

Research Article

Hot-Rolled Steel Strip Surface Inspection Based on Transfer Learning Model

Hao Wu ^{1,2} and Quanquan Lv ²

¹Anhui Province Key Laboratory of Special Heavy Load Robot, Maanshan 243032, China

²School of Mechanical Engineering, Anhui University of Technology, Maanshan 243032, China

Correspondence should be addressed to Hao Wu; wuhaomoses@qq.com

Received 4 December 2020; Accepted 9 May 2021; Published 29 May 2021

Academic Editor: Giovanni Diraco

Copyright © 2021 Hao Wu and Quanquan Lv. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the production process of steel strips, the detection of surface defects is very important. However, traditional methods of defect detection bring problems of low detection accuracy and dependence on subjective judgment. In this study, the surface defects of steel strips are detected by a classic convolutional neural network method that is improved by the use of a transfer learning model. This model has the advantages of shorter training time, faster convergence, and more accurate weight parameters. The transfer learning model obtained through experiments secures better results in defect detection than the classic convolutional neural network method, as its accuracy of training and testing has reached about 98%. Finally, a model based on a full convolutional neural network (FCN) is proposed for segmenting the defective areas of steel strips.

1. Introduction

As an important steel product, the hot-rolled strip is widely used in manufacturing industry [1–3]. During the process of producing hot-rolled strips, factors such as the declining performance of rolling equipment and fluctuations in the production process may cause different types of surface defects to appear on the surfaces of hot-rolled strips, which will result in financial losses. In practice, inspection of the strips is usually carried out visually by individuals, a method that is unreliable and time-consuming. With the development of computer vision and pattern recognition, a series of automated methods have been applied to the inspection of strip steel [4, 5]. Chengming et al. [6] proposed a research method for surface quality inspection of cold-rolled strips based on a back propagation (BP) neural network. First, the features were extracted by a wavelet transform, the nonlinear characteristics of a pattern-recognition method based on a BP neural network were used, and then images of five kinds of typical cold-rolled strip surface defects were studied. The average recognition rate was 92%. Jiahui et al. [7] proposed a method of applying a Gabor filter for the surface defect detection of strip steel. Using the Gabor filter's characteristics

of frequency selection and direction selection, the evaluation function was introduced to maximize the difference of energy response between images containing defects and defect-free images, and the image with defects was segmented. Ke et al. [8] applied the hidden Markov tree model to the on-line detection of the surface defects of the strip steel, respectively, modeling and realizing multiscale defect segmentation using this model. When different types of defects are represented by the same “defect model,” the segmentation accuracy of the model for common defects on the strip surface reaches 94.4%.

The traditional methods of manual detection and classical pattern recognition described above are highly dependent on the experience parameter-setting of operators or algorithms. These methods are often only suitable for the inspection of strip steel under specific conditions. If the external environment changes slightly, such as a change in illumination intensity, the change of image size and the increase of noise carried by signal source will cause the detection to fail. In recent years, due to the excellent performance of deep-learning technology in many visual tasks [9, 10], a method based on deep learning is proposed for the automatic detection of defects on steel surfaces, replacing manual inspection and traditional pattern-recognition methods.

Yanxi et al. [11] proposed a detection algorithm for defects on strip steel surfaces based on convolutional neural networks (CNN). By establishing a CNN model with the introduction of deep-learning knowledge and the creation of data sets, the automatic extraction and detection of strip steel surface defects have been realized. The effectiveness of the algorithm has been verified by experiments.

He et al. [12] proposed a hierarchical learning framework based on convolutional neural networks to classify the defects of hot-rolled steel and introduced a multiscale receiving field (MSRF) to be used together with the pretraining model concept-v4 to extract multiscale features. At the same time, a group of small automatic encoders (AE) has been trained to reduce the size of extracted features adaptively, so as to avoid overfitting the training set. Experiments on samples captured from two hot-rolling production lines show that the proposed framework achieves 97.2% and 97% classification rates, respectively, which is much higher than the traditional method.

Other deep-learning algorithms applied to the location and classification of strip surface defects include fast r-CNN [13], yolov3 [14], and SSD [15], which can realize the function of locating the position of defects.

Therefore, because the machine vision and deep-learning technology can realize the automation and intelligentization of the product surface defect detection process, this technology has great practical application value. It has been applied to the surface defect identification of steel as mentioned above and the wood defect inspection, glass surface defect inspection, shape defect inspection of small electronic components and surface defects of magnetic sheet, etc. For manufacturers, this technology can not only better guarantee the production quality of products but also greatly improve the production efficiency of products and greatly improve the competitiveness of industrial production.

In this paper, the convolutional neural network model is first introduced, and then, a transfer learning model with higher accuracy is proposed. For the detection of surface defects of hot-rolled strips, we compared a convolutional neural network with a transfer learning model. The training time of the transfer learning model [16] is shorter, its convergence speed is faster, the weight parameters are more accurate, and the accuracy of its detection of the surface defects of strip steel is also higher.

2. The Structure of a Convolutional Neural Network

A convolutional neural network is mainly composed of several layers, including an input layer, convolution layers I and IV, pooling layers II and IV, a full connection layer V, and a softmax layer VI, as shown in Figure 1.

As its name implies, the input layer is responsible for the input tasks of the whole convolutional neural network. When the input is an image, the input represents the pixel matrix of the image. The input layer transforms the pixel matrix or three-dimensional matrix into another three-dimensional matrix, layer by layer through different neural network structures, until it finally reaches the full connection layer.

The convolution layer is the most important part of a convolutional neural network. The information received by the convolution layer is only a small part of the upper layer neural network. However, although only a small piece of information is accepted, the convolution layer will process the information at a deeper level, so as to obtain higher abstract features.

The role of the pooling layer is completely different from that of the convolution layer. When the pooling layer receives the node matrix from the previous layer, it will reduce the size of the matrix. This process therefore reduces the parameters of the whole neural network and the number of nodes in the full connection layer.

The full connection layer is located in the last two layers of the convolutional neural network. It mainly gives the classification results for the information processed by the multilayer convolution layer and the pooling layer. In the convolutional neural network, the convolution layer and pooling layer are equivalent, so this is a tool to automatically extract image features and enhance information content, while the full connection layer takes over the final work of collection and classifies and summarizes the final information.

The softmax layer is also used for classification problems, and this layer gives the probability of the current information being classified into each category.

3. Transfer Learning

Transfer learning is a kind of learning process that uses the similarity of data, tasks or models to apply a model learned in an old field to a new field.

Transfer learning has emerged to solve the contradiction in the current state between big data and less annotation and weak computing. Under the current state, we possess a great deal of data information consisting of images, languages, and text, but most of these data are unprocessed, and only a few of them have been correctly labeled manually. Similarly, big data also requires more powerful equipment for storage and calculation, but this condition cannot be met by most ordinary scholars. To solve these contradictions, the theory and application of transfer learning have emerged as the times require. Through the methods of migrating data annotation, model migration, and adaptive learning, migration learning can meet the user requirements of large amounts of data annotation, fine-tuning after model migration, and personalized and flexible adjustment of the model. The transfer learning model used in this paper is an image classification model trained on the large image database named ImageNet. It often takes weeks or even months to train the model on a powerful computer with multiple GPUs. The model can classify over a thousand kinds of images. After training the model, we only need to understand the input and output layer and how to fine tune it. Therefore, after using large sample data sets, the same model solves the classification problem well. Then, the weight parameters used in training the model can effectively complete the tasks of feature extraction and classification on small data sets, as shown in Figure 2. The upper part is the model pretrained on the large data set, and the lower part only needs to adjust the network

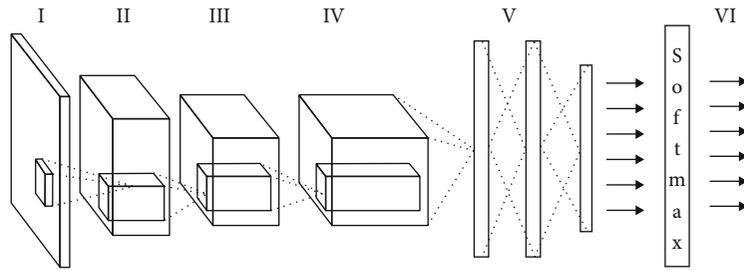


FIGURE 1: Architecture of a convolutional neural network.

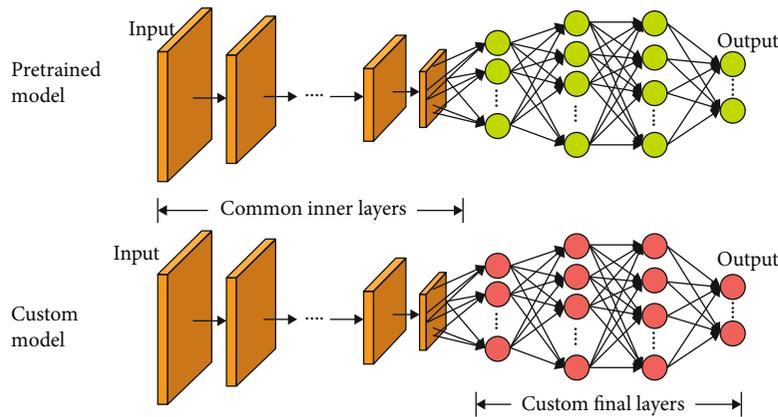


FIGURE 2: Transfer learning model [17].

output structure according to the actual needs. The model is applied to the small data set in the lower part.

3.1. FCN. A fully convolutional neural network (FCN) is used here for segmentation. Figure 3 is a schematic diagram of an FCN network. The full connection layer is behind the CNN network, and the full volume build-up layer is changed behind the FCN network. FCN has made this change based on this. A classification network is a whole reasoning process from picture to probability value—that is, when the input is a picture, the output is a number. However, a FCN can input any size and then produce the corresponding size output through an effective reasoning and learning process—that is, the input is a picture, and the output is a picture. Any input size is ideal. This was also our original intention in choosing an FCN model.

The FCN network has achieved very good results in three relatively comprehensive data sets. There are five downsamplings in FCN. The first three downsampling blocks are called the shallow network, and the last two are deep networks. In the field of semantic segmentation [18, 19], there will be a more important problem, that is, local information and global information. The local information receptive field is smaller, but the global information receptive field is larger. The size of the receptive field sometimes affects the semantic segmentation of images, in the way that a large window and a small window have different views. In the segmentation of small objects, if the receptive field is large, it will lead to too much information to separate the object

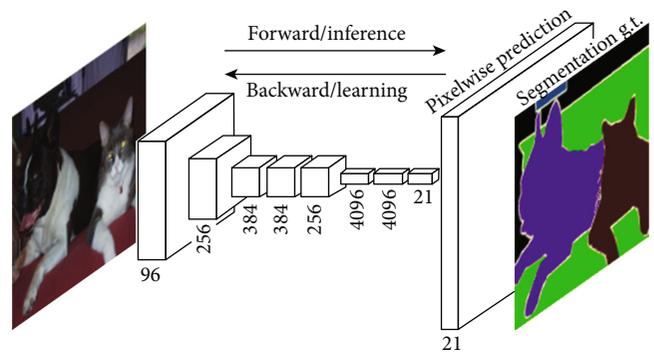


FIGURE 3: Schematic diagram of fully convolutional neural network [20].

accurately, and vice versa. Local information is helpful to segment smaller targets, while global information is generally used to segment larger targets. Then, the FCN defines a jump connection, which combines semantic information from the deep layer and representation information from the shallow layer, fusing these two sources of information to segment more accurately and precisely.

4. Experimental Analysis and Results

4.1. Data Set. In the production process of hot-rolled strips, surface defects will inevitably be caused, for a variety of reasons. As well as improving the production conditions and

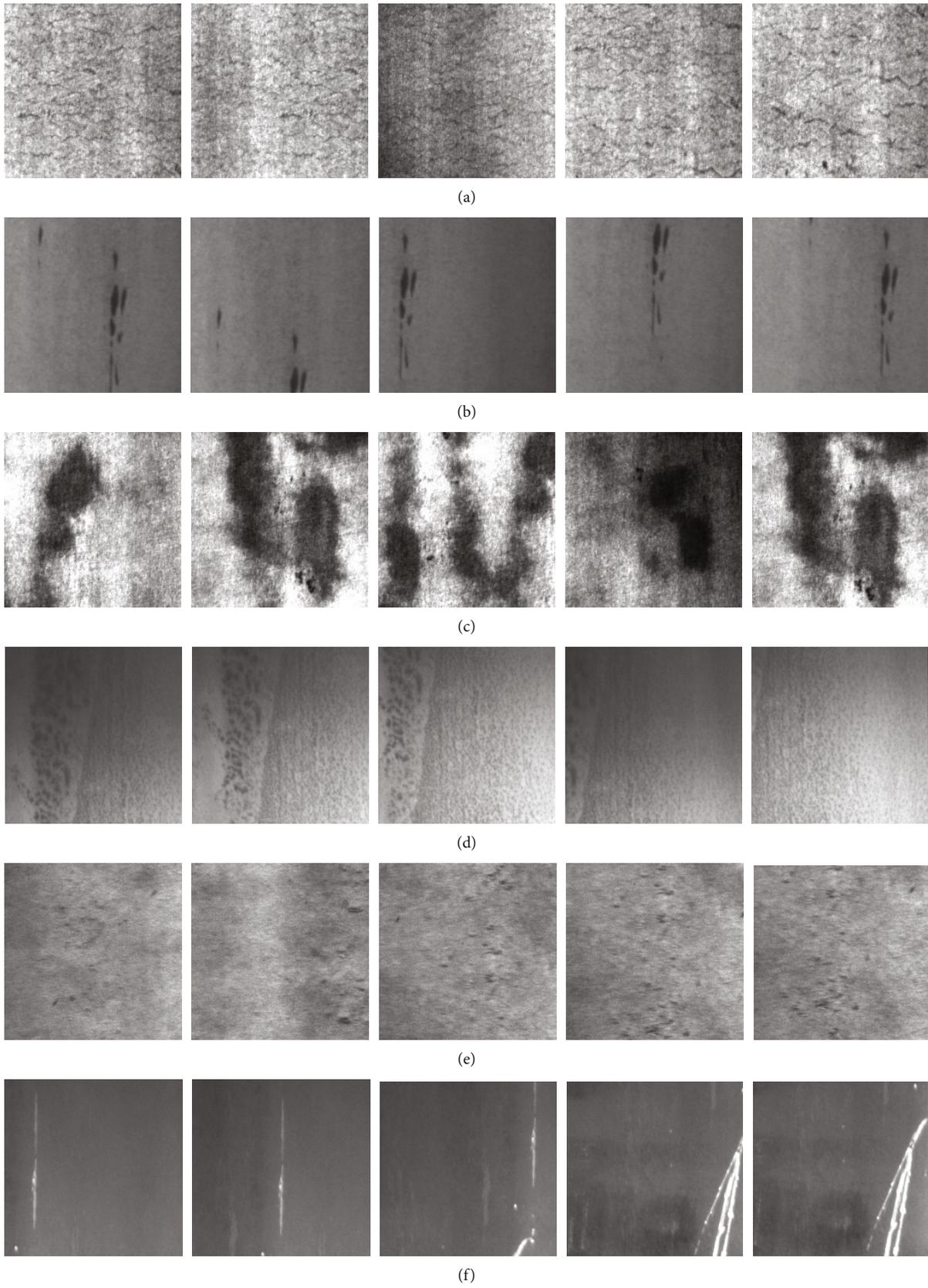


FIGURE 4: Six types of surface defects of hot-rolled strip: (a) cracking; (b) inclusion; (c) patches; (d) pitted surface; (e) rolled in scale; and (f) scratches.

production process and reducing the possibility of surface defects, it is also necessary to do a good job of surface defect detection. It is therefore particularly important to use good defect detection methods. In order to use a surface detection method based on a convolutional neural network, a complete and effective data set is required. The data set of surface defects of hot-rolled strips used in this experiment is from Northeast University [21], as shown in Figure 4. It includes six types of common surface defects of hot-rolled strips: rolled in scale (RS), patches (PA), cracking (Cr), pitted surface (PS), inclusions (in), and scratches (SC). It is difficult to collect the surface defect images of hot-rolled strips, so there are only 300 images for each type of defect, a total of 1800 defect images.

The second training set contains a total of 600 defect images, and the test set contains 300 defect images. The third training set contains a total of 1200 defect images and a test set of 600 defect images. In this experiment, we use the Google collaborative notebook with the TensorFlow platform to load the Google cloud hard disk. We can directly import the data set from the Google cloud hard disk by modifying the path and realizing the cloud GPU operation.

4.2. Experiment Based on a Convolutional Neural Network Model.

The experiment is divided into training sets and a test set. The first training set contains 300 defect images (average score of six categories, 50 images of each category) and 150 defect images of the test set (25 images of each category). In this experiment, the training times of Experiment 1, Experiment 2, and Experiment 3 are 78 minutes, 160 minutes, and 300 minutes, respectively, when the epoch is 20. The training accuracy of the model is still the measure of test accuracy. With the increase of epoch, they all show an increasing trend, but the growth results are different. When the epoch is 20, the training accuracy is 0.8733, 0.9517, and 0.9498, and the test accuracy is 0.4867, 0.7067, and 0.6933, respectively. The training accuracy of each group of experiments can reach about 90%, but the accuracy of the model test is only about 70%.

4.3. Experiment Based on a Transfer Learning Model.

Previous experiments based on a convolutional neural network model enjoyed good training results, but their evaluation test results were not very good.

Given that transfer learning has many advantages, including no need to retrain the model, faster training speed, and less time spent, a strip detection experiment based on a transfer learning model was carried out. The transfer learning method replaces the last full connection layer of the model pretrained on ImageNet data set and trains a new full connection layer with the output of a bottleneck layer, so as to deal with the classification problem for hot-strip defects.

The number of the experimental data set is 1800, there are a total of six types of hot-rolled steel strip defect sample pictures, and the experimental data set is the same as the one used before with the convolutional neural network model, so the training effects can be compared to a certain extent. The parameter-settings are as follows: the learning rate is 0.01, the batch is 100, and the step is variable. By

TABLE 1: Experimental results of transfer learning model.

Step	Validation acc	Test acc
20	58.00%	68.20%
40	85.00%	81.70%
60	84.00%	92.40%
80	92.00%	93.50%
100	98.00%	92.30%
120	97.00%	94.00%
140	93.00%	95.60%
160	96.00%	96.00%
180	94.00%	96.10%
200	93.00%	97.30%
220	97.00%	97.40%
240	96.00%	96.20%
260	98.00%	95.80%
280	98.00%	97.30%
300	98.00%	97.50%
320	99.00%	97.80%
340	98.00%	98.20%
360	100.00%	98.90%
380	96.00%	97.00%
400	100.00%	96.40%

changing different steps, that is, by changing different training steps, we can observe the change of model accuracy with the step increases.

As shown in the experimental results in Table 1, we can visualize the changes in accuracy of the training process and the test process in the experiment, so as to obtain intuitive results, as shown in Figures 5 and 6.

Figure 5 shows the variation of training accuracy. It can be seen from the figure that training accuracy increases from 58% at the beginning to 98–100% at the end. At the same time, it can be seen that the training accuracy of the steps has been stable since 80%, which indicates that the training of the data set is good.

Figure 6 shows the variation of test accuracy. From these figures, it can be seen that the test accuracy increases from 68.2% at the beginning to 97–99% at the end. At the same time, it can be seen that the test accuracy of the step has been stable since 60, which is more than 92%, indicating that the test precision of the data set is still very good. Therefore, under the experimental conditions, the overfitting phenomenon that occurred in previous experiments is not observed. For example, there is no similar situation in which the model training is not good enough due to the small number of data sets, which implies a good classification effect for six types of defects. By comparing the convolutional neural network model with the transfer learning model, it can be seen that the training accuracy of the two parts of the experiments without transfer learning can reach more than 90%. However, under the same data set conditions, the training time of the convolutional neural network is significantly longer than that of transfer learning, and the test accuracy is far lower than that of the transfer learning method. Therefore,

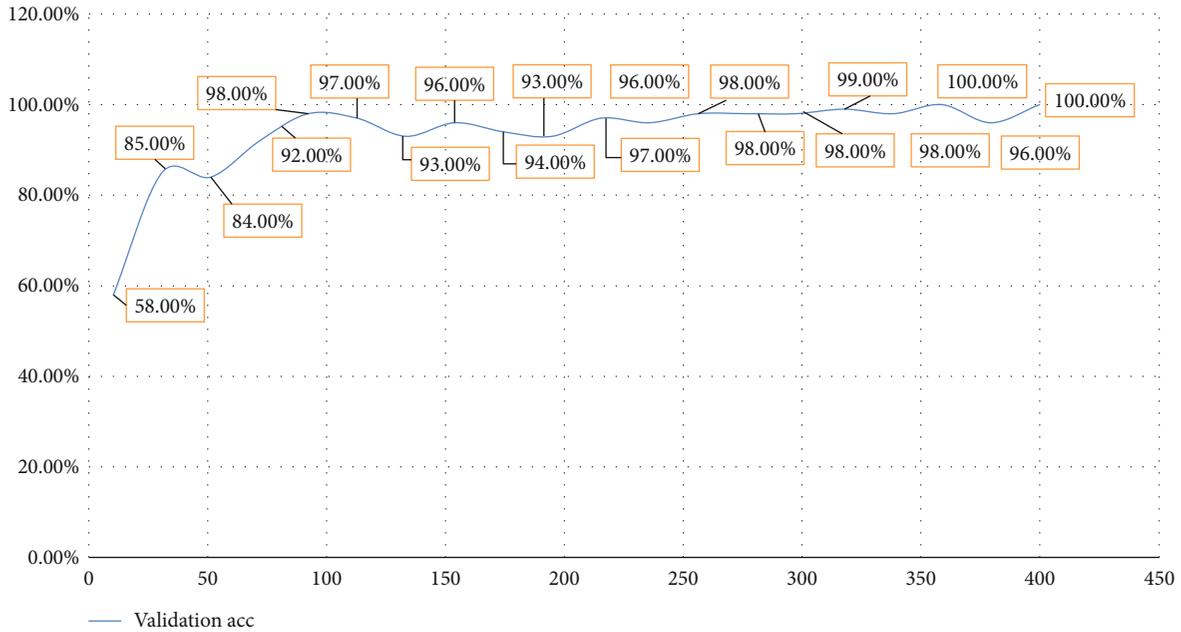


FIGURE 5: Variation chart of training accuracy.

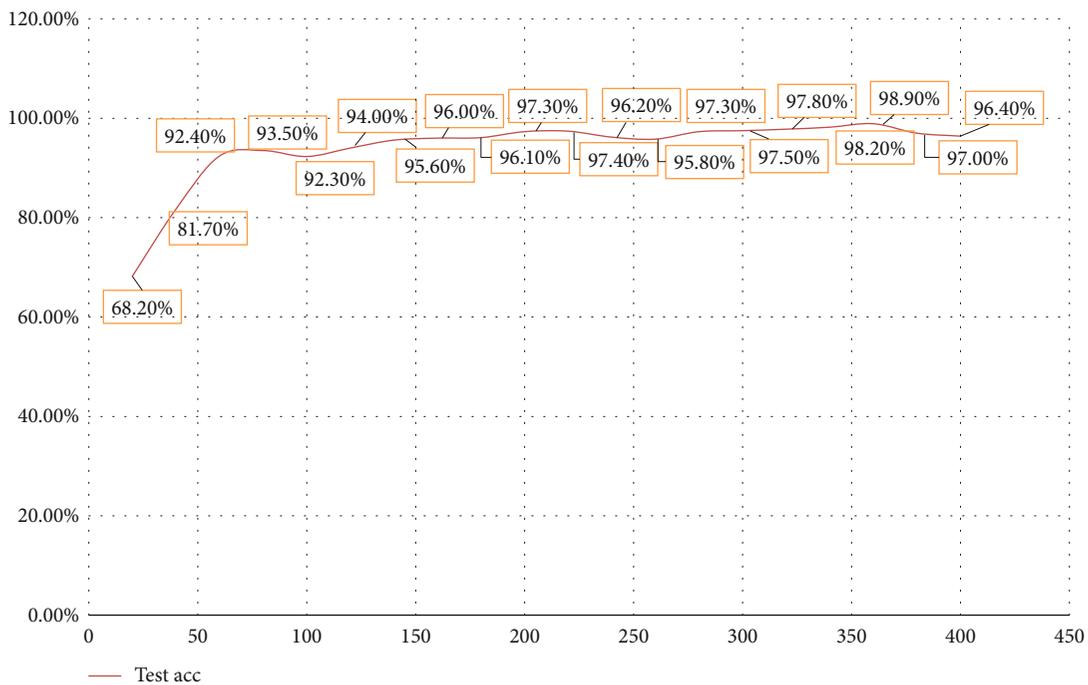


FIGURE 6: Test accuracy variation diagram.

this is an effective method for classifying and predicting the defects in hot-rolled strips by using transfer learning, which is convenient for training and testing, and has high accuracy for small data sets.

4.4. Using FCN to Segment Defects in Steel Strip. First, we use Colab to train the FCN model. Two kinds of steel plates are selected, about 500 pieces in total, and the training set test set is divided in the ratio of 4 : 1. Then, 20000 iterations are

carried out to see the effect of segmentation for scratch and plaque defects.

As shown in Figure 7, for the scratch category, the four images (d)–(g) are iterated 20000 times. Although they are still a little larger than the original figure, the contour has become more accurate. For the (f) scratch image, we can see that the scratch on the far left is very small, and the segmentation of the model is not accurate, and the boundary on the right is relatively fuzzy. Two points can be made here.

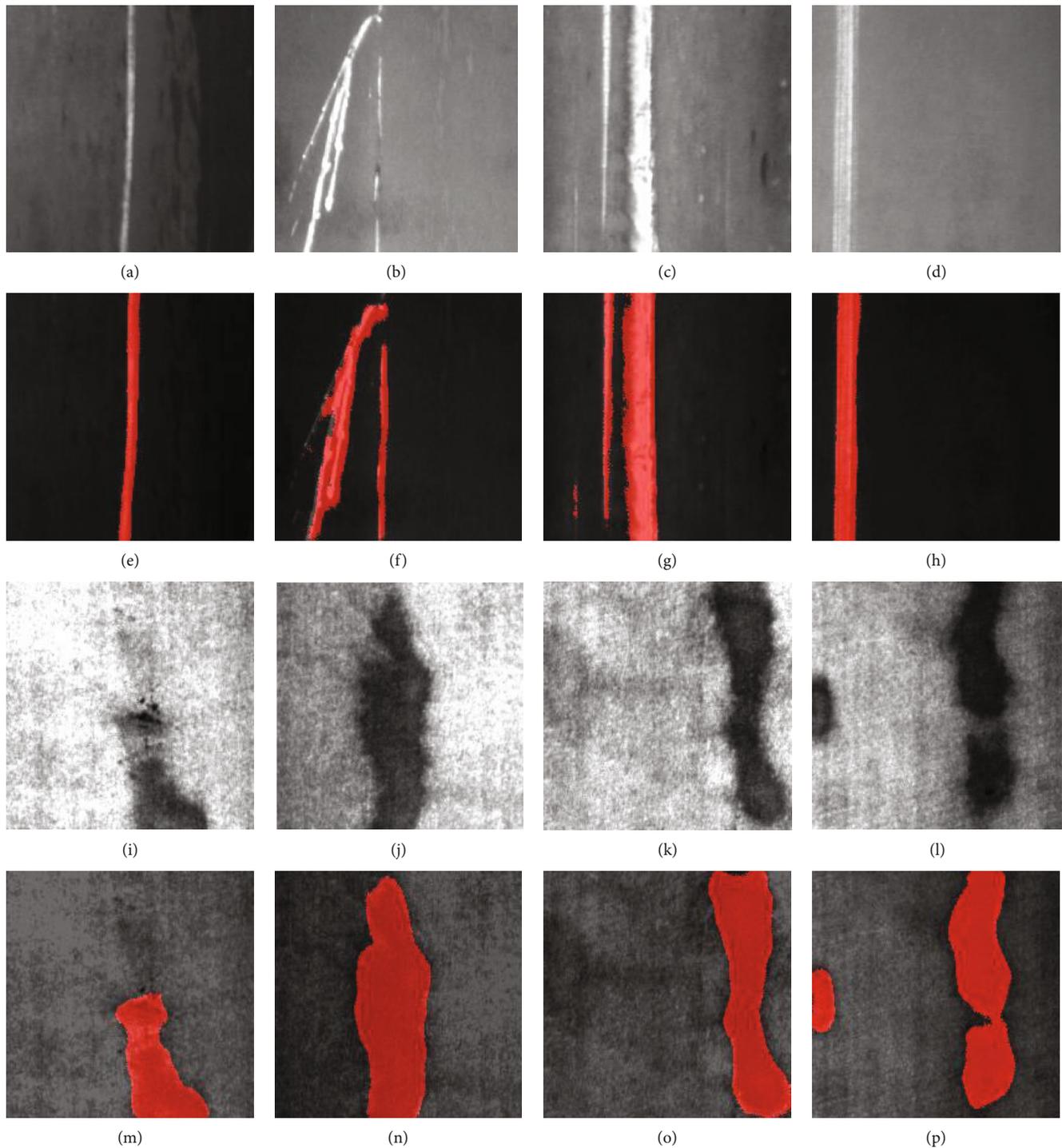


FIGURE 7: Experimental results of defect segmentation.

The first is that the pixel resolution value of the steel plate itself is relatively poor, and the boundary is not obvious, which leads to unsatisfactory segmentation. The second point is that the size of the input image, which leads to a relatively large receptive field and finally makes the image boundary segmentation effect, may not be ideal. Therefore, no matter what model is used, it will have shortcomings and deficiencies.

In the experiment with the steel plate data set model, the segmentation image is still relatively successful, but the steel plate defect is not segmented with clear contours after all, so we do not have high segmentation accuracy for the defective parts of the steel plate. We think that the model segmentation and marking can be carried out through the comparison in the subsequent production data set, so that the model can provide more accurate segmentation.

5. Conclusion

In order to solve the problems of traditional hot-strip defect detection methods, such as slow detection speed, low detection accuracy, and the dependence of parameter-setting on subjective experience, this study applied a convolutional neural network to the detection of surface defects of hot-rolled strips. On this basis, the convolutional neural network model is studied and improved, and a transfer learning model with higher-optimization training accuracy is proposed. The transfer learning model offers a shorter training time, faster convergence speed, and more accurate weight parameters. The experimental results show that the transfer learning model has higher accuracy than the convolutional neural network method in the detection of strip defects. The transfer learning model and FCN defect segmentation method we proposed can not only be applied to strip steel detection but also be easily extended to other surface defect detection areas.

Data Availability

NEU surface defect database: http://faculty.neu.edu.cn/yunhyan/NEU_surface_defect_database.html.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported in part by the China National Key Research and Development project (2017YFE0113200), National Natural Science Foundation of China (No. 51605004), Anhui Provincial Natural Science Foundation (1808085QE162 and 2108085ME166), and Open Project of Anhui Province Key Laboratory of Special and Heavy Load Robot (TZJQR007-2021).

References

- [1] H. Yingli and H. Ying, "Research on an online detection and recognition algorithm for surface defects of strip steel," *Photoelectron Laser*, vol. 26, no. 2, pp. 320–327, 2015.
- [2] Y. He, K. Song, H. Dong, and Y. Yan, "Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network," *Optics and Lasers in Engineering*, vol. 122, pp. 294–302, 2019.
- [3] H. Di, X. Ke, Z. Peng, and Z. Dongdong, "Surface defect classification of steels with a new semi-supervised learning method," *Optics and Lasers in Engineering*, vol. 117, pp. 40–48, 2019.
- [4] C. H. Yong, S. Xu, L. Kun, and S. He-Xu, "Surface defect detection of steel strip based on spectral residual visual saliency," *Optical Precision Engineering*, vol. 24, no. 10, pp. 2572–2580, 2016.
- [5] X. Tao, W. Hou, and D. Xu, "A survey of surface defect detection methods based on deep learning," *Acta Automatica Sinica*, vol. 46, pp. 1–18, 2020.
- [6] C. M. Wang, Y. H. Yan, J. Li, H.-Y. Fu, and G.-C. Sun, "Surface quality detection of cold rolled strip based on BP neural network," *Mechanical Design and Manufacturing*, vol. 6, pp. 106–108, 2007.
- [7] C. Jiahui, Y. Yunhui, and D. Dong, "Application of Gabor filter in strip surface defect detection," *Journal of Northeast University (Natural Science Edition)*, vol. 31, no. 2, pp. 257–260, 2010.
- [8] K. Xu, S. Min, Y. Chaolin, and Z. Peng, "Application of hidden Markov tree model to on-line detection of surface defects [J]," *Journal of Mechanical Engineering*, vol. 49, no. 22, pp. 34–40, 2013.
- [9] L. Wang and G. Shengqi, "Strip surface defect recognition based on deep learning algorithm," *Journal of Xi & Apos; An University of Engineering*, vol. 2017, pp. 669–674, 2017.
- [10] G. Jiachen, L. Gao, and L. Luchi, "Application of target detection algorithm based on deep learning in cold rolling surface defect detection," *Metallurgical Automation*, vol. 43, no. 6, pp. 19–22, 2019.
- [11] Y. Yang and Z. Meng, "Surface defect detection of steel strip based on CNN," *Heavy Machinery*, vol. 2019, pp. 25–29, 2019.
- [12] D. He, K. Xu, and D. Wang, "Design of multi-scale receptive field convolutional neural network for surface inspection of hot rolled steels," *Image and Vision Computing*, vol. 89, pp. 12–20, 2019.
- [13] W. Haiyun, W. Jianping, and L. Fuhua, "Study on surface defect detection of metal sheet and strip using faster R-CNN with multilevel feature," *Mechanical Science and Technology for Aerospace Engineering*, vol. 40, no. 2, pp. 262–269, 2021.
- [14] X. U. Qiang, Z. H. U. Hongjin, F. A. N. Honghui, Z. H. O. U. Hongyan, and Y. U. Guanghui, "Study on detection of steel plate surface defects by improved YOLOv3 network," *CEA*, vol. 56, no. 16, pp. 265–272, 2020.
- [15] J. Chen, M. Liu, Q. S. Fu, and Z. Yao, "Surface defect detection method of hot rolled steel strip based on deep learning," *Automation and Information Engineering*, vol. 40, no. 4, pp. 11–16, 2019.
- [16] F. Zhuang, P. Luo, Q. He, and Z. Shi, "Research progress of transfer learning," *Acta Software Sinica*, vol. 26, no. 1, pp. 26–39, 2015.
- [17] "Image Classification using Transfer Learning in PyTorch," <https://www.learnopencv.com/image-classification-using-transfer-learning-in-pytorch/>.
- [18] J. Černecký, K. Valentová, E. Pivarčiová, and P. Božek, "Ionization impact on the air cleaning efficiency in the interior," *Measurement Science Review*, vol. 15, no. 4, pp. 156–166, 2015.
- [19] P. Božek, "Registration of holographic images based on integral transformation," *Computing & Informatics*, vol. 2013, no. 31, pp. 1369–1383, 2012.
- [20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, Boston, MA, USA, 2015.
- [21] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1493–1504, 2020.