



HAL
open science

Defect Detection and Classification of Strip Steel Based on Improved VIT Model

Lina Xing, Tinghui Li, Honghui Fan, Hongjin Zhu

► **To cite this version:**

Lina Xing, Tinghui Li, Honghui Fan, Hongjin Zhu. Defect Detection and Classification of Strip Steel Based on Improved VIT Model. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.327-335, 10.1007/978-3-031-03948-5_26 . hal-04178731

HAL Id: hal-04178731

<https://inria.hal.science/hal-04178731v1>

Submitted on 8 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Defect Detection and Classification of Strip Steel based on Improved VIT Model

Lina Xing, Tinghui Li, Honghui Fan, Hongjin Zhu

¹ Jiangsu University of Technology, Changzhou 213001, China

² 1801 Zhongwu Avenue, Changzhou, Jiangsu, China

XingLina7512@163.com

Abstract: With the further development of industry, strip steel occupies an important position in industrial production and is widely used in various manufacturing fields. It is especially important to monitor the quality of strip steel production. In order to improve the detection rate of defective strip steel for its complex and varied surface defects and other characteristics, this paper proposes a defect classification algorithm based on the SFN-VIT (Improved Shuffle Network Unite Vision Transformer) model to classify six types of defects in strip steel and compare it with other classification algorithms based on convolutional neural network. The experimental data show that the proposed SFN-VIT model outperforms the traditional machine learning algorithm model and achieves an average accuracy of 91.7% for defect classification on the NEU-CLS dataset (Tohoku University strip steel surface defect categories dataset), which is a 5.1% improvement compared to the traditional classification algorithm.

Keywords: Defect Classification, Strip Steel, Transformer, Machine Vision

1 Introduction

Strip steel is widely used in aerospace, automotive manufacturing, building structures, machinery manufacturing, household appliances and electronic instrumentation production because of its high surface flatness and good mechanical properties [1].

With the increasing requirements of industrial production, the total demand and production quality of strip steel is also increasing [2]. Due to the influence of cold rolling process, production materials, processing equipment and many other technologies, the surface of strip steel will inevitably produce defects such as inclusions, scratches, pressed-in oxide, cracks, pockmarks, patches and so on, which directly affect the corrosion resistance and wear resistance of strip steel, so the detection of surface defects on strip steel is an important part of product quality monitoring. The research on strip steel surface defect detection at home and abroad is mainly based on traditional machine vision and based on deep learning, while in recent years deep learning algorithms gradually show high performance in computer vision tasks beyond traditional machine vision algorithms, and the research on strip steel surface defect detection based on deep learning has become mainstream and achieved better results. Therefore, the SFN-Vision Transformer model is used to detect strip steel surface defects according to their characteristics.

The Vision Transformer model is based on the following considerations for the detection and classification of cold-rolled strip surface defects: Vision Transformer is an image classification algorithm based on the Transformer model, while Transformer lacks inductive preferences such as translation invariance and locality. In other words, Vision Transformer lacks a priori experience that convolutional structures have, therefore it requires more data to learn such structures. When the Vision Transformer model is pre-trained on a large-scale dataset and then migrated to a small-scale dataset, the results achieved are comparable to most outstanding convolutional neural networks and are suitable for small-scale cold-rolled strip defect datasets. To the greatest extent, it avoids the prone to overfitting of convolutional neural networks to small data sets. Furthermore, the strip surface defect dataset has a large variation of intra-class defects, while inter-class defects have many similarities. Moreover, the strip surface defect image will be affected by the image sensor itself and the surrounding environment such as high temperature and bright light, these characteristics bring challenges to the strip surface defect category recognition. We optimize the defect classification algorithm for the above problem as follows: sparse representation technique is used to denoise the strip surface defect images, which enhances the feature extraction effect of the subsequent network model.

2 Related Work

To improve the detection rate of defective strips, many scholars have studied it in recent years and proposed various methods for improving the performance of classification models [3]. Literature [4] proposed a new semi-supervised learning method based on convolutional self-encoder and semi-supervised generative adversarial network, which improved the performance of SGAN with limited training samples as a solution to the situation that strip steel sample images exist mostly unlabeled, and improved the classification rate by about 16% compared with traditional detection methods; literature [5] used YOLOv3 as the feature extraction network, and the DenseNet Block for feature fusion and reuse before predicting the whole layer of features, and detected on six strip steel surface defect datasets with a detection accuracy of 85.7%, achieving good results; literature [6] proposed a tree neural network classifier, consisting of four different neural networks, which divided the defects into specific defects by area type, disk type, and area and linear detail types, and used the SFFS algorithm to select the optimal features for the classifier and achieved better results; literature [7] established a lightweight search algorithm using dominant non-uniform patterns (DNUPs), proposed a mixed-mode code mapping mechanism for encoding, and performed feature extraction in the framework of local binary patterns (GCLBP), and finally used a nearest neighbor classifier to achieve histogram matching, and experimental data showed that better experimental results were achieved; literature [8] first binarized the image data, combined with the original defective image for the extraction of geometric features, grayscale features and shape features, and based on the support vector machine classification model with Gaussian radial basis as the kernel function, and determined the model parameters through cross-validation, the experimental classification accuracy was better than that of the traditional BP neural network-based classification model.

3 Research Methods

In this paper, the strip steel surface defect detection model is improved based on the Vision Transformer model. The image data are denoised using a sparse representation technique before being fed into the classification model, and the VIT (Vision Transformer) model is improved using a modified ShuffleNetv2 network as a feature extraction network.

3.1 De-Noise Processing

The images acquired during the actual production of strip steel can generate noise due to environmental factors such as the image sensor itself, high temperature, and bright light [9]. Noise can interfere with the classification results and affect the effect of feature extraction, so it is essential to perform noise reduction on the images before performing the detection [10]. The sparse representation technique is introduced to denoise cold rolled strip images. While removing the image noise, the noise is not a sparse component of the signal, it can minimize the damage to the original image information [11]. The sparse representation technique mainly consists of two parts: sparse coding and dictionary learning. In other words, the dictionary atoms that can linearly represent the image data are generated in the data set, thus forming an overcomplete dictionary that can extract the multidimensional features of the image, and the sparse vector is obtained by sparse coding of the image with the trained overcomplete dictionary, while the noise is not linearly represented by the overcomplete dictionary and cannot be sparse, so the noise is discarded in the form of residuals in the reconstruction process.

In this paper, KSVD algorithm is used for denoising, and the sparse representation model is used to optimize the minimum error by fixing the number of sparse coefficients. K-means and SVD algorithms are integrated in KSVD algorithm, and after determining the number of iterations K , each iteration is decomposed using SVD algorithm, and in the process of updating the dictionary, only one atom in the dictionary and its corresponding sparse coding vector are updated each time, the other atoms remain unchanged. The other atoms are kept unchanged, and the sparse coding factor is also updated when all the atoms in the dictionary are updated to complete this iterative process. The algorithm steps are as follows:

- (1) Randomly select n atoms in the strip defect dataset to initialize the learning dictionary.
- (2) Based on the determined sparse representation model, the sparse encoding matrix X is calculated using the OMP algorithm.
- (3) Optimize the error matrix using the SVD algorithm as in Equation (1) and update the atoms in the dictionary.
- (4) Determine whether all the atoms in the dictionary have been updated, if not continue to step (3); if all the atoms in the dictionary have been updated, determine whether the number of iterations has been reached, if not continue to step (2); if the

number of iterations has been reached, get the updated dictionary, use the updated dictionary to calculate the sparse encoding and reconstruct the image with steel defects, and then perform weighted superposition on the image blocks to get the denoised image.

$$E_k = Y - \sum_{j \neq k} d_j x_T^j \quad (1)$$

3.2 SFN-Vision Transformer

Convolutional neural network models have been the dominant model in computer vision since the explosion of deep learning [12]. Convolution has a natural advantage in image problems by virtue of equivalence and localization, but by continuously stacking convolutional layers to obtain large perceptual fields, while Transformer has a long memory length and can parallelize processing [13][14]. Vision Transformer applies Transformer to image classification tasks [15]. VIT model consists of three main modules, Linear Projection of Flattened Patches, Transformer Encoder, and MLP Head. The Transformer Encoder module requires a vector matrix as input, thus mapping patches divided into images into one-dimensional vectors by linear mapping, adding category vectors and position encoding, feeding them into the Transformer Encoder module of the repetitive stack Encoder Block module L times, and getting classification results according to the corresponding data generated by extracting category vectors.

The VIT model will be optimized in the following two aspects: before reshaping the image data into a flattened two-dimensional fragment sequence, feature extraction is performed on the image data, and we replace the feature extraction network with a lighter ShuffleNetv2 network and improve the ShuffleNetv2 model. Combining the SE unit with Shuffle Block, we add the SE module on the shortcut branch of Shuffle Block. That is, a squeezing operation is first performed on it by global average pooling, compress the channel feature information into weight factors, normalize them with two fully connected layers, and finally multiply the normalized weight factors with the weighted input features to achieve weight rescaling. The structure of Shuffle Block with a step of 1 is shown in Fig.1. By adding SE unit modules to suppress the feature channel information on the Shuffle Block shortcut branch that has little effect on the current task, strengthening important feature information, and further enhancing the performance of the network model. The optimized SFN-Vision-Transformer network model is shown in Fig.2.

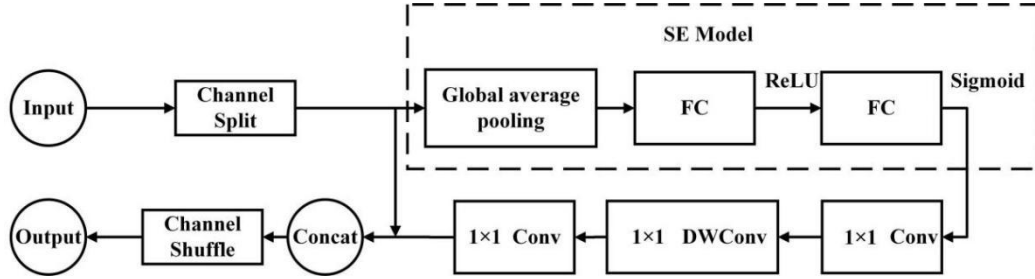


Fig. 1. Shuffle Block structure diagram

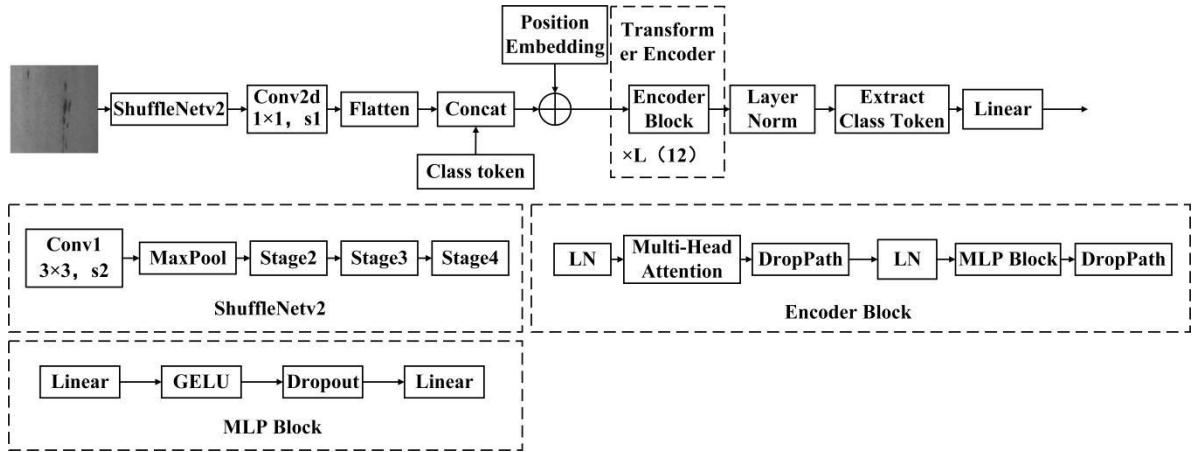


Fig. 2. SFN-Vision Transformer network structure diagram

4 Experimental Results and Analysis

4.1 Data Set

The strip surface defect dataset used in this paper is a NEU-CLS image file for defect classification task disclosed by Tohoku University, containing six typical strip surface defects, namely, inclusions, scratches, pressed-in oxide, cracks, pockmarks and patches, each containing 300 images, each with the size of 299×299 , Fig.3 shows the sample images of six strip defects. Due to the lack of positive samples in the NEU-CLS dataset, 300 images of qualified strip steel were obtained as positive samples by the camera with the size of 299×299 .

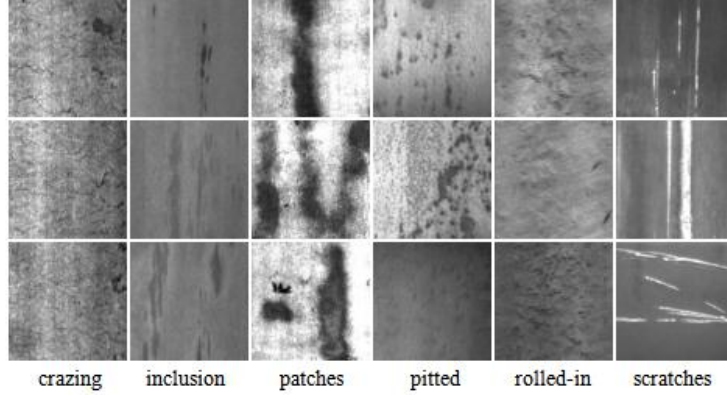


Fig. 3. Strip surface defect map

4.2 Experimental Setup and Evaluation Indexes

Experimental Environment and Model Parameters. The experiment is conducted based on the following platform and environment: code implementation using Python, specific experimental environment configuration is: processor Intel Core i5-10400, 16G memory devices, operating system Win10 Professional 64-bit.

In order to reduce the number of parameters of the model, the Vision Transformer model is selected as ViT-Base, and the Patch Size is 32×32 , The time of repeatedly stacked Encoder Blocks in Transformer Encoder is 12, the dimension of each token sequence after passing through the Embedding layer is 768, the number of nodes of the first fully connected MLP Block in Transformer Encoder is 3072, and the number of heads in Transformer The number of heads of Multi-Head Attention is 12.

Evaluation Indicators. Since the strip surface defect dataset used in this experiment contains six defects, which is a multi-classification task, the multi-classification problem is transformed into multiple binary classification problems for evaluation, specifically containing Accuracy and Recall rate of each defect, where Accuracy is the proportion of the number of correctly classified samples to the total number of samples, as in Equation (2), and Recall rate is the proportion of all positive classes that are predicted to be positive, as in Equation (3) [16].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Among them, TP (True Positive) refers to the number of positive classes judged as positive; FN (False Negative) refers to the number of positive classes misjudged as negative; FP (False Positive) refers to the number of negative classes misjudged as positive; TN (True Negative) refers to the number of negative classes judged as negative [17].

4.3 Experimental Results

Since the VIT model can get better results on large datasets, thus the proposed model was pre-trained on ImageNet dataset, and after obtaining the pre-training parameters, the training was carried out using the strip surface defect dataset, and the tuning of the model parameters was completed. the final partial results of strip defect classification on the proposed model are shown in Fig.4:

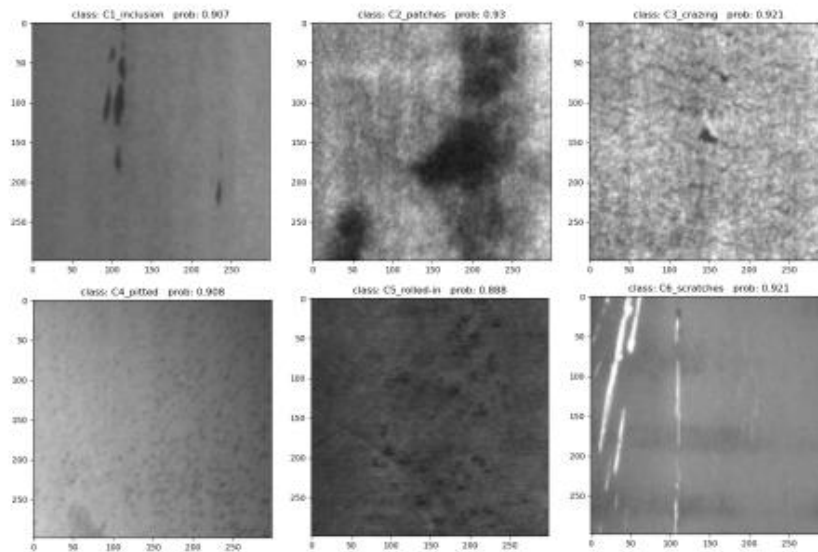


Fig. 4. Some strip steel surface defects detection results

The experiments were conducted in the context of a dataset containing six types of defective images, 300 images of each defective and qualified image strips, and the dataset was divided into a training set and a validation set in a 2:1 ratio, and the average accuracy reached 91.8% and the average recall rate reached 89.2% after 500

iterations with batch-size set as 5, as shown in Table 1.

Table 1. Various types of defects detection and evaluation table

Type of defect	Accuracy/%	Recall rate/%
Crazing	92.1	90.4
Inclusion	91.6	88.7
Patches	92.9	90.8
Pitted	91.7	89.7
Rolled-in	89.8	85.3
scratches	92.4	90.2

The proposed algorithm is compared with other classification model algorithms based on the average accuracy and average recall rate evaluation metrics, and the results are shown in Fig.5. After pre-training on large ImageNet datasets, the generalization ability of the Vision Transformer model is greatly improved, large-scale data trained on Vision Transformer can outperform the inherent inductive bias that convolutional neural networks have, such as localization and translational covariance, and the required computational resources are also significantly reduced, and its detection capability is better compared with the convolutional neural network model. The SFN-VIT model based on the combination of convolutional neural network and Transformer can be calculated in parallel using Transformer and the number of operations required to calculate the correlation between different positions does not increase with the increase of distance, which can produce the advantages of a more explanatory model. At the same time, the use of convolution characteristics can have the advantage of better expressiveness in structured data such as images, so as to learn image features more efficiently, and its detection results are better than those of the convolution network model and the improved Vision Transformer Model.

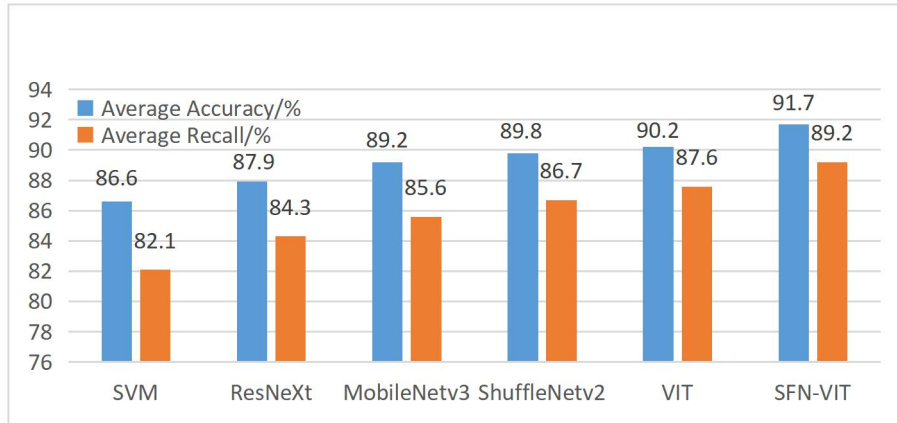


Fig. 5. Comparison chart of classification models

The traditional industrial methods for detecting defects in strip steel are eddy current detection, infrared detection and magnetic leakage detection. Eddy current inspection method uses eddy current detector on the surface of the strip to observe the changes in the induced current to determine the presence or absence of defects. However, using this method for inspection requires a long period of heating to process the strip to be inspected into a uniform temperature field, so the method has great limitations and is very difficult to achieve for cold rolled strip. Infrared detection method uses electric current through the surface of the detected object to produce different temperatures, resulting in infrared differences to achieve defect detection. When passing through the defective area, a difference in temperature and infrared light is generated that is different from the normal area. But the method is not adapted to large depth defects, and limited types of defects can be detected. Leakage detection method uses the change in magnetic flux density to detect defects, and this technology can detect the size of defects, but the method can only detect ferromagnetic materials and the influence of the surrounding environment is large. There are limitations in the detection occasions. The machine vision-based inspection method can solve the above problems, so the improved VIT model of strip steel surface defect detection method proposed in this paper has some practical significance.

5 Summary

In this paper, the classification of strip steel surface defects was performed based on

the NEU-CLS dataset. By improving the Vision Transformer model, pre-training on ImageNet data, and optimizing the parameters of the network model, a classification model with high accuracy is obtained. It can be seen that using the attention mechanism can make the model pay more attention to the features of the defective part, and at the same time, noise reduction is performed on the image to achieve the purpose of improving the classification accuracy. Many improvements need to be made in the future, such as the introduction of BackBone for feature extraction to the VIT model, which will increase the number of parameters of the network model, and further optimization of the algorithm, which will be completed in future work.

Acknowledgements

This work was supported by the National Natural Science Foundation of China(61806088), by the Natural science fund for colleges and universities in Jiangsu Province(20KJA520007), by Graduate Practice Innovation Project Fund for Jiangsu university of Technology(XSJCX21_51).

References

1. Tang Ming, Li Yuanyuan, Yao Wei, Hou Lingyu, Sun Qichun, Chen Jiahang. A strip steel surface defect detection method based on attention mechanism and multi-scale maxpooling[J]. Measurement Science and Technology, 2021, 32(11):
2. Wenyan Wang, Kun Lu, Ziheng Wu, Hongming Long, Jun Zhang, Peng Chen, Bing Wang. Surface Defects Classification of Hot Rolled Strip Based on Improved Convolutional Neural Network:Instrumentation, Control and System Engineering[J]. ISIJ International, 2021, 61(5):
3. Support Vector Machines; New Support Vector Machines Study Findings Have Been Reported from University of Science and Technology (Multi-class Classification Method for Strip Steel Surface Defects Based On Support Vector Machine With Adjustable Hyper-sphere)[J]. Journal of Robotics & Machine Learning, 2019:
4. He Di, Xu Ke, Zhou Peng, Zhou Dongdong. Surface defect classification of steels with a new semi-supervised learning method[J]. Optics and Lasers in Engineering, 2019, 117:
5. Jiaqiao Z . Surface defect detection of steel strips based on classification priority YOLOv3-dense network[J]. Ironmaking & Steelmaking, 2020.
6. Moon C I, Choi S H, Kim G B, et al. Classification of Surface Defects on Cold Rolled

- Strip by Tree-Structured Neural Networks[J]. Transactions of the Korean Society of Mechanical Engineers A, 2007, 31(6): 651-658.
7. Luo Q , Sun Y , Li P , et al. Generalized Completed Local Binary Patterns for Time-Efficient Steel Surface Defect Classification[J]. IEEE Transactions on Instrumentation and Measurement, 2018,PP:1-13.
 8. Hu H, Li Y, Liu M, et al. Classification of defects in steel strip surface based on multiclass support vector machine[J]. Multimedia tools and applications, 2014, 69(1): 199-216.
 9. Wan Xiang, Zhang Xiangyu, Liu Lilan. An Improved VGG19 Transfer Learning Strip Steel Surface Defect Recognition Deep Neural Network Based on Few Samples and Imbalanced Datasets[J]. Applied Sciences, 2021,11(6):
 10. Dongyan, Cui, Kewen, Xia, Aslam, Nelofar, Jingzhong, Hou. Defect Classification Recognition on Strip Steel Surface Using Second-Order Cone Programming-Relevance Vector Machine Algorithm[J]. Journal of Computational and Theoretical Nanoscience, 2016, 13(9):
 11. Classification of Surface Defects of Hot Rolled Strips: Effectiveness of SVM over HOG and GLCM Combined Features[J]. International Journal of Innovative Technology and Exploring Engineering, 2020, 9(3):
 12. Engineering; Researchers at Central South University Target Engineering (Surface Defect Classification for Hot-rolled Steel Strips By Selectively Dominant Local Binary Patterns)[J]. Journal of Engineering, 2019:
 13. Chao WANG, Yu-ting LIU, Ya-ning YANG, Xiang-yu XU, Tao ZHANG. Research on Classification of Surface Defects of Hot-rolled Steel Strip Based on Deep Learning[C]//Proceedings of 2019 2nd International Conference on Informatics, Control and Automation (ICA 2019), 2019:375-379.
 14. Science - Science and Engineering; Investigators at University of Putra Malaysia Report Findings in Science and Engineering (Surface Defects Classification of Hot-rolled Steel Strips Using Multi-directional Shearlet Features)[J]. Journal of Engineering, 2019:
 15. Mohammed Waleed Ashour, Fatimah Khalid, Alfian Abdul Halin, Lili Nurliyana Abdullah, Samy Hassan Darwish. Surface Defects Classification of Hot-Rolled Steel Strips Using Multi-directional Shearlet Features[J]. Arabian Journal for Science and Engineering, 2019,44(4):
 16. Qiwu Luo, Xiaoxin Fang, Yichuang Sun, Li Liu, Jiaqiu Ai, Chunhua Yang, Oluyomi Simpson. Surface Defect Classification for Hot-Rolled Steel Strips by Selectively Dominant Local Binary Patterns.[J]. IEEE Access, 2019,7:
 17. Kolisnik Brendan, Hogan Isaac, Zulkernine Farhana. Condition-CNN: A hierarchical

multi-label fashion image classification model[J]. Expert Systems With Applications, 2021, 182: