t_h = \log(h/h_a), \\ t_x^* = (x^* - x_a)/w_a, & t_y^* = (y^* - y_a)/h_a, \quad t_w^* = \log(w^*/w_a), \quad t_h^* = \log(h^*/h_a), \end{cases} \tag{4}$$

where $t_i = \{t_x, t_y, t_w, t_h\}^T$ is the offset predicted by the anchor, and $t_i^* = \{t_x^*, t_y^*, t_w^*, t_h^*\}^T$ denotes the offset between the anchor and the true border. The loss function of the regression used in RPN is given by

$$L_r = \sum_i P_i^* L(t_i, t_i^*)/N_{anc}, \tag{5}$$

$$smooth_{L1}(t_i - t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2/\beta, & |t_i - t_i^*| < 1/\beta, \\ |t_i - t_i^*| - 0.5 * \beta, & Otherwise, \end{cases} \tag{6}$$

where N_{anc} is the number of anchors, P_i^* is the category of each anchor prediction, L is the $smooth_{L1}$ function, and β is a parameter that is used to control the transformation of the function. During the training, we set the intersection over union (IoU) to be the overlap rate between the predicted frame and the true frame of the ROI, and if the $IoU > 0.7$, then, the ROI is a positive sample, and $p_i^* = 1$; if the $IoU < 0.3$, then the ROI is a negative sample, and $p_i^* = 0$.

There are few samples for rails surface defect detection; thus, there are not enough samples for network training, which results in the phenomenon of overfitting. In order to avoid overfitting and obtain better network parameters, an improved FPN was pretrained by using the MS COCO dataset, and the learned parameters and model of the FPN were transferred to the rail surface detection model using transfer learning. In our work, the FPN was used for the optimization detection model; so, all the convolutional blocks of ResNet50 were used as the backbone for transfer learning, so that small-scale anchor boxes were generated from the feature maps extracted from the fifth convolutional block. Then, the detection effect of small objects was improved. The order of features extracted by the neural network was from a low to high level. The low level refers to the characteristics of strong universality, such as texture, edge, and other information, and the high level refers to the information features of the overall category of the target. Therefore, when transfer learning, the input convolutional layer was frozen to retain the low-level feature recognition information model, and the convolutional layer close to the output was finetuned to identify the rail defect feature information.

3.2. Defect Classification and Identification

Based on RepMet, a multimodal network was established to extract the feature information of the ROI, obtain different feature vectors, and distinguish the defects by measuring the distances between the feature vectors of the mode and the feature vectors of the DML.

The input of RepMet was a feature map with a fixed-size type of ROI, which was extracted by the improved FPN. Then, the convolution layer was used to further extract the ROI features to better distinguish different categories and ensure that the features extracted from the same types were uniform. The process of extraction can be defined as

$$F_i = f(x_i), \quad i \in O, \tag{7}$$

where f is the convolution operation, x_i is the feature map of the extracted ROI, O is the set of the ROI, and F_i is the feature map after convolution processing. Figure 2 shows the structure of the DML embedded module; the feature vector module E is composed of three FC layers, and each layer is followed by a ReLU to use nonlinear processing, where $FC_{v1} = 512d$, $FC_{v2} = 256d$, $FC_{v3} = 128d$, and $FC_{vt}, t \in \{1,2,3\}$ denotes the t-th FC layer of the feature vector module. Finally, the ROI was dealt by the feature vector module, and the feature vector $E_i = E(F_i)$ was obtained to extract the common characteristics of all feature information.

A multimodal network is shown in Figure 2, which extracted the same and different information using the same and different categories, respectively. Similarly, the feature map F_i was input to the high-dimensional FC layer denoted by FC_m for nonlinear processing. In order to extract richer feature information, the FC layer was set to $FC_m = 1024d$, which was convenient to use and learn the feature information of each mode. In each modal network, three FC layers with ReLU were used, where $FC_{n1} = 512d$, $FC_{n2} = 256d$, $FC_{n3} = 128d$, $n \in \{1, \ldots, N\}$, N denotes the number of classes, and $FC_{ns}, s \in \{1,2,3\}$ denotes the s-th FC layer of each mode network. After the nonlinear processing, the feature vector E_{ij} of the multimodal network was obtained, where the distance between E_i and E_{ij} is defined as

$$D_j(E_i) = D(E_i, E_{ij}) = \sqrt{\sum_{i=1}^{N} (E_i - E_{ij})^2}, j = 1, \cdots, K, \tag{8}$$

where K is the number of modes. All class distributions are assumed to be mixtures of isotropic multivariable Gaussian distributions; $D_j(E_i)$ is also used to calculate the extracted probabilities of j-th class in j-th mode, which can be defined as

$$P_j(E_i) \propto \exp\left(-(D_j^2(E_i))/(2\sigma^2)\right), \tag{9}$$

where σ^2 is the variance. By Equations (8) and (9), the posterior probability of the defective category in the ROI can be described as

$$P(C|X) = P(C|G) = \min_{j=1,\ldots,K} P_j(E_i), \tag{10}$$

where C denotes the i-th class, its minimum value is the minimum distance of all modal calculations, G is the mixture coefficient, and X is the covariance of the modes. After calculating the posterior probability of the defective category in the ROI, the posterior probability of the background category needed to be further calculated. The foreground probability was used to calculate the background probability, defined as

$$P(B|X) = P(B|G) = 1 - \min_{j=1,\ldots,K} P_j(E_i), \tag{11}$$

where B is the background category.

Next, the embedded loss function L_{em} and cross-entropy loss function L_{CE} were used in the loss function of classification and recognition, where L_{em} ensures that the distance is small between the class and the correct mode and is large between the class and the incorrect mode. L_{em} is defined as

$$L_{em} = ReLU(\min_{j} D_j(E_i) - \min_{j,i \neq i^*} D_j(E_i) + \alpha), \tag{12}$$

where i^* is the label of the correct class, α is the error between the nearest distance from E_i to the correct class and E_i to the error class. By Equations (10) and (11), the L_{CE} is defined as

$$L_{CE} = -\sum_{i=1}^{N} P(C|X) \log P(B|X). \tag{13}$$

Finally, let $L_t = L_{em} + L_{CE}$, which was used to reversely adjust the network parameters of the classification and recognition in the case of few samples.

4. Experiment

In this section, we describe how the proposed method was verified using the miniImageNet dataset, and the constructed defect dataset was used to detect and classify defects.

4.1. Experiment Dataset

In this subsection, the miniImageNet dataset and the rail surface defect dataset are introduced for use in the comparison and ablation experiments.

(1) MiniImageNet dataset

The miniImageNet dataset is a benchmark meta-learning and few-shot learning dataset, which contains 100 categories, and each category includes 600 samples. The miniImageNet dataset includes all samples of the ImageNet dataset.

(2) Defect dataset of rail surface

A common dataset given by [47] is introduced in this subsection, and the dataset consists of images of rail surfaces with at least one defect. Moreover, in this dataset, there are two types of images: one type from fast rails and the other from normal or heavy-duty rails. Then, in order to facilitate the study of the rail surface dataset, the images were cropped, and the defect types are shown in Figure 5. The figure shows that the defect images were divided into five categories: crack, regular circle, irregular, small, and blur. The crack defects are long and narrow cracks across the rail surface; the regular circle defects refer to round defects on the rail surface; irregular defects mean that the surface defects may be caused by many fine-grained shapes; small dotted defects refer to very tiny rail surface defects, and the defect can be observed when the image is enlarged; blurred defects mean that the eye cannot clearly see the outline of the rail defects.

4.2. Evaluation Metrics

In this subsection, the evaluation metrics used in this paper are introduced. In object detection, the classification target is classified as a positive or negative sample, and the prediction result is classified as true or false. Finally, there are four types of samples: true positive (TP) , true negative (TN) , false negative (FN) , and false positive (FP) .

Precision is the ratio of the TP to the sum of samples predicted to be positive, as shown in

$$Precision = TP/(TP + FP). \tag{14}$$

Recall is the ratio of the TP to the number of true positive samples, as shown in

$$Recall = TP/(TP + FN). \tag{15}$$

In addition, the evaluation metrics are given by

$$AP = \int_0^1 p(r)dr, \tag{16}$$

$$mAP = \left(\sum_{q=1}^{Q} AP_q\right)/Q, \tag{17}$$

where the average precision (AP) is the area value enclosed by the coordinate axis and the curve of precision and recall, p is the precision, r is the recall, q represents the categories, and Q is the number of categories. The mean average precision (mAP) is the ratio of the AP to the number of classes. The mAP is used as the evaluation standard in this paper.

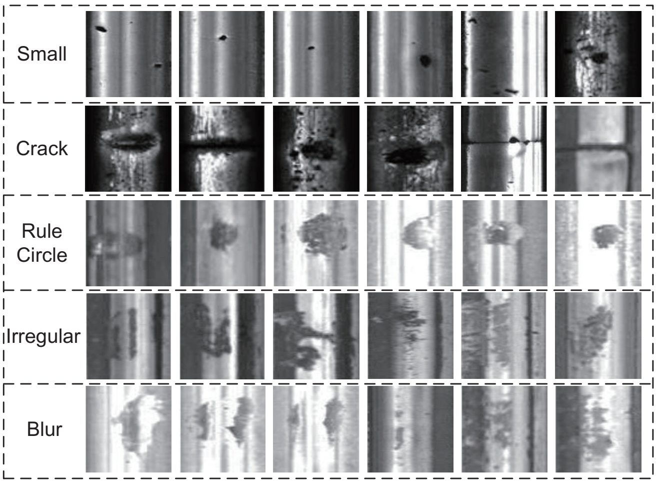

Figure 5. The types of rail defects.

4.3. Experimental Results of the miniImageNet Dataset

In the experiment, the 5-way 1-shot and 5-way 5-shot modes were used to train the network. That is, one sample (1-shot) and five samples (5-shot) were selected from each category of the defect dataset to train the network. The loss curves under the different training times are given in Figure 6. The curves in Figure 6 show that the training loss values of the methods in the figure decreased gradually, and the curves converged gradually during the training process, with the curves using RepMet and our method being more stable.

(1) Performance evaluation

In order to compare the proposed method and traditional deep learning, the methods of the training network with the miniImageNet dataset and learning from scratch are usually used to test the performance of the network. Our experiment adopted the training method on the miniImageNet dataset, and the experimental results are listed in Table 1. Table 1 shows that the mAP of our method was better than the other methods except for the MBSS method [48] from the state of the art in the 5-way 1-shot and 5-way 5-shot modes, which implies that our method is also satisfactory. Our method was 5.7% and 8.34% lower than MBSS in the 5-way 1-shot and 5-way 1-shot tasks, respectively. The backbone network used by MBSS is ResNet12, whose network model complexity is lower than that of our method.

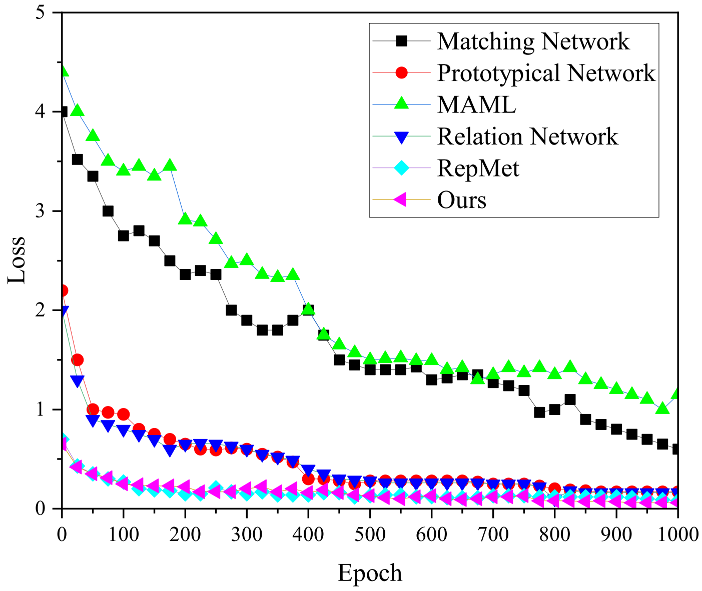

Figure 6. The training loss of the different methods.

Table 1. The mAP of different methods on the miniImageNet dataset.

Method	mAP of 5-Way 1-Shot	mAP of 5-Way 5-Shot
Matching Network	43.56%	55.31%
Prototypical Network	49.42%	68.20%
MAML	48.70%	63.11%
Relation Network	50.44%	65.32%
RepMet	56.90%	65.80%
MBSS	65.40%	81.76%
Ours	58.70%	73.42%

(2) Ablation experiments

In order to verify the effectiveness of the improved method in this paper, ablation experiments were used, and the experimental data and hyperparameters were the same as in the above experiments. The ablation experiment is an experiment, which only compares the improved part, and the others remain unchanged. The methods of the experiment were as follows: CBAM was not used in the 'non-CBAM' experiment; FPN was not used in the 'non-FPN' experiment, but we only used the last layer of features to extract the ROI; deformable convolution was not used in the 'non-DC' experiment; the pretrained model was used to extract the ROI, but the ROI extraction module was not finetuned in the 'non-FT' experiment; in the 'IOU-DML' experiment, the module of the ROI extraction combined the DML structure of RepMet. The results of the ablation experiment are given in Table 2.

Table 2. MiniImageNet dataset for the mAP results of the ablation experiments.

Method	mAP of 5-Way 1-Shot	mAP of 5-Way 5-Shot
non-FPN	31.45%	40.27%
non-DC	48.53%	61.01%
non-CBAM	58.02%	71.72%
non-FT	57.13%	69.84%
IOU-DML	56.98%	69.21%
Ours	58.70%	73.42%

By comparing Tables 1 and 2, the results can be summarized as follows: the mAP of our method was low when the FPN and DC were not used; when CBAM was not used, the ROI module was finetuned, and the loss function was not used in the module, or if the ROI module was replaced with the DML network in RepMet, then the mAPs of these experiments, as shown in Table 2, were slightly higher than RepMet in Table 1. In addition, it can be seen that the FPN and DC directly influenced the mAP of our method, and when the ROI, loss function, and CBAM were added, the performance of the network was enhanced.

4.4. Experimental Results of the Defect Dataset

In order to verify the effectiveness of our method in the rail surface defect dataset, related detection experiments were carried out.

The CBAM was tested to observe the effect of the attention module on the demonstration. Figure 7 gives the comparisons of the class activation maps between the CBAM and non-CBAM, which showed that the method including the CBAM had a significant effect on the defect features and was more effective at detecting defects.

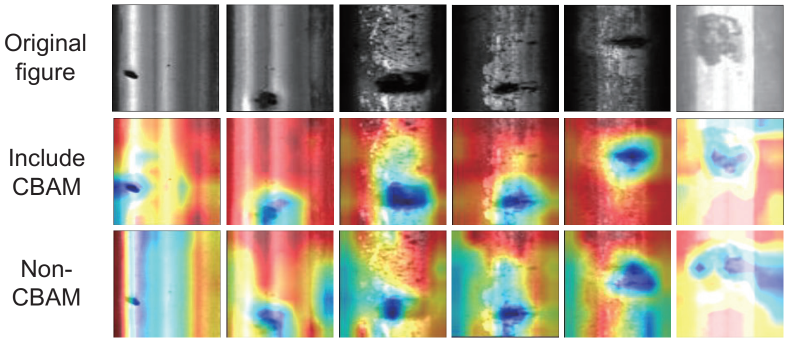

Figure 7. Comparsions of class activation maps between CBAM and non-CBAM.

In the experiment on the rail dataset, 5-way 1-shot and 5-way 5-shot modes were used to train and evaluate the performance of the network. Table 3 shows the evaluation results of the rail surface defect dataset mAP under different methods. In the case of the 5-way 1-shot and 5-way 5-shot, from Table 3, with the increase in the number of samples, the mAP of each method improved. Our method was improved over the RepMet method. In particular, compared with RepMet, the mAP of our method increased by 6.5% in the case of the 5-way 1-shot and 6.14% in the case of the 5-way 5-shot. Finally, the test results of the rail classification on five types of defects are shown in Figure 8.

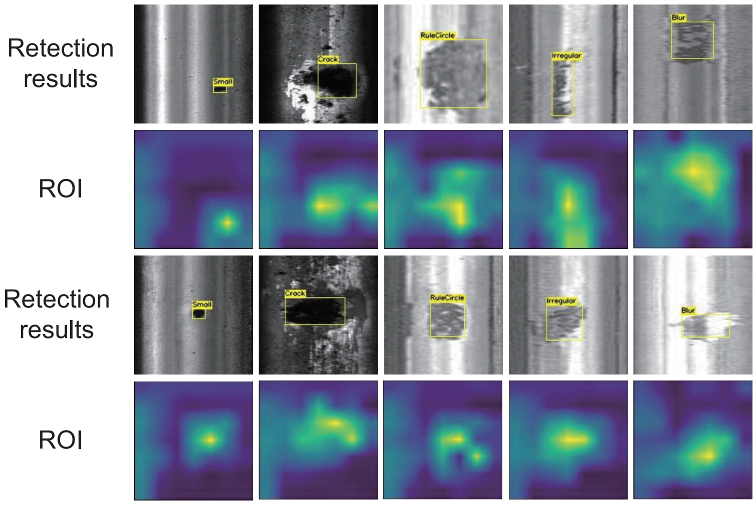

Figure 8. Test results of the rail classification.

Table 3. The mAP of the rail surface defect dataset using different methods.

Method	mAP of 5-Way 1-Shot	mAP of 5-Way 5-Shot
Matching Network	35.83%	46.92%
Prototypical Network	43.33%	48.45%
MAML	41.36%	49.72%
Relation Network	45.20%	51.36%
RepMet	51.22%	59.63%
Ours	54.55%	63.29%

5. Conclusions

Aiming at the complex rail surface, this paper proposed a rail surface defect detection method based on an improved FPN and metric learning. Based on FPN, the deformable convolution was replaced by traditional convolution to deal with the problem of the different defect sizes and easy deformation. The CBAM was added to enhance the weight of the defect features and reduce the amount of calculation. The RPN was added to extract the features and locate the bounding boxes of the small sample defects. In order to solve the problem of having only a few samples, which cannot meet the training requirements of the model, the improved FPN was pretrained on the MS COCO dataset, and the trained parameters were transferred to the rail defect detection model. Based on metric learning, a multimode network was established to classify the defects by calculating the distance between each modality and the embedded feature vector. The effectiveness of our method was verified by comparison experiments and ablation experiments. The results showed that the mAP (5-way 5-shot) of our method was 73.42% on the miniImageNet dataset and 63.29% on the rail defect dataset. Due to the complex features and high interclass similarity of the rail surface defects, the mAP values of our method on the rail defect dataset were all lower than those of the miniImageNet dataset. In terms of the limitations, our method was easily affected by the sample size and focused more on extracting the ROI, which needs to be further improved in the classifier part. In the future, we aim to develop and improve different methods to locate and classify rail surface defects. In addition to this, we will investigate data augmentation methods to enrich the current dataset and further improve the accuracy of classification.

Author Contributions: Conceptualization, Z.H. and S.G.; methodology, Z.H. and S.G.; software, Z.H.; validation, Z.H. and S.G.; formal analysis, Z.H. and J.L.; investigation, Z.H.; resources, Z.H. and X.A.; data curation, Z.H.; writing—original draft preparation, Z.H.; writing—review and editing, Z.H., S.G. and Y.H.; visualization, Z.H.; supervision, Z.H. and X.A.; project administration, Z.H.; funding acquisition, Z.H. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially funded by the National Natural Science Foundation of China under Grant 62103076 and partially funded by the Key Science and Technology Program of Henan Province (No. 222102220054, 232102210356).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The current research datasets can be downloaded at http://icn.bjtu.edu.cn/Visint/resources/RSDDs.aspx (accessed on 5 June 2022), the miniImageNet datasets at https://www.kaggle.com/datasets/arjunashok33/miniimagenet (accessed on 4 September 2022), and the code is provided at https://github.com/poetteop/rail_defect_detection (accessed on 25 April 2023).

Acknowledgments: The authors thank the editors and anonymous reviewers for their valuable suggestions.

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

References

1. Jamshidi, A.; Faghih-Roohi, S.; Hajizadeh, S.; Nunez, A.; Babuska, R.; Dollevoet, R.; Li, Z.; De Schutter, B. A big data analysis approach for rail failure risk assessment. *Risk Anal.* **2017**, *37*, 1495–1507. [CrossRef] [PubMed]
2. Ma, L.; Wang, W.; Guo, J.; Liu, Q. Study on wear and fatigue performance of two types of high-speed railway wheel materials at different ambient temperatures. *Materials* **2020**, *13*, 1152. [CrossRef]
3. Kou, L. A Review of Research on Detection and Evaluation of the Rail Surface Defects. *Acta Polytech. Hung.* **2022**, *19*, 167–186. [CrossRef]
4. Magel, E.; Mutton, P.; Ekberg, A.; Kapoor, A. Rolling contact fatigue, wear and broken rail derailments. *Wear* **2016**, *366*, 249–257. [CrossRef]
5. Oliveira, H.; Correia, P.L. Automatic road crack detection and characterization. *IEEE Trans. Intell. Transp. Syst.* **2012**, *14*, 155–168. [CrossRef]
6. Ge, H.; Chua Kim Huat, D.; Koh, C.G.; Dai, G.; Yu, Y. Guided wave–based rail flaw detection technologies: State-of-the-art review. *Struct. Health Monit.* **2022**, *21*, 1287–1308. [CrossRef]
7. Zhong, Y.; Gao, X.; Luo, L.; Pan, Y.; Qiu, C. Simulation of laser ultrasonics for detection of surface-connected rail defects. *J. Nondestruct. Eval.* **2017**, *36*, 70.
8. Zhu, J.; Tian, G.; Min, Q.; Wu, J. Comparison study of different features for pocket length quantification of angular defects using eddy current pulsed thermography. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 1373–1381. [CrossRef]
9. Piao, G.; Li, J.; Udpa, L.; Qian, J.; Deng, Y. Finite-Element Study of Motion-Induced Eddy Current Array Method for High-Speed Rail Defects Detection. *IEEE Trans. Magn.* **2021**, *57*, 1–10. [CrossRef]
10. Piao, G.; Li, J.; Udpa, L.; Udpa, S.; Deng, Y. The effect of motion-induced eddy currents on three-axis MFL signals for high-speed rail inspection. *IEEE Trans. Magn.* **2021**, *57*, 1–11. [CrossRef]
11. Jia, Y.; Liang, K.; Wang, P.; Ji, K.; Xu, P. Enhancement method of magnetic flux leakage signals for rail track surface defect detection. *IET Sci. Meas. Technol.* **2020**, *14*, 711–717. [CrossRef]
12. Kurhan, D.; Havrylov, M. The Mathematical Support of Machine Surfacing for the Railway Track. *Acta Tech. Jaurinensis* **2020**, *13*, 246–267. [CrossRef]
13. Ye, J.; Stewart, E.; Roberts, C. Use of a 3D model to improve the performance of laser-based railway track inspection. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2019**, *233*, 337–355. [CrossRef]
14. Niu, M.; Song, K.; Huang, L.; Wang, Q.; Yan, Y.; Meng, Q. Unsupervised saliency detection of rail surface defects using stereoscopic images. *IEEE Trans. Ind. Inform.* **2020**, *17*, 2271–2281. [CrossRef]
15. Ng, A.K.; Martua, L.; Sun, G. Dynamic modelling and acceleration signal analysis of rail surface defects for enhanced rail condition monitoring and diagnosis. In Proceedings of the 2019 4th International Conference on Intelligent Transportation Engineering (ICITE), Singapore, 5–7 September 2019; pp. 69–73.
16. Sysyn, M.; Nabochenko, O.; Kluge, F.; Kovalchuk, V.; Pentsak, A. Common Crossing Structural Health Analysis with Track-Side Monitoring. *Komunikacie* **2019**, *21*, 77–84. [CrossRef]
17. Zhang, Y.; Luo, L.; Zhang, Y.; Gao, X.; Long, J. Interlaced scanning by laser ultrasonic for defects imaging of train rail surface. In Proceedings of the Eleventh International Conference on Information Optics and Photonics (CIOP 2019), Xi'an, China, 20 December 2019; SPIE: Bellingham, DC, USA, 2019; Volume 11209, pp. 373–381.
18. Ramzan, B.; Malik, S.; Ahmad, S.; Martarelli, M. Railroads surface crack detection using active thermography. In Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, 12–16 January 2021; pp. 183–197.
19. Gan, J.; Wang, J.; Yu, H.; Li, Q.; Shi, Z. Online rail surface inspection utilizing spatial consistency and continuity. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 2741–2751. [CrossRef]
20. Hu, Z.; Zhu, H.; Hu, M.; Ma, Y. Rail surface spalling detection based on visual saliency. *IEEJ Trans. Electr. Electron. Eng.* **2018**, *13*, 505–509. [CrossRef]
21. Min, Y.; Xiao, B.; Dang, J.; Yue, B.; Cheng, T. Real time detection system for rail surface defects based on machine vision. *EURASIP J. Image Video Process.* **2018**, *2018*, 1–11. [CrossRef]
22. Yang, H.; Wang, Y.; Hu, J.; He, J.; Yao, Z.; Bi, Q. Segmentation of Track Surface Defects Based on Machine Vision and Neural Networks. *IEEE Sens. J.* **2022**, *22*, 1571–1582. [CrossRef]
23. Sresakoolchai, J.; Kaewunruen, S. Railway defect detection based on track geometry using supervised and unsupervised machine learning. *J. Civ. Struct. Health Monit.* **2022**, *21*, 1757–1767. [CrossRef]
24. Ming, G.; Zhou, B.; Luo, X.; Ling, R.; Zhou, M. *Rail Surface Defect Detection Method Based on Deep Learning Method with 3D Range Image*; Springer Nature: Singapore, 2023; pp. 45–59.
25. Elhanashi, A.; Lowe, D., Sr.; Saponara, S.; Moshfeghi, Y. Deep learning techniques to identify and classify COVID-19 abnormalities on chest x-ray images. In Proceedings of the Real-Time Image Processing and Deep Learning 2022, Orlando, FL, USA, 27 May 2022; SPIE: Bellingham, DC, USA, 2022; Volume 12102, pp. 15–24.
26. 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]

27. Shang, L.; Yang, Q.; Wang, J.; Li, S.; Lei, W. Detection of rail surface defects based on CNN image recognition and classification. In Proceedings of the 2018 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon, Republic of Korea, 11–14 February 2018; pp. 45–51.
28. Liu, J.; Teng, Y.; Ni, X.; Liu, H. A fastener inspection method based on defective sample generation and deep convolutional neural network. *IEEE Sens. J.* **2021**, *21*, 12179–12188. [CrossRef]
29. Gibert, X.; Patel, V.M.; Chellappa, R. Deep multitask learning for railway track inspection. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 153–164. [CrossRef]
30. Feng, J.H.; Yuan, H.; Hu, Y.Q.; Lin, J.; Liu, S.W.; Luo, X. Research on deep learning method for rail surface defect detection. *IET Electr. Syst. Transp.* **2020**, *10*, 436–442. [CrossRef]
31. Yang, H.; Wang, Y.; Hu, J.; He, J.; Yao, Z.; Bi, Q. Deep learning and machine vision-based inspection of rail surface defects. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 1–14. [CrossRef]
32. Ni, X.; Ma, Z.; Liu, J.; Shi, B.; Liu, H. Attention Network for Rail Surface Defect Detection via Consistency of Intersection-over-Union (IoU)-Guided Center-Point Estimation. *IEEE Trans. Ind. Inform.* **2021**, *18*, 1694–1705. [CrossRef]
33. Liu, Y.; Xiao, H.; Xu, J.; Zhao, J. A Rail Surface Defect Detection Method Based on Pyramid Feature and Lightweight Convolutional Neural Network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–10. [CrossRef]
34. Zhang, Z.; Wang, H.; Wang, N. Sample extraction and expansion method with feature reconstruction and deformation information. *Appl. Intell.* **2022**, *52*, 15916–15928. [CrossRef]
35. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [CrossRef]
36. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems (NIPS)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
37. Gao, T.; Han, X.; Liu, Z.; Sun, M. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 6407–6414.
38. Lv, Q.; Song, Y. Few-shot learning combine attention mechanism-based defect detection in bar surface. *ISIJ Int.* **2019**, *59*, 1089–1097. [CrossRef]
39. Xu, W.; Gan, Y.; Su, J. Bidirectional matrix feature pyramid network for object detection. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 8000–8007.
40. Li, D.; Li, R. Mug defect detection method based on improved Faster RCNN. *Laser Optoelectron. Prog.* **2020**, *57*, 353–360.
41. Li, C.J.; Qu, Z.; Wang, S.Y.; Bao, K.H.; Wang, S.Y. A Method of Defect Detection for Focal Hard Samples PCB Based on Extended FPN Model. *IEEE Trans. Compon. Pack. Manuf. Technol.* **2021**, *12*, 217–227. [CrossRef]
42. Dong, H.; Song, K.; He, Y.; Xu, J.; Yan, Y.; Meng, Q. PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection. *IEEE Trans. Ind. Inform.* **2019**, *16*, 7448–7458. [CrossRef]
43. Yang, L.; Wang, Z.; Gao, S. Pipeline magnetic flux leakage image detection algorithm based on multiscale SSD network. *IEEE Trans. Ind. Inform.* **2019**, *16*, 501–509. [CrossRef]
44. Wu, J.; Le, J.; Xiao, Z.; Zhang, F.; Geng, L.; Liu, Y.; Wang, W. Automatic fabric defect detection using a wide-and-light network. *Appl. Intell.* **2021**, *51*, 4945–4961. [CrossRef]
45. Karlinsky, L.; Shtok, J.; Harary, S.; Schwartz, E.; Aides, A.; Feris, R.; Giryes, R.; Bronstein, A.M. Repmet: Representative-based metric learning for classification and few-shot object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5197–5206.
46. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
47. Yu, H.; Li, Q.; Tan, Y.; Gan, J.; Wang, J.; Geng, Y.; Jia, L. A coarse-to-fine model for rail surface defect detection. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 656–666. [CrossRef]
48. Cheng, J.; Hao, F.; He, F.; Liu, L.; Zhang, Q. Mixer-Based Semantic Spread for Few-Shot Learning. *IEEE Trans. Multimed.* **2023**, *25*, 191–202. [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.