

Research on aluminum defect classification algorithm based on deep learning with attention mechanism

Wen Zhang, Shibao Sun, Huanjing Yang

College of Information Engineering, Henan University of Science and Technology, Luoyang 471000, China

Abstract: Product quality is an important indicator for determining the quality of industrial products. Defects on the surface of aluminum profiles are inevitably caused in the actual production process due to the influence of various factors such as environment and equipment, and these defects seriously affect the quality of aluminum profiles. The focus and difficulty of research have shifted to how to quickly and accurately identify and classify surface defects in aluminum profiles. To address this issue, this paper proposes an aluminum defect classification algorithm that uses an attention mechanism in conjunction with the traditional Inception V4 network model for deep learning image classification, to accurately identify and classify aluminum defect areas. Experiments and comparative analysis are performed on the aluminum defect recognition dataset from the Alias Tianchi platform, and the results show that the algorithm with the addition of the attention mechanism improves accuracy by 1.24% over the original model.

Keywords: Deep learning; Defect classification; Attention mechanism; Inception V4.

1. Introduction

With the development of the industrial manufacturing field in the world, China has become a major manufacturing country with a large number of industrial products to be produced every day. With the increasing demand, the quality of industrial products is becoming more and more demanding, however, no matter the process of production or transportation, there is no way to avoid the defects on the surface of the products [1]. For example, due to the complexity of the production environment and the limitations of the processing equipment, the surface of aluminum profiles, as a basic material in industrial manufacturing, is prone to cracks, abrasions, peeling, pits, scratches, miscellaneous colors, dirty spots, and other defects during production and transportation, which can seriously affect the quality of aluminum profiles [2]. Therefore, it is important to automate the classification and identification of defects in aluminum profiles.

The existing defect recognition classification methods are the manual recognition method, single mechanism recognition method, infrared recognition method, magnetic particle recognition method, eddy current recognition method, magnetic leakage recognition method, machine vision recognition method [3-7], and other seven methods. However, the first six methods mentioned above have the shortcomings of low efficiency and accuracy due to the limitation of the principle. Among the methods of machine learning recognition, such as support vector machines and decision trees are based on manual to feature extraction and classification of defects, which can only learn some low-level features and cannot learn detailed abstract high-level features [8], and the accuracy rate is low in the recognition and classification of subtle defects.

With the development of deep learning, convolutional neural networks (CNNs) are widely used in image classification, target detection [9-10], speech recognition, and intelligent robotics [11], etc. Convolutional neural networks (CNNs) extract features differently from manual extraction, but learn features and extract them from the data itself, with

better results and greater robustness. Shanshan Xu et al [12] implemented wood defect recognition classification by convolutional neural network and this method does not require complex pre-processing of images and can recognize many kinds of wood defects with high correct rate and efficiency. Liu, Meng Ke et al [13] achieved the classification of track surface defects recognition by the convolutional neural network, which solved the problem that traditional machine vision recognition technology relies on manual experience. Zhiyang Wu et al [14] proposed a convolutional neural network-based monochrome fabric defect recognition algorithm for the problem of high leakage rate and low efficiency of manual recognition of fabric defects in fabric production enterprises, and the experimental results showed that this method could achieve high accuracy and speed. tian Wang et al. [15] proposed a convolutional neural network to automatically extract features to distinguish between defect-free and defective images for product quality control. Gui Zhong Fu et al. [16] proposed a deep learning-based approach through a compact and effective convolutional neural network model that emphasizes the training of low-level features and incorporates multiple sensory fields.

However, due to the uneven distribution of surface defects of aluminum profiles and the fact that aluminum profile defects belong to small target information, the defect classification using only the above-mentioned deep convolutional neural network method is easy to ignore some detailed information, especially when some defects on the surface of aluminum profiles are small and easy to classify inaccurately, which leads to poor results.

Therefore, this paper focuses on the current practical problem of classifying aluminum surface defects by selecting the classical deep learning network classification model Inception V4 and adding the attention mechanism scSE module to its model for fusion, which makes the network model learning easier to capture the local key information on feature map channels and space and suppress irrelevant information, to further improve the accuracy of the model in classifying aluminum defects.

2. Related Work

2.1. Inception V4 Network

For convolutional neural networks, an effective way to obtain efficient network performance is to increase the breadth and depth of the network, but the blind increase will lead to a dramatic increase in network parameters, which can easily cause overfitting [17]. GoogLeNet [18] Inception V1 was proposed by the Google team in 2014 and won first place in that year's ImageNet competition classification task won the first place. He was the first to propose the Inception structure based on LeNet-5. The core of Inception is to use convolutional kernels of different sizes, which makes the existence of different sizes of perceptual fields, and finally, achieve the fusion of different scale features by stitching. The effect is to increase the depth and width of the network while also reducing the number of parameters. Different versions of Inception V2-V4 [19-21] have emerged based on this network with continuous optimization. In this paper, we use the Inception V4 network as the main architecture, and the network principle of Inception V4 is described in detail below. The network is composed of a Stem module, Inception-A module, Inception-B module, Inception-C module, Reduction-A module, and Reduction-B module. Each module is highly tunable. The width and depth of the network are increased without increasing the network parameters, thus increasing the accuracy and not overfitting. The network structure is shown in Figure 1.

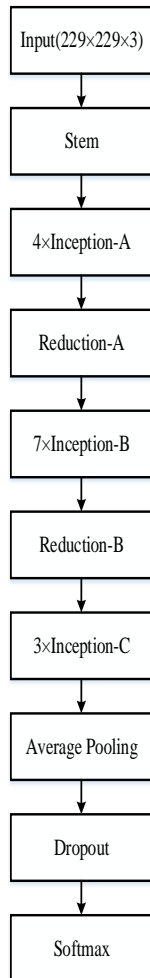


Figure 1. Inception V4 structure diagram

2.2. SCSE Attention Mechanism

The attention mechanism mimics the study of human brain vision, which has access to a large amount of information from the outside world at anytime and anywhere, and from a large amount of information, quickly locates relatively important information and ignores irrelevant information [22]. Therefore, it is important to introduce the attention mechanism into deep learning networks to achieve similar functions in defect recognition classification.

The scSE module is based on the evolution of the SE module in SENet [23]. three variants of the SE module of ROYAG et al [24], namely, the cSE channel compression spatial excitation module, the sSE spatial compression channel excitation module, and the scSE module formed by combining the cSE and sSE modules in parallel. And it is experimentally demonstrated that such a module can enhance meaningful features and suppress useless features.

The main idea proposed by the cSE module is shown in Figure 2, Firstly the input feature map $U=[u_1, u_2, \dots, u_c]$, each channel $u_i (u_i \in \mathbb{R}^{H \times W})$, U obtains the vector $z (z \in \mathbb{R}^{H \times W})$ after passing through the global pooling layer (GAP). The value at the k -th channel can be expressed as:

$$z_k = \frac{1}{H \times W} \sum_i^H \sum_j^W u_k(i, j) \quad (1)$$

where H, W is the size of the feature map, C is the number of channels, (i, j) is the coordinate on the feature map. Then the vector is fully connected twice, W_1, W_2 are the weights of the fully connected layer, and then the process of Relu function enhances the independence between each channel and Sigmoid normalization to obtain $\sigma(\hat{z})$, where the \hat{z} value can be expressed as:

$$\hat{z} = W_1(\delta(W_2 z)) \quad (2)$$

Finally, it is multiplied with the unprocessed feature information of the original channel to obtain the calibrated feature map. In this way, the information within the unimportant channels will be reduced and suppressed, while the information within the important channels will remain almost unchanged and the disguised phase is enhanced. The whole process can be expressed as:

$$\hat{U}_{cSE} = F_{cSE}(U) = [\sigma(\hat{z}_1)u_1, \sigma(\hat{z}_2)u_2, \dots, \sigma(\hat{z}_c)u_c] \quad (3)$$

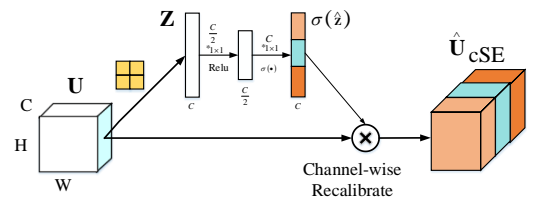


Figure 2. Channel attention mechanism

(2) The main idea of sSE module is shown in Figure 3, sSE is a variant on the basis of cSE, cSE to improve the ability of important channel feature information of the network by compressing spatial information, then, in turn, it can also compress channel information to improve the ability of important spatial feature information of the network, so sSE was born.

For the input feature map $U=[u_1, u_1, u_1, 2, \dots, u_i, j, \dots, u_H, W]$ ($u_i, j \in \mathbb{R}^{1 \times 1 \times c}$), H, W are the dimensions of the feature map respectively, (i, j) is the spatial location of the feature map, the vector q is obtained by squeezing the feature information of the space through a 1×1 convolution with a channel number of 1. The vector q can be expressed as:

$$q = W_{sq} * U \quad (4)$$

Where $W_{sq} \in \mathbb{R}^{1*1*c*1}$, The obtained feature map q is then normalized by the sigmoid function to obtain (\cdot) , which corresponds to the spatial feature information of the (i,j) pixel points in the feature map respectively, and finally weighted with the input original feature map. The whole process can be expressed as:

$$\hat{U}_{sSE} = F_{sSE}(U) = [\sigma(q_{1,1})u^{1,1}, \dots, \sigma(q_{i,j})u^{i,j}, \dots, \sigma(q_{H,W})u^{H,W}] \quad (5)$$

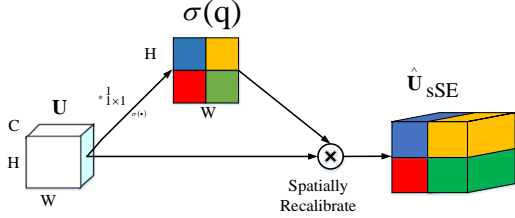


Figure 3. Spatial attention mechanism (sSE)

(3) The main idea proposed by the scSE module is shown in Figure 4, which is a tandem of the cSE and sSE modules, and the feature extraction is carried out separately by both the channel and space of the feature map, and then the weighted summation is obtained after the feature map information obtained is more specific and targeted, making the final extracted feature information also more focused. Its formula can be expressed as:

$$U_{scSE} = U_{cSE} + U_{sSE} \quad (6)$$

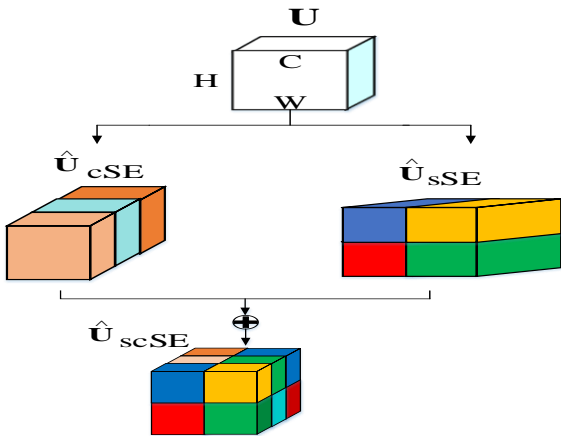


Figure 4. Channel-space attention mechanism (scSE)

3. Method

The difficulty of classifying aluminum defects lies in the inconsistent shape of the surface defects and the difference in size. For the surface defects of aluminum are very small and some small scratches, this paper incorporates the attention mechanism scSE module in the module of Inception V4 network model to pay more attention to these regions of interest, extract more important feature information for learning and then improve the accuracy of the model classification. The structure diagram of the improved algorithm is shown in Figure 5.

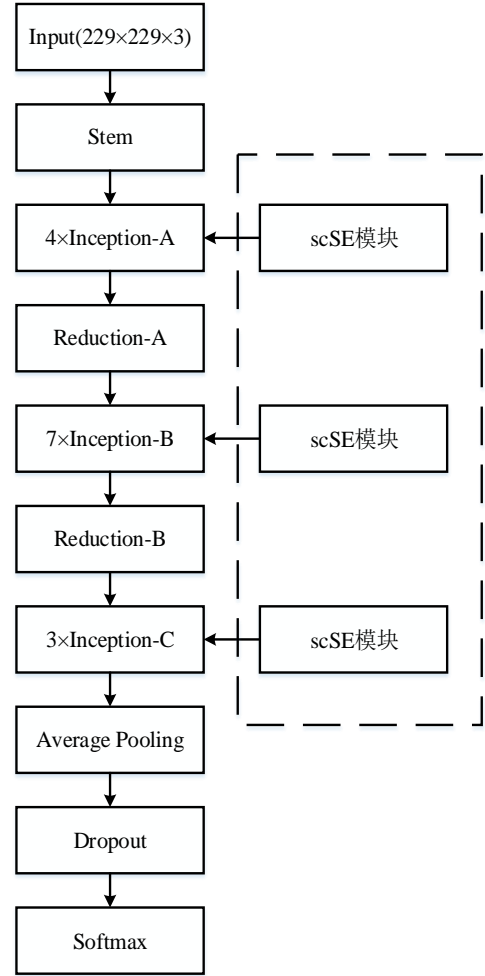


Figure 5. Inception V4-scSE structure diagram

Inception V4 investigates two modules, the Inception Module and Reduction Module, the attention module scSE is added to Inception-A, Inception-B, and Inception-C modules respectively for fusion, but the number of parameters thus generated does not affect the network training time and the real-time demand for classification of aluminum defects. The Inception Module is added to increase the width and depth of the network to obtain more feature information, and the Reduction Module is added to reduce the computational effort. The Reduction Module was added to reduce the computational effort.

The algorithm in this paper uses the cross-entropy loss function commonly used in classification problems in the network training process, Its formula can be expressed as:

$$L = -\frac{1}{N} \sum_{i=1}^N (y_{(i)} \log \hat{y}_{(i)} + (1 - y_{(i)}) \log (1 - \hat{y}_{(i)})) \quad (7)$$

Where N denotes the size of the number of samples, and i denotes the i -th sample, $y_{(i)}$ is the true value, $\hat{y}_{(i)}$ is the predicted value.

4. Experimental Analysis

4.1. Datasets

This paper uses a dataset from the 2018 Guangdong Industrial Smart Manufacturing Big Data Innovation Competition - Aluminum Profile Surface Defect Recognition dataset, which is hosted by AliCloud Tianchi. The dataset contains no defect samples and ten categories of defect samples (divided into non-conductive, scuffed, cross-strip pressure dent, orange peel, bottom leakage, bruise, pit, convex powder, cracked coating, and dirty spot). The

following are images of the ten categories of defect samples, respectively, as shown in Figure 6.



Figure 6. Pictures of aluminum defect categories

4.2. Evaluation Criteria

The experiments in this paper belong to a classification task, so the common criteria of classification, Accuracy, Precision and Recall, are used to evaluate the performance of the algorithm. First of all, the concepts of TP, TN, FP, and FN are concerned. TP is predicted as a positive sample and predicted correctly, TN is predicted as a negative sample and predicted correctly, FP is classified (predicted) as the actual negative sample as a positive sample, and FN is classified (predicted) as the actual positive sample as a negative sample.

Table 1. Classification confusion

Actual Sample	Prediction Sample	
	Positive Sample	Negative Sample
Positive Sample	TP(True Positives)	FN(False Negatives)
Negative Sample	FP(False Positives)	TN(True Negatives)

1)Accuracy

Accuracy is the most commonly used evaluation metric in classification, and it is the proportion of all correctly predicted samples (both positive and negative classes) to the total.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

2)Precision

The precision rate is the percentage of all predicted positive classes that are truly positive out of all predicted positive classes, and is specific to our prediction results.

$$P_{precision} = \frac{TP}{TP + FP} \quad (9)$$

3)Recall

The recall is the proportion of all predicted positive classes that are actually positive out of the total number of positive classes.

$$R_{recall} = \frac{TP}{TP + FN} \quad (10)$$

4.3. Results and Analysis

This experiment is based on Windows operating system, trained on Pytorch deep learning framework, using GeForce GTX1650Ti GPU. the size of all images in the experiment is set to 400×400, the epoch in the experiment is set to 50, the stochastic gradient descent algorithm is used, and the initial learning rate is 0.001.

To verify the performance of the algorithm proposed in this paper, the improved algorithm Inception V4-scSE is compared and analyzed with the network model InceptionV4 and the two most commonly used deep learning network models Xception [25] and shufflenet-v2 [26] on the aluminum defect recognition dataset by these three criterias accuracy, precision and recall.

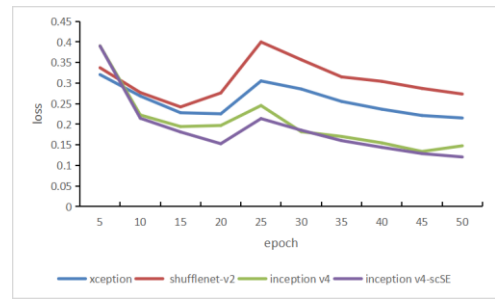


Figure 7. Comparison of loss curves different model validation
Loss curves

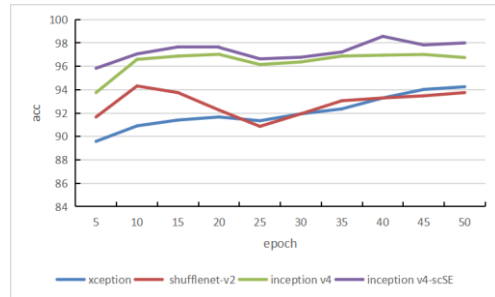


Figure 8. Comparison of acc curves for different model validation

As shown in Figure 7 and Figure 8, the curves of loss function and accuracy of the proposed algorithm model in this paper are compared with other network models at the time of validation. From the graphs, it can be seen that the lowest loss value of the proposed algorithm Inception V4-scSE is around 0.1, and the accuracy rate is around 98%. The next lowest loss value of Inception is around 0.15, and the accuracy is around 96%. It can also be seen from the figure that the loss values and accuracy rates of these four algorithms gradually level off after the 25th epoch.

Table 2. Comparison of different evaluation criterias under different network models

Method	Accuracy	Precision	Recall
Xception	94.25	91.7	92.5
Shufflenet-V2	93.75	85.7	85.4
Inception V4	96.76	92.0	94.4
Inception V4-scSE	98.00	97.8	96.0

As can be seen from Table 2, the Inception V4-scSE network model proposed in this paper has 1.24% higher accuracy, 5.8% higher precision, and 1.6% higher recall than the original model Inception v4. And it is also compared with other deep learning network models Xception and Shufflenet-V2, and both the accuracy, precision and recall rates are significantly better than the performance of these network models.

5. Conclusion

In this paper, we propose a deep learning aluminum defect classification algorithm incorporating an attention mechanism to address the practical problem of identifying and classifying defects generated by aluminum in the production process and various factors. The accuracy of classifying aluminum defects is improved by adding an attention mechanism module to the deep neural network model Inception V4 to enhance the sensitivity of the network to small defects. Experiments are also conducted on the aluminum defect recognition dataset from Ali Tianchi to

demonstrate the effectiveness of the algorithm.

However, this algorithm only classifies a single defect on an image, and in reality, there are often multiple defects on a single image, so in-depth research is needed for this multi-label defect classification task in the future, making it possible to truly solve the practical problems that exist in reality and improve the practicality and robustness of the algorithm.

References

- [1] Xu Qianxin, et al. "Improved YOLOv3 network for surface defect detection of steel plates." *Computer Engineering and Applications* 56.16(2020):265-272.
- [2] Duan, Chunmei, and Taochuan Zhang. "Two-stream convolutional neural network based on gradient image for aluminum profile surface defects classification and recognition." *IEEE Access* 8 (2020): 172152-172165.
- [3] Gupta, Ranjeetkumar, et al. "A Review of Sensing Technologies for Non-Destructive Evaluation of Structural Composite Materials." *Journal of Composites Science* 5.12 (2021): 319.
- [4] Tu Zhenyue. "Application of magnetic particle inspection in boiler inspection." *Applied Energy Technology* 11(2019):3.
- [5] Huang, Feng-Ying. "Quantitative assessment method for eddy current detection of cracks on rail surface." *China Railway Science* 38.2 (2017): 6.
- [6] Ege, Yavuz, and Mustafa Coramik. "A new measurement system using magnetic flux leakage method in pipeline inspection." *Measurement* 123 (2018): 163-174.
- [7] Duan, Chunmei, and Taochuan Zhang. "Two-stream convolutional neural network based on gradient image for aluminum profile surface defects classification and recognition." *IEEE Access* 8 (2020): 172152-172165.
- [8] Saeed, Faisal, et al. "A robust approach for industrial small-object detection using an improved faster regional convolutional neural network." *Scientific Reports* 11.1 (2021): 1-13.
- [9] Li, Zewen, et al. "A survey of convolutional neural networks: analysis, applications, and prospects." *IEEE transactions on neural networks and learning systems* (2021).
- [10] Yang, Jing, et al. "Real-time recognition method for 0.8 cm darning needles and KR22 bearings based on convolution neural networks and data increase." *Applied Sciences* 8.10 (2018): 1857.
- [11] Yang, Guanci, et al. "Convolutional neural network-based embarrassing situation detection under camera for social robot in smart homes." *Sensors* 18.5 (2018): 1530.
- [12] Shanshan Xu, Yingang Liu, and Sheng Xu. "Wood defect recognition based on convolutional neural network." *Journal of Shandong University (Engineering Edition)* 43.2 (2013): 23-28.
- [13] Liu, Meng Ke, Wu, Yang, and Wang, Xun. "Implementation of convolutional neural network based track surface defect detection technology." *Modern Computer: Professional Edition* 000.029(2017):65-69.
- [14] Wu, Zhiyang et al. "A fast detection algorithm for monochrome fabric defects based on convolutional neural networks." *Journal of Computer-Aided Design and Graphics* 30.12 (2018): 9.
- [15] Wang, Tian, et al. "A fast and robust convolutional neural network-based defect detection model in product quality control." *The International Journal of Advanced Manufacturing Technology* 94.9 (2018): 3465-3471.
- [16] Fu, Guizhong, et al. "A deep-learning-based approach for fast and robust steel surface defects classification." *Optics and Lasers in Engineering* 121 (2019): 397-405.
- [17] Wang D, Cai B-B, and Zai C-F. "An Inception-V4-based method for car parking status detection." *Computer Age* 003(2022):000.
- [18] Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- [19] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning*. PMLR, 2015.
- [20] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [21] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." *Thirty-first AAAI conference on artificial intelligence*. 2017.
- [22] Li, Yubo, et al. "Application of YOLOv5 Based on Attention Mechanism and Receptive Field in Identifying Defects of Thangka Images." *IEEE Access* 10 (2022): 81597-81611.
- [23] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [24] Roy, Abhijit Guha, Nassir Navab, and Christian Wachinger. "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks." *International conference on medical image computing and computer-assisted intervention*. Springer, Cham, 2018.
- [25] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [26] Ma, Ningning, et al. "Shufflenet v2: Practical guidelines for efficient cnn architecture design." *Proceedings of the European conference on computer vision (ECCV)*. 2018.