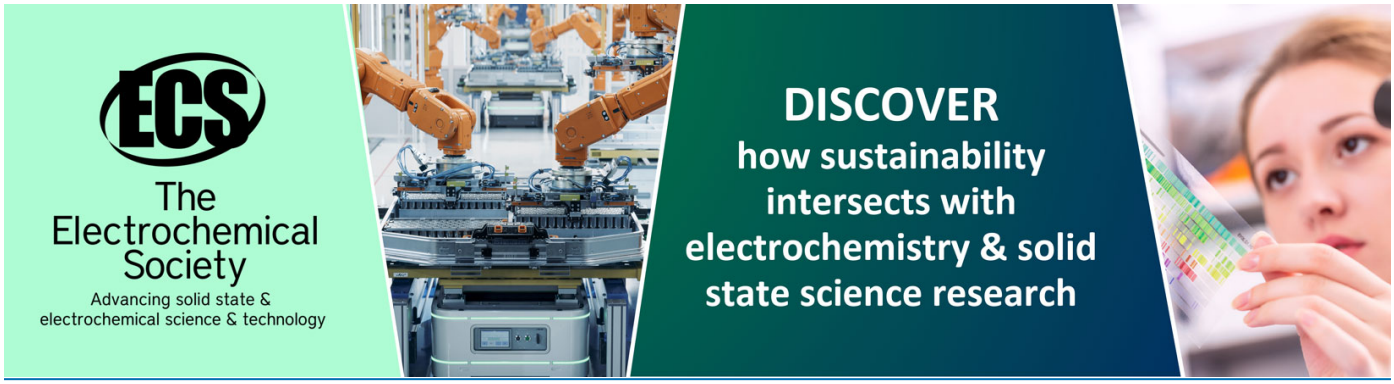
Detection of Surface Defects of Steel Plate Based on ViT



Detection of Surface Defects of Steel Plate Based on ViT

Jiangling Fan*, Xufeng Ling, Jingxin Liang

Shanghai Normal University TIANHUA College, 1661 Shengxin North Road, Jiading District, Shanghai, 201815, China
Email: fanjiangl@163.com

Abstract. A self-attention-based method termed as Vision Transformer (ViT) is applied to efficiently detect the Surface Defects of Steel Plate. The defect image is divided to N*N patches, each of which corresponds to a word, and the whole image data is used as a sentence or paragraph in NPL. A ViT framework is constructed by a learnable module with sequence length of L and 12 multi-head attention layers. We train the proposed model on the surface defects dataset. The experiment results show empirically that ViT has superior performance compared to alternative approaches.
Keywords: Steel Plate Surface Defects, Vision Transformer (ViT), Self-Attention

1. Introduction

Steel plate is an indispensable raw material in industry. Owing to the differences of rolling process, rolling equipment, slabs and other reasons, the surface of steel plate appears various defects during the steel plate producing process. The appearance of steel plate is affected by defects, and the fatigue resistance, wear resistance and corrosion resistance are also reduced. In the production process, online monitoring of steel plate status, timely detection of steel plate surface defects, controlling and improvement of surface quality have always been the goal of industry [1, 2, 3].

Modern industrial production needs intelligent detection with high resolution, high rate and nondestructive. However, the traditional surface defect detection methods cannot meet the requirements. At present, computer vision (CV) technology uses multiple CCD cameras to collect the surface pictures of steel plate on the production line in real time. By comprehensive utilization of modern information technologies such as image processing, pattern recognition, neural network et al., the surface defects detected and identified automatically.

The traditional CV based defect detection methods have advantages in specific problems, but they are not satisfactory in recognition accuracy, robustness and generalization ability. With the development of Artificial Intelligence (AI), the method based on Deep Learning is widely used in defect detection because of its ability of fitting arbitrary complex functions and better feature extraction. Deep Learning is a kind of Deep Neural Network (DNN), which can automatically learn and extract the input data features, and solve the problem of complexity and uncertainty of artificial features extraction in traditional methods. Deep Learning method has stronger recognition accuracy and better generalization ability. Xiao Shuhao et al. applied Transfer Learning of DNN to product surface quality detection, which solved the problem of insufficient training samples and inadequate training calculation of Deep Learning [4]. He Yu, Song Kechen et al. proposed a weakly supervised-learning-based defect detection network for the practical defect detection task of hot rolled steel plate surfaces. The result shows that the method could detect defects with incomplete labels, and obtain the classification error rate of 0. 68% and the localization error rate of 17. 75%, which are

better than the other related methods [5]. Zhang Tao and others conducted a comprehensive study on the development and application of surface defect detection technology based on Deep Learning, compared the advantages and disadvantages of mainstream surface defect detection technology based on Deep Learning, pointed out the existing problems of the technology, and prospected the future development trend [6, 7].

Deep Learning technology has been developing rapidly, and transformer method based on self-attention has caused wide researchers. Transformer was originally mainly used in Natural Language Processing (NLP), which has achieved excellent results. Researchers compare Vision Transformer (ViT) with Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) for Deep Learning, which can achieve good learning effect and greatly reduce the computational resources required for training [8, 9, 10]. In this paper, a ViT based method for defect detection of steel plate surface is designed.

2. Methods

As shown in figure 1, the ViT network is composed of Embedding, Positional Embedding, Transformer module (composed of 12 attention layers connected in series) and Multi-Layer Perception (MLP) act as a classifier.

Firstly, the input image size is 224 × 224, and it is divided into 49 (7 ×7) smaller patches with 32 × 32. Each patch is regarded as a word in NLP. These small patches are concatenated together, which are equivalent to sentences and paragraphs in NLP. Different image sizes correspond to different sentence lengths. The image patches are embedded through a word embedding matrix, and the location information is also embedded through a positional embedding matrix. Then the two embedded vectors are added to form an input sequence. Each patch is transformed into a vector with 384 dimensions.

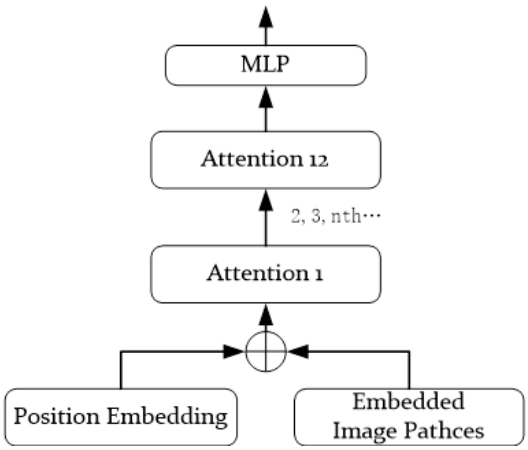


Figure 1. Transformer Architecture.

Secondly, a self-attention based ViT model is constructed, which has a self-attention module with 12 layers in series. Self-attention is the basic unit of ViT, and its main principle is to model the relationship between each input value of input sequence. After learning and synthesizing the overall information of the input sequence, the self-attention module can predict the current input value from multiple other input values. The output self-attention Z is shown in formula (1).

$$Z = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

The input image vector is first transformed into three different vectors: the query vector q, the key vector k and the value vector v with dimension $d_q = d_k = d_v = 384$. Vectors derived from different inputs are then packed together into three different matrices, namely, Q, K and V. It should be noted that the

Q, K and V need to be trained. In order to get multiple complex relationships of different positions in the sequence, multi-head self-attention (MHSA) structure is introduced, and each head corresponds to one relationship. Considering the computational limitation, the embedding dimension is set to 384 and divided into 8 head each of which has 48 dimensions. The multi-head attention module is shown in figure 2.

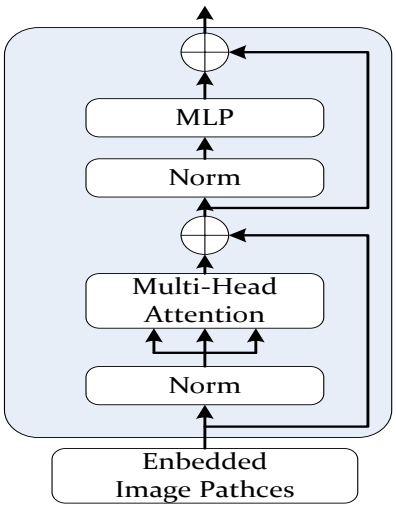


Figure 2. Multi-Head Attention Module.

Next, the self-attention module with 12 layers can complete feature extraction. The low-level self-attention can capture the basic features of the sequence, and the high-level self-attention can get more complex relationships. The last layer is MLP, which mainly completes the classification, fitting and other tasks of downstream.

Finally, ViT trains the image dataset of steel plate surface defects and attains the desired classification effect. As shown in the figure 3.

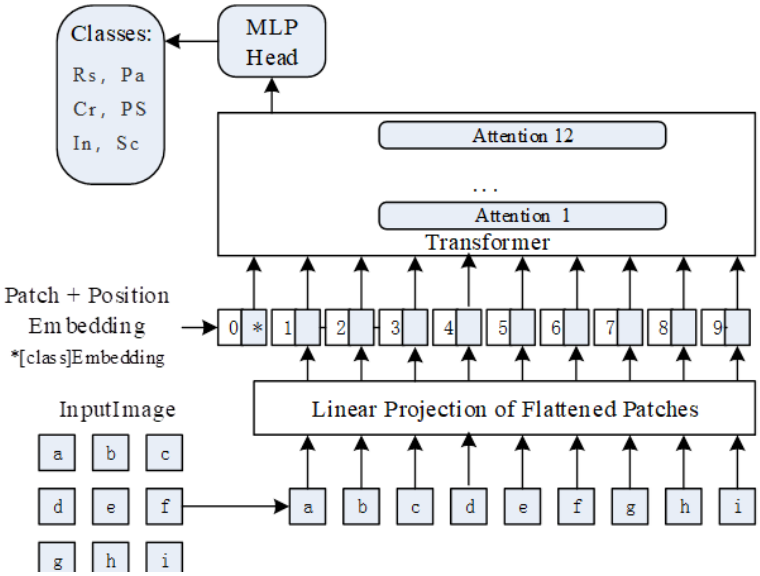


Figure 3. Sketch Map of ViT.

The structure of ViT is exactly the same as traditional transformer, and it needs lots of training. Now we can get ViT pre-trained models, which are trained for the surface defect database. Fine-tuning the pre-trained model for new tasks is mainly through the MLP. The advantage is that the amount of training is small, only a small amount of training calculation is needed, which greatly improves the efficiency.

3. Experiment and Result

The experimental dataset of this paper is based on the surface defect dataset published by Northeastern University. As shown in figure 4 [11] the experimental surface defects concludes six different types, Such as rolled-in scale, patches, crazing, pitting surface, inclusion and scratches. Each type of defect contains 300 samples, and each sample's original resolution is 200×200.

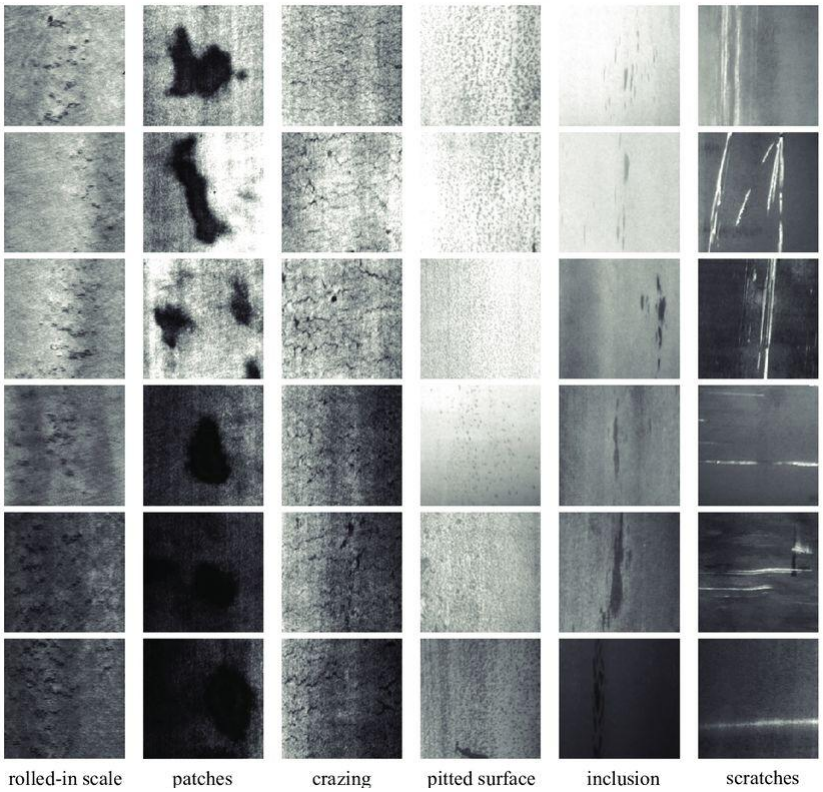


Figure 4. Steel Plate Surface Defects Dataset.

The experimental platform is Pytorch. In the process of system training, the images in training dataset are enhanced by random cutting, deformation, normalization and other means to achieve better training effects. During the initialization for the parameters of the experimental model, Batch size is set to 128, the total number of training rounds (epoch) is set to 500. The optimizer is Adam, and the learning rate parameter is 1e-4. The proportion of training set, verification set and testing set is 60%, 20% and 20% respectively.

The change curve of loss value is shown in figure 5. With the increase of epoch, the loss values of training set and verification set decrease gradually.

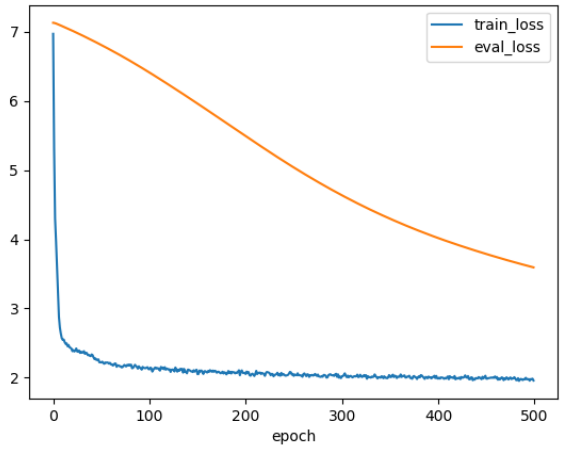


Figure 5. Training loss and verification loss of ViT.

The training set and verification set are used to adjust the hyper parameters of CNN framework for steel plate surface defects detection. When Epoch is set to be 500, the accuracy change curve of training set and verification set is shown in figure 6. With the increase of epoch times, the accuracies of training set and verification set are improved. When the Epoch value is greater than 100, the accuracy of the verification set increases quickly, and the accuracy of training set does not increase significantly. The accuracy of final verification set is 88.93%.

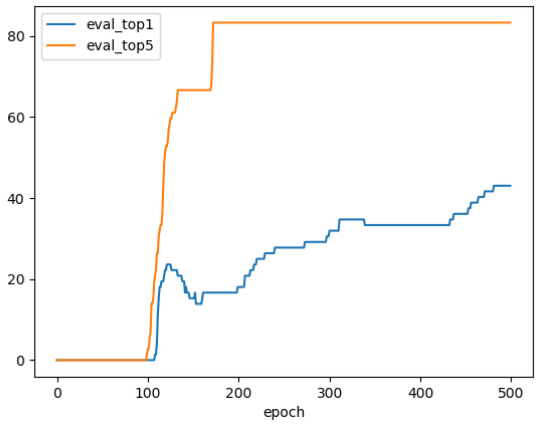


Figure 6. Comparison of training ACC and verification ACC.

4. Experiment

Comparative experiments are made to compare ViT with ResNet18 and ResNet50. The comparative experiment was carried out on the platform of Pytorch. The details are shown in the table 1 below.

Table 1. Performance Comparison.

	Validation ACC	Patch Size	Parameters(M)
ResNet18	86.52%	224	11.47
ResNet50	88.23%	224	23.36
ViT	88.93%	224	20.58

Compared with ResNet model based on CNN, the accuracy of ViT is higher for detection of surface defects on steel plate.

5. Conclusion

In this paper, ViT model is used to train and detect the surface defects dataset of steel plate. The experimental result shows that compared with CNN based ResNet18 and ResNet50, ViT has better defect detection ability. ViT has better foreground in the field of defects detection. The next step is to improve the robustness and generalizability of ViT and make it into an embedded system for practical application.

References

[1] Guo Zh B 2020 *Research on Detection and Classification of Steel Plate Surface Defects* Inner Mongolia University of Science &Technology.
[2] He D 2021 *Application of Deep Learning Method in Detecting Steel Structure Defects and Chars* University of Science and Technology Beijing.
[3] Ou J G 2020 *Research on On-line Detection Algorithm of Steel Plate Surface Defects Based on Deep Convolutional Neural Network* South China University of Technology.
[4] Xiao Sh H, Wu L and He W 2020 Application of deep learning in surface quality detection *Machinery Design & Manufacture* (01): 288-292.
[5] He Y, Song K Ch, Zhang D F, et al. 2021 Weakly-supervised steel plate surface defect detection algorithm by integrating multiple level features *Journal of Northeastern University (Natural Science)* **42**(05): 687-692.
[6] Zhang T, Liu Y T, Yang Y N, et al. 2020 Review of surface defect detection based on computer vision *Science Technology and Engineering* **20**(35): 14366-14376.
[7] Li Sh B, Yang J, Wang Zh, et al. 2019 Review of development and application of defect detection technology *Acta Automatica Sinica* **45**(11): 1-18.
[8] Han K, Wang Y H and Chen H T 2020 A survey on visual transformer arXiv- CS - Computer Vision and Pattern Recognition DOI: arxiv-2012.12556.
[9] Khan S, Naseer M and Hayat M 2021 Transformers in vision: A survey arXiv - CS - Machine Learning DOI: arxiv-2101.01169.
[10] DosoViTskiy A, Beyer L and Kolesnikov A 2020 An image is worth 16x16 words: Transformers for image recognition at scale arXiv - CS - Machine Learning DOI: arxiv-2010.11929.
[11] Song K and Yan Y 2013 A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects *Applied Surface Science* **285**(21): 858-864.