Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 5, No. 1, 2025

Multi-Model Convolutional Neural Network Fusion for Enhanced Metal Surface Defect Detection and Classification

Lucas Ferreira
School of Electrical and Computer Engineering, University of São Paulo, São Paulo, Brazil
lucas75@usp.edu.br

Abstract:

The rapid growth of China's non-ferrous metal industry highlights the increasing demand for high-precision defect detection in metal structures used in aerospace, automotive, and high-speed rail industries. Traditional defect detection methods, such as eddy current, magnetic leakage, infrared, and ultrasonic techniques, often face limitations in material adaptability and defect classification accuracy. In response to these challenges, this study proposes a novel approach leveraging multi-model convolutional neural network (CNN) fusion for metal surface defect detection. By optimizing existing CNN architectures and integrating a feature fusion algorithm, the proposed method effectively classifies 12 types of metal surface defects, including scratches, orange peel, and bottom leakage. Transfer learning is employed to train single CNN models, which are subsequently fused to enhance classification accuracy. Experimental results demonstrate that the multi-model fusion network outperforms individual CNN models in terms of accuracy, recall rate, and F1 score, validating the superiority of this approach for comprehensive metal defect detection.

Keywords:

Non-destructive Testing; Defect Classification; Deep Learning; Feature Fusion.

1. Introduction

In recent years, with the sustained and healthy growth of our national economy, China's non-ferrous metal industry has maintained a sustained and rapid development trend, the output has continued to grow, and the volume of import and export trade has also increased. Plate metal structure is widely used in various mechanical equipment, especially alloy sheet, which is widely used in national pillar industries such as aerospace, high-speed rail, aircraft, automobile manufacturing, etc. Cracks, scratches, bottom leaks, layer cracks and orange peel will occur in sheet metal processing. Long-term working will lead to the development of micro-defects in the metal structure into macro-cracks, which will eventually lead to structural fracture and serious catastrophic accidents. Therefore, it is of great significance to detect defects with metal surface.

At present, the traditional defect detection methods commonly used in the world are: (1) eddy current detection method; (2) Method of magnetic leakage detection; (3) Infrared detection method; (4) Method of ultrasonic detection; However, the traditional defect detection technology mainly relies on mechanical inspection method, which has limitations on inspection materials and types of defects. It also requires high professionalism of inspectors, long inspection period and poor applicability. With the improvement of computer performance, defect target recognition technology based on machine vision and deep learning[1] emerges. For example, Wang[2] et al. designed a multi-layer Convolutional Neural Network (CNN) for defect detection of six kinds of metal defects. They sampled features in the original map using sliding window method and then classified small image blocks of each kind of image. Although this method is suitable for six kinds of metals, its classification effect on metal surface defects is poor. Mei[3] and others put forward a method for texture image defect detection based on image pyramid hierarchy and convolution denoising self-encoder. This method is very effective on the repetitive background texture image set of fabric, but the effect of data set on surface of metal surface processing parts is not ideal.

Based on the above research, it is found that there are still some problems in metal surface defect detection technology at home and abroad. On the basis of the existing research on metal surface defect identification at home and abroad, the original convolution neural network model is optimized and 12 kinds of metal surface defects such as non-conductive, scratch, orange peel, bottom leakage, bump and dirt point are classified and identified by feature fusion algorithm. Without too much loss of detection speed and at the same time taking into account the improvement of accuracy, high-quality detection results can be obtained.

2. Data Acquisition and Enhancement

2.1 Dataset Introduction

The metal surface defect data used in the classification network in this paper are all real industrial metal data in the South China Sea, and are the monitoring data of defective metals in actual production. The data set is single-label data, that is, the data containing only one defect category. However, multiple defects of the same type may appear in a picture. The data includes normal samples and 11 defect samples, totaling 12 categories, and the total number of samples is 2136.

2.2 Dataset Enhancements

2136 pictures of metal surface defects are divided into a training set containing 1879 pictures and a test set containing 257 pictures. Because different convolutional network models have different requirements for data size, for the original data set, 2560 × 1920 size image, adjust the resolution of VGG16 and ResNet-50 input images to 224 × 224, adjust XCeption and Inception V3 to 299 × 299. In addition, only using transfer learning can not completely overcome the problem of insufficient data. In order to reduce the risk of over-fitting, online data enhancement technology is used to enhance the original metal data training set. The data enhancement is shown in Figure 1:

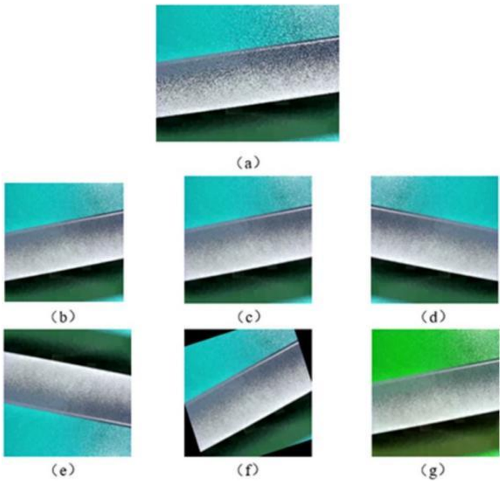


Figure 1. Two or more references

In Figure 2, a is the original data, and b and c are adjusted to 224 respectively × 224 and 299 × 299, and perform other enhancement operations on the adjusted data. D and e are obtained by horizontal and vertical inversion, f is obtained by rotating at any angle, and g is obtained by color transformation.

2.3 Analysis of Four Single Models

Some classical population convolution neural network models, such as VGG, Inception, Xception, ResNet, are more commonly used classification convolution neural network models at this stage. In order to explore the classification ability of single model for metal surface defects and facilitate the selection of single model for subsequent feature fusion of multiple models, several classical classification models used are analyzed first.

2.3.1 VGG Network

VGG network[4] is a deep convolution neural network jointly developed by well-known researchers and enterprises, which is an improvement based on AlexNet[5]. VGG-Net has five structures, each with a similar structure and 224 input image sizes × 224, VGG-16 reuses the small 3 several times × 3-convolution core is used to replace the larger convolution core in AlexNet to extract more complex and representative features. Stacking of multiple non-linear layers in the network can increase the depth of the network and further realize extraction of more complex features.

2.3.2 Inception Network

Inception[6] Network architecture was first proposed by Szegedy et al. Inception network is a parallel computation of multiple convolution kernels. Its core idea is to use the thinly connected NIN (Network In Network) idea of biological nerve[7], which greatly reduces the parameters and network complexity. When the output data of the previous layer in Inception-v1 is used as the input of the current layer, it will pass through 1 × 1,3 × 3,5 × 5 Convolution operation, and in order to limit the number of convolution input channels and reduce the amount of computation, it is still in 3 × 3 and 5 × 5 Add additional 1 before volume accumulation and after maximum pooling 1 × 1 Convolution. Finally, the output of all operations is spliced in the depth direction and transmitted to the next level. The improvement strategies adopted by Inception v3 based on Inception v1 are as follows: 1. Use convolution size decomposition technology, that is, replace large convolution operations with small convolutions; 2. Asymmetric decomposition of convolution kernel, that is, N × N Convolution is replaced by 1 × N convolution and N × 1 Convolution, the parameter quantity of this asymmetric decomposition is 2/ N of the original convolution, and the overall parameter quantity of the Inception v3 network is less than that of the Inception v1 network model.

2.3.3 Xception Network

The Xconcept[8] model was proposed by Fran ç ois Chollet in 2016. XCeption is an extension of the Inception architecture and an improved model for InceptionV3. The main improvement is to replace the standard Inception module with the deeply separable convolution. This module improves the accuracy of the model without increasing the complexity of the network. At the same time, XCeption further uses the idea of Inception. In addition to dividing the input data into several compressed data blocks, it also maps the spatial correlation for each output channel, and executes 1 × 1 Deep convolution to obtain the correlation between channels, thus decoupling the cross-channel correlation and spatial correlation. The parameter quantity of XCeption is the same as that of InceptionV3, but the use of model parameters is more efficient.

2.3.4 ResNet Network

ResNet[9] network model can train very deep network by using residual module and conventional gradient descent algorithm, and will not worry about gradient disappearance and other problems. The residual network is different from the traditional sequential network model. It adds the y=x identity mapping layer. When input x propagates forward, it passes through two convolutions and then adds with input x. That is, the multiplication of the original matrix is transformed into the process of addition. The deep derivative of the back propagation will also be directly transmitted back through the path of the identity mapping, which can make the network not degenerate with the depth increasing. Figure 2 shows a residual block of the residual network. ResNet consists of many such residual modules.

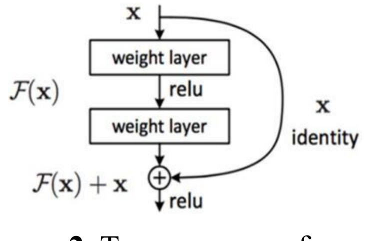

x
weight layer
relu
weight layer
$\mathcal{F}(\mathbf{x})$
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 2. Two or more references

For the task of metal surface defect classification, the VGG-16, ResNet-50, Inceptionv3 and

XCeption models with different characteristics are selected for training and classification comparison, and finally the model features with better recognition effect are selected for fusion, in order to achieve better classification effect.

2.4 Single Model Experiment and Result Analysis

2.4.1 Hyperparameter Setting

In the deep learning model, there are two types of parameters. One is the parameters of the model itself, such as the parameters in the convolution layer, pooling layer, and full connection layer. These parameters are automatically learned after the model training. The other is superparameters, such as the number of convolution layers, learning rate, epoch, optimizer selection, loss function, batch size, and drop out settings. These superparameters need to be set by yourself. The superparameters are related to the performance of the final model. After many experiments and adjustments, the model parameters for single defect classification on metal surface are finally determined as shown in Table 1:

Table 1. Hyperparameters of the single model

parameter	VGG16	ResNet-50	Xception	Inception V3
Input ImageSize	224×224	224×224	299×299	299×299
Learning Rate	0.001~0.0001	0.001~0.0001	0.001~0.0001	0.001~0.0001
Epoch	50	50	50	50
Dropout	0.5	0.5	0.5	0.5
Batch_size	40	40	32	32
Optimizer	RmSprop			
Loss Function	Cross Entropy			

2.4.2 Experimental Training

For the training of the convolution neural network model, when the amount of data is insufficient and the amount of model parameters is huge, the transfer learning idea is often used for fine-tuning training. In order to prevent the over-fitting of the training due to the lack of metal surface defect data, and to save the training time, the four single models mentioned, VGG16, ResNet-50, Inception V3 and XCeption, are used for fine-tuning training in the same way combined with the idea of transfer learning, The training process is as follows: First, four pre-training models trained by ImageNet are used as feature extraction networks, then the output neurons of the classifier of the pre-training model are removed, and the number of outputs is changed to the number of types of metal surface defects (12 types), and the Dropout layer is added to prevent over-fitting. Finally, selectively freeze part of the network layer, send the metal surface defect training set into the model for retraining, and constantly adjust the parameters and the number of frozen layers to achieve the best recognition effect.

In order to avoid ab initio training, reduce the amount of calculation, and make full use of the feature extraction ability of the pre-training network learned from the source data, it is necessary to selectively freeze the number of network layers during training. For the frozen network part, the network parameters are not updated, but only the unfrozen part. After repeated experiments, for VGG16, ResNet-50 Inception V3 and XCeption have frozen the first 34 layers, the first 168 layers, the first 249 layers and the first 126 layers respectively. Figure 3 shows the freezing layer setting of VGG16. The freezing layer setting principle for the other three models is similar.

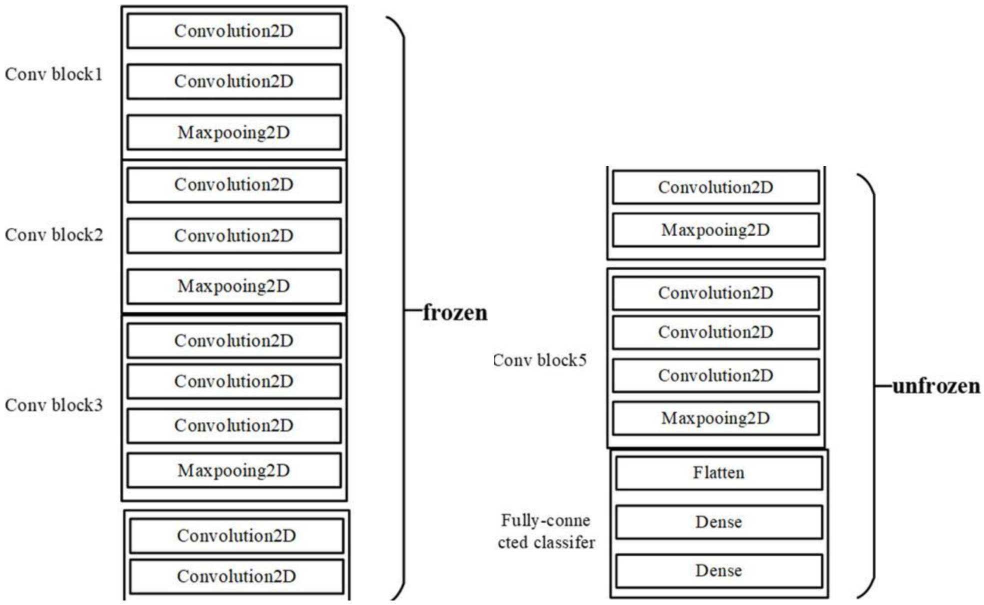
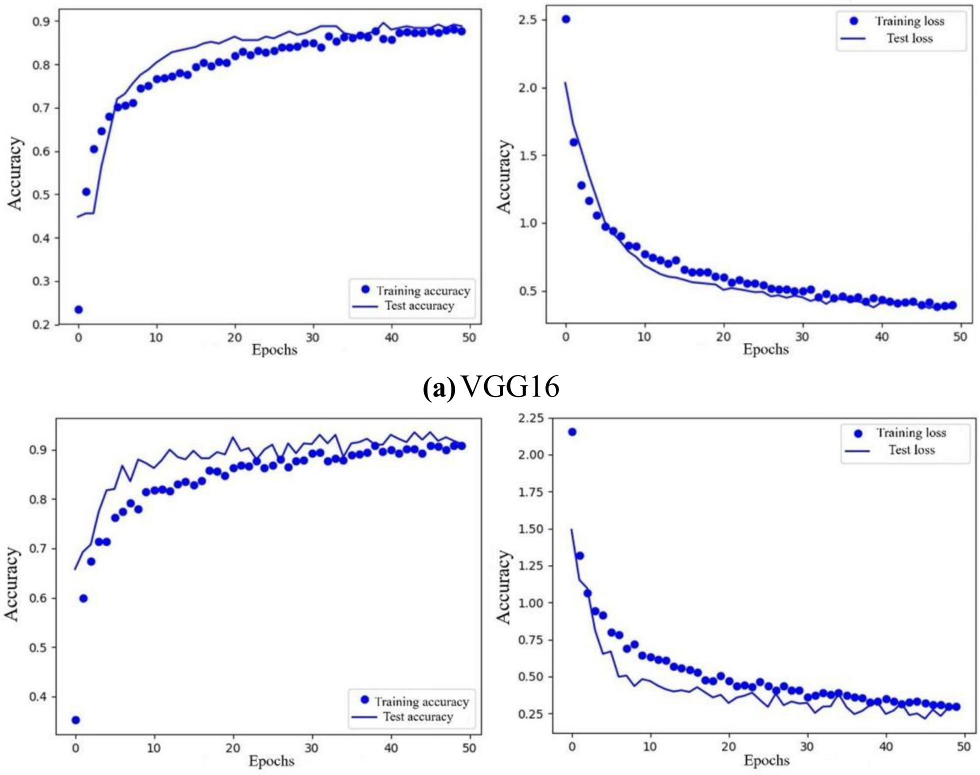


Figure 3. Freeze layers setting of VGG16

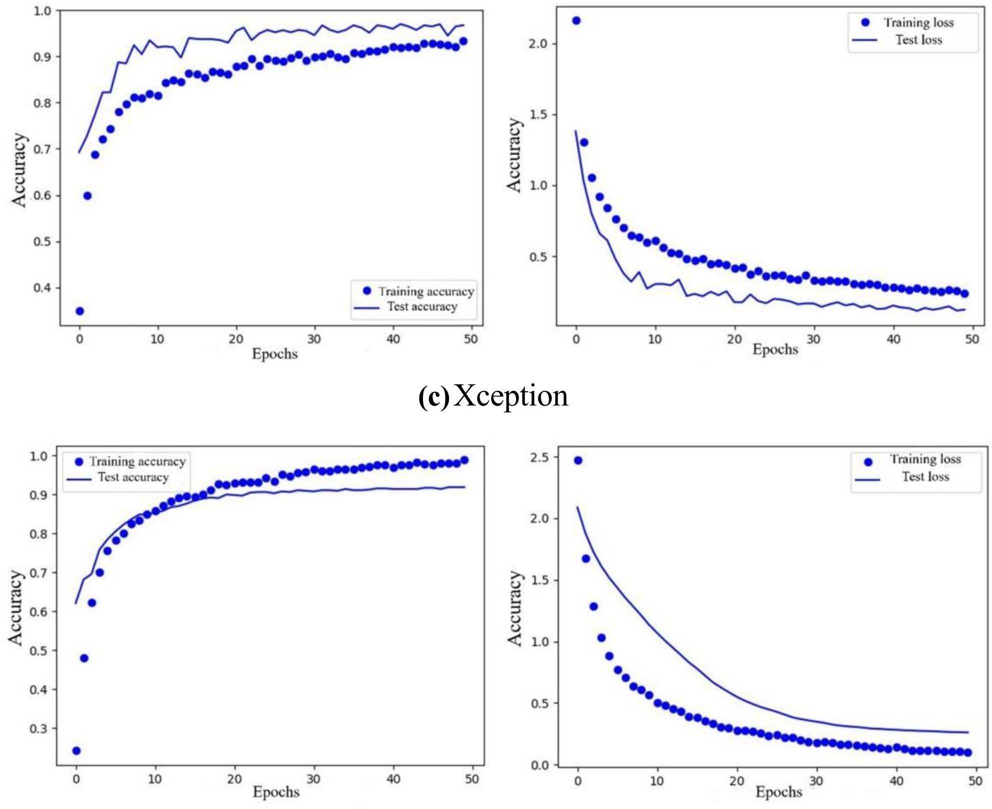
2.4.3 Comparison of Experimental Results

For the 50 epochs to be trained, the data in the test set will be tested once after each epoch is trained, and the accuracy and loss values of the training and testing will be recorded at the same time. The 12 metal defects will be trained and tested on the training set and test set of metal defects. The accuracy and loss values of the training and testing during the whole training process are shown in Figure 4.



(a) VGG16

(b) InceptionV3



(c) Xception

(d) ResNet-50

Figure 4. Single model accuracy and loss value

In Figure 4, the left side shows the accuracy of the model in the training set and test set, and the right side shows the loss value in the training set and test set. It can be seen from the training loss curve of each model that the training process of the four models is good, the loss value drops steadily, and there is no excessive jitter and over-fitting phenomenon.

Table 2. Single model performance comparison

Network model	Accuracy(/%)	Precision(/%)	Recall (/%)	F1 score (/%)
VGG16	88	84	82	83
InceptionV3	91	90	87	88
Xception	93	92	91	91
ResNet-50	92	91	91	91

Table 2 shows the performance comparison of the final single models predicted by the test set. From the above chart, it can be seen that ResNet-50, Xception and InceptionV3 are the best models. Their final classification and recognition accuracy rates on the test set are 92%, 93% and 91% respectively, and the accuracy rates are also higher than 90%. The single model with the worst classification effect is VGG16. The classification accuracy of the model in the test set is 88%, and the F1 score, accuracy and accuracy are the lowest.

3. Research on Multi-model Feature Fusion Classification Method

3.1 Structure of Feature Fusion Model

Four convolutional neural network models are used to directly use the extracted features to classify 12 kinds of metal defects. Among them, three models perform well in classification, including

ResNet-50, InceptionV3 and Xception. It is proposed to use the extracted features of these three models to fuse and then build a feature fusion model for classification.

Multi-model feature fusion network is shown in Figure 5. First, the input image is enhanced and adjusted to the size required by the single model as input. Then, the fine-tuned single model is used as the feature extraction part of the metal surface defect image, and the 2048 dimensional feature vectors will be output after the image is input. The three feature vectors extracted from the three single models are superimposed and expanded into 6144 dimensional vectors. Finally, the feature vectors of 6144 are sent into the designed classifier for classification and recognition. Finally, 12 kinds of vectors are output. In order to further prevent the occurrence of over-fitting phenomenon, a Dropout layer is added in the layer before the output, with a random deactivation ratio of 0.5.

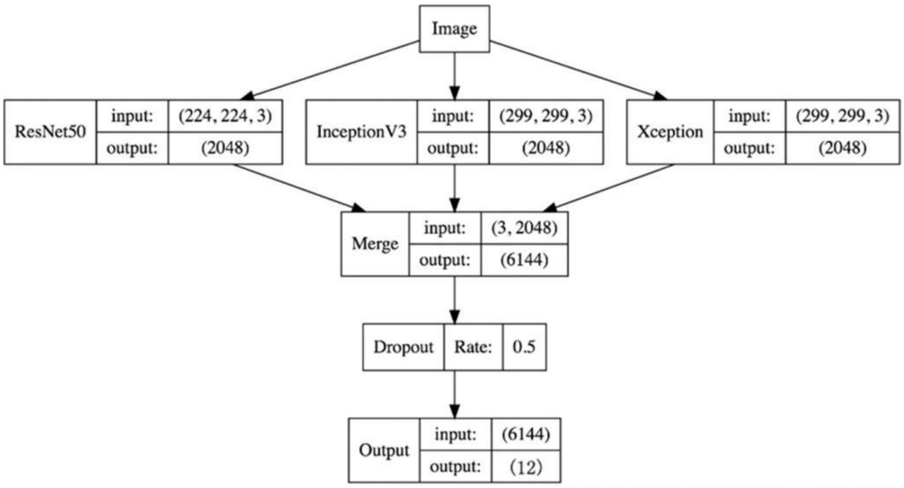


Figure 5. Multi-model Feature Fusion Network

3.2 Feature Fusion Model Experiment and Result Analysis

Similarly, the multi-model feature fusion network model is different from the single model. It does not need to re-train three single models, but only needs to train the classifier part of the output, so there is no need to change the super-parameters, that is, the super-parameters of the multi-model are the same as the super-parameters of the single model. A total of 50 epochs need to be trained, and 12 kinds of metal defects need to be trained and tested on the training set and test set of metal surface defects, The accuracy rate and loss value of training and testing of multi-model feature fusion in the whole training process are shown in Figure 6.

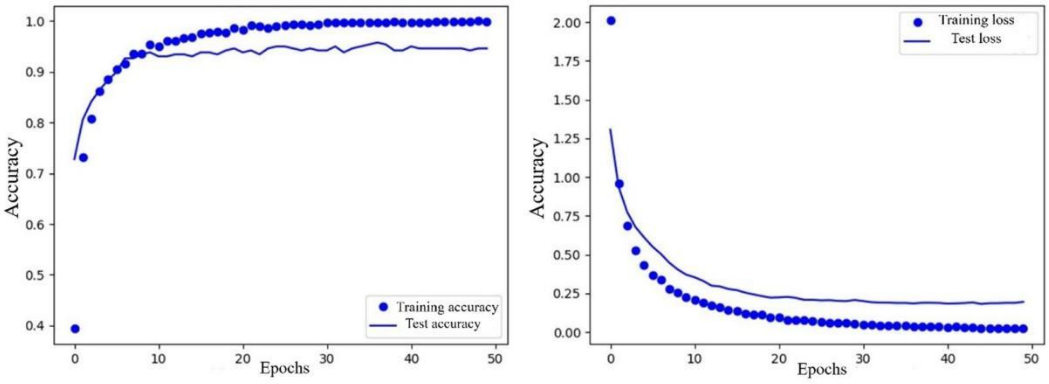


Figure 6. Accuracy and loss value of feature fusion model

Finally, the loss value of the multi-model feature fusion network in the metal defect test set is only 0.19, and the training loss curve tends to be stable after rapid decline, without over-fitting. And from the training loss curve, it can be seen that the training process of the feature fusion model is more stable than that of the single model, and the convergence speed is faster. Finally, after the prediction of the prediction set, the performance comparison between the obtained evaluation index and the optimal single model Xcept is as follows:

Table 3. Performance comparison between single model and feature fusion model

Network model	Accuracy(/%)	Precision(/%)	Recall (/%)	F1 score (/%)
Multi-model feature fusion network	95	95	95	95
Xception	93	92	91	91

As can be seen from Table 3, the multi model feature fusion network performs better in all indicators and has better robustness than the Xeption model, with an accuracy rate of 95%, 2% higher than the single model Xeption with the best performance, and 3%, 4% and 4% higher respectively than the accuracy rate, recall rate and F1 Score, which fully verifies the effectiveness of the feature fusion model. In order to further verify the performance of InceptionV3, Xception and Resnet-50 fusion models in the detection of metal surface defects, the trained models are used to predict and count 257 test set samples. The prediction results of each category are shown in Table 4:

Table 4. Prediction effect

Network model	Precision(/%)	Recall (/%)	F1 score (/%)
Non-conductive	100	100	75
other	80	84	75
Pulverized powder	75	75	75
Scratch	91	91	91
Orange peel	67	77	72
Cross bar indentation	100	100	100
normal	100	98	99
Coating cracking	100	100	100
Bottom Leak	98	100	99
Bruising	100	77	87
Dirty spots	91	100	95
Pit-up	100	100	100
Weighted Average\Total	95	96	95

In Table 4, we can see that the accuracy of most categories is more than 85%, non-conductive, cross bar indentation, normal, bumping and pitting reach 100%, recall rate and F1 index of each category are more than 90%, that is, the model has low risk of miss detection and false detection, at the same time, it has high recognition rate and better robustness, which verifies that multi-model feature fusion has strong classification and recognition ability.

4. Conclusion

In view of the limitation of single convolution neural network model in metal surface defect classification, a method based on multi-convolution neural network model fusion is proposed to classify and identify metal surface defects, and good classification results are achieved.

First, four single-model metal defect classification networks are trained by using the method of transfer learning, and then three single models with better classification effect are selected to fuse, and the enhanced data are sent into the multi-model feature fusion model to train the metal surface defect classification network based on Multi-model feature fusion. Finally, the single model and multi-model network are compared by classification experiment. The experimental results show that the final multi-model feature fusion network model improves the classification accuracy, accuracy, recall rate and F1 Score of metal surface defect test set compared with the highest single model, and the multi-model feature fusion network shows good classification performance for 12 metal surface defect categories, which verifies the effectiveness and superiority of the multi-model fusion classification method for metal surface defect classification network.

References

[1] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7533): 436-444.

[2] Wang T, Chen Y, Qiao M, et al. A fast and robust convolutional neural network-based defect detection model in product quality control[J]. International Journal of Advanced Manufacturing Technology, 2017, 94(9): 3465-3471.

[3] Mei S, Yang H, Yin Z. An Unsupervised-Learning-Based Approach for Automated Defect Inspection on Textured Surfaces[J]. IEEE Transactions on Instrumentation and Measurement, 2018:1266-1277.

[4] S. Liu and W. Deng, "Very deep convolutional neural network-based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.

[5] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[6] Szegedy C, Vanhoucke V, loffe S, et al. Rethinking the Inception Architecture for Computer Vision[J]. IEEE, 2016:2818-2826.

[7] Lin M, Chen Q, Yan S. Network in Network[J]. Computer Science, 2013,4(5):1-10.

[8] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.

[9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.