华南理工大学

South China University of Technology

《统计学习与数据科学》 课 程 论 文

(2023-2024 学年 暑假 学期)

Defect recognition and localisation model based on improved Alexnet-SVM and Fast RCNN

学生姓名：蔡永瑄

提交日期： 2024 年 08 月 28 日 学生签名：蔡永瑄

学　　号	2002130320038	座位编号	
学　　院	数学学院	专业班级	信息管理与信息系统
课程名称	统计学习与数据科学课程设计	任课教师	夏立
教师评语：			
本论文成绩评定： _____分			

Defect recognition and localisation model based on improved Alexnet-SVM and Fast RCNN

Yonghuang Cai

School of Mathematics, South China University of Technology, Guangzhou510630
2967787845@qq.com

Abstract

In this study, a metal surface defect recognition and localization model based on improved AlexNet-SVM and Fast R-CNN is proposed. Metal surface defect detection is crucial to product quality, and the application of lightweight deep learning methods in this field has attracted increasing attention. Firstly, through image preprocessing and data enhancement, we use the improved AlexNet network to extract metal surface features, and achieve high accuracy defect recognition by global average pooling and SSA optimized SVM classifier, with an accuracy of 88.79%, while keeping the model lightweight, and the storage space is only 10.23MB. The FLOPs is 63.23MB. Furthermore, combining the Faster R-CNN model and Region Proposal Network, we realize the pixel-level automatic defect localization, and the recognition accuracy is as high as 95.23%.

Keywords: Alexnet, SSA_SVM, Fast RCNN, Defect recognition, Lightweight

1 Introduction

In modern industrial production, the quality of metal and plastic products is directly related to the market competitiveness of products and the use experience of consumers. Small defects in the surface of a product, although imperceptible, can cause irreversible damage to the performance and durability of the product. With the development of automation technology, automatic surface defect detection technology has gradually become an important and challenging field in manufacturing [1].

Traditional surface defect detection methods, such as those based on template matching or edge detection, are often limited by environmental factors and the diversity of defect types, which are difficult to adapt to the complex and changing industrial environment. In recent years, the rise of deep learning technology provides a new solution for automatic defect detection. Domen Tabernik, Maticoc-Suc and Danijel Skooc-caj and other researchers established a surface defect detection model based on deep learning [2], which demonstrated the recognition ability of high accuracy even with a small amount of training data. However, existing models still face some challenges in practical applications, especially when deployed on resource-constrained devices[3].

This study aims to develop a lightweight defect identification and localization model to meet the needs of efficient and accurate surface defect inspection on low-cost handheld devices with limited storage space and computing power. We first conduct an in-depth analysis of metal surface features and construct a network model consisting of AlexNet and SSA-SVM, which improves the model's ability to recognize defects under different lighting and background conditions through image preprocessing and data augmentation techniques. Furthermore, we adopt the Faster R-CNN model combined with the Region Proposal Network to achieve the precise location of the defect area,

which greatly improves the accuracy and practicability of the detection.

After the introduction section of this study, we will organize the rest of the paper in the following structure. In Section 2, we detail our methodology, including the design of the model, the ensemble strategy of the improved AlexNet-SVM and Fast R-CNN. Section 3 shows the experimental validation process we conducted on the Kolektor defective product dataset, including a detailed description of data preprocessing, model training, and test results. Finally, in Section 4, we summarize the main findings of this paper, discuss the advantages and potential applications of our approach, and present directions for future work.

2 Methodology

2.1 Classification

The network in this paper consists of a metal surface feature extractor and a metal surface classifier. To determine whether the metal representation is defective or not, this is obviously a binary classification problem, for the Kolektor image dataset, we use AlexNet lightweight deep learning network as the basis, use SPP to improve the size and number of kernels of AlexNet, and go through the migration learning to extract the metal surface target feature matrix, and at the same time, combine with the SVM binary classification model, under the sparrow search algorithm SSA optimisation to obtain the best recognition parameters of SVM, and finally get a recognition network with AlexNet-SPP+SSA-SVM to achieve lightweight metal surface defect detection.

2.1.1 AlexNet model

(1) Convolutional layer

Suppose there are K convolutional kernels, where the capital letter C denotes the total number of channels in the input, and $x^c(i,j)$ denotes the region in the ith row and jth column of the cth channel of the input. Then the kth output of the layer can be represented as

$$y_k(i,j) = \sum_{c=1}^{c} f_k^{c*} x^c(i,j) + b_k \tag{1}$$

(2) Activation layer

The activation layer improves the learnability of the network and enriches the network structure, where the activation functions in Alexnet are all ReLU functions, which can be expressed as

$$y_k(i,j) = \begin{cases} x_k(i,j), x_k(i,j) \geqslant 0 \\ 0, x_k(i,j) < 0 \end{cases} \tag{2}$$

The shape of the ReLU function is

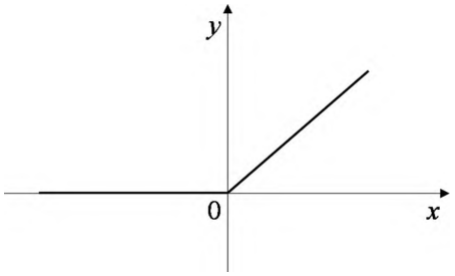

Figure 1: ReLU function

(3) Ponding layer

The role of the pooling layer is to reduce the size of the extracted feature matrix, and if pooling is not taken, the problem of learning too large a parameter occurs.The pooling of the Alexnet network are all maximal pooling, which can be expressed as

$$y_k(I,J) = max\big(x_k(I+i-1, J+j-1)\big) \tag{3}$$

(4)Linear layer

The linear layer, also known as the fully connected layer, serves to flatten the multi-dimensional array obtained by convolution into one-dimensional data, reducing the data dimensionality.

AlexNet is an improved model of CNN network architecture, which innovatively uses ReLU as the activation function to solve the gradient dispersion problem of the original function Sigmoid, and at the same time replaces the average pooling with the maximum pooling to overcome the blurring effect brought by the average pooling. Finally, the AlexNet network is faster in training and extracting feature information with less storage space.

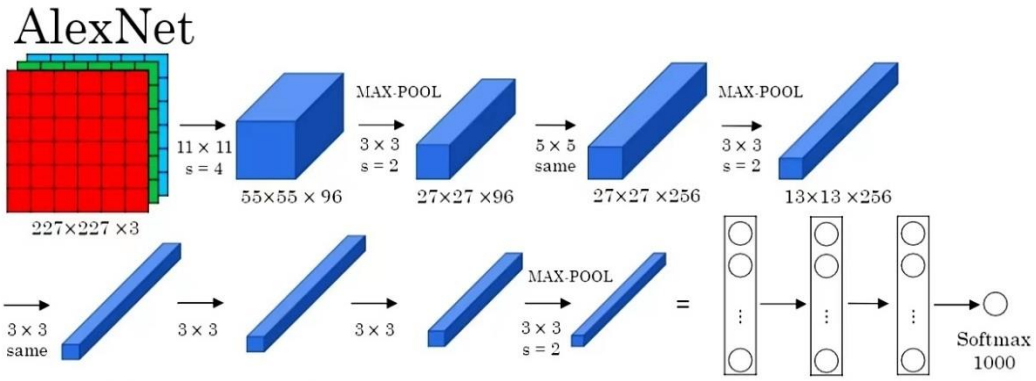

Figure 2: Alexnet network structure

Where CL1 to CL5 are convolutional layers and FCL6 to FCL8 are fully connected layers. The model image input is 227×227×3, and the number of neurons in the fully connected layers FCL6 to FCL7 is 4096.Assuming K convolutional kernels, the kth output of the layer can be expressed as

$$N = \frac{W - F + 2P}{S} + 1 \tag{4}$$

Where W is the input image size, F is the size of the convolution kernel or pooling kernel, P is the number of pixels , and S is the step size.

2.1.2 SSA-SVM model

The classification performance of Support Vector Machine (SVM) is greatly affected by the selection of its own parameters. Combining the Sparrow Search Algorithm (SSA) with the optimisation of the penalty parameter (C) and kernel parameter (g) of the SVM can improve the classification performance of the SVM, and accelerate the classification of the presence of defects on the metal surface, so the SSA-SVM model is constructed by combining the extracted features of

the metal surface.

(1) SSA

During each iteration, the location update of the discoverer is described as

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} exp\left(-\frac{i}{\alpha \mathrm{iter}_{max}}\right), & R_2 < \mathrm{ST} \\ X_{i,j}^{t} + Q\mathrm{L}, & R_2 \geqslant \mathrm{ST} \end{cases} \tag{5}$$

where t is the current number of iterations; $iter_{max}$ is the maximum number of iterations; $X_{i,j}^{t}$ is the position information of the ith sparrow in the jth dimension, $\alpha \in (0,1]$ is a random number; R_2 and ST denote the warning value and the safety value respectively, $R_2 \in [0,1]$, $ST \in [0.5,1]$; Q is a random number obeying normal distribution; L is a $1d$ matrix where each element within the matrix is all $1. ST \in [0.5,1]; Q$ is a normally distributed random number; L is a $1d$ matrix, where each element of the matrix is 1

The follower's location update is described as

$$X_{i,j}^{t+1} = \begin{cases} Q\exp\left(\frac{X_{worst}^{t} - X_{i,j}^{t}}{i^2}\right), & i > \frac{n}{2} \\ X_{P}^{t+1} + \left|X_{i,j}^{t} - X_{P}^{t+1}\right| A^{+} \cdot L, & others \end{cases} \tag{6}$$

where X_{p} is the optimal position currently occupied by the discoverer, and $X_{w}orst$ denotes the current global worst position; and A is a $1d$ matrix where each element is randomly assigned a value of 1 or -1,and $A^{+} = A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}$,where A^{+} is the pseudo-inverse matrix.

(2) SVM

SVM has excellent classification results for small-sample data classification problems, aiming at finding an optimal hyperplane that maximises the distance between the nearest partition of different data and the hyperplane. Meanwhile, it has good generalisation ability and robustness by mapping the input space where the sample points are located to the high-dimensional feature space using the principle of Structural Risk Minimisation (SRM) instead of Empirical Risk Minimisation (ERM)[4].

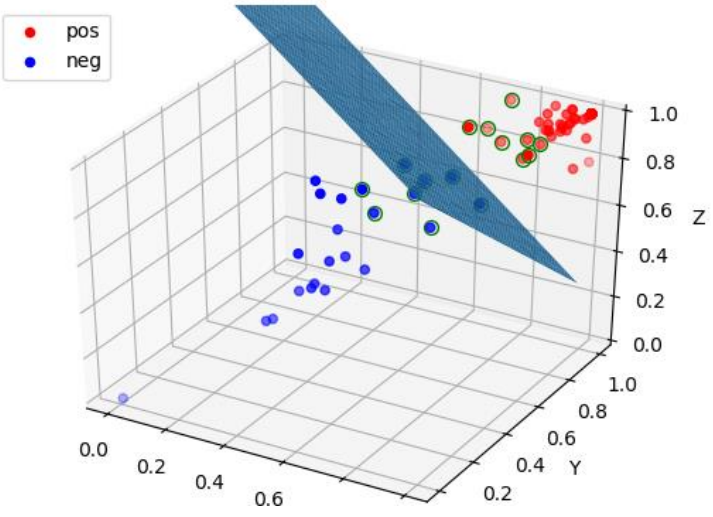

Figure 3: SVM

Assume a training dataset Q on a feature space, where x_i is the i-th feature vector, also known as a sample,y_i is the class labelling of $x_i, x_i \in R^n, y_i \in \{-1,1\}, i = 1,2,\ldots,N$.

$$Q = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\} \tag{7}$$

The learning objective of SVM is to find a separating hyperplane in the feature space that can classify instances into different classes, the separating hyperplane (w,b) corresponds to the equation

$$w \cdot x + b = 0 \tag{8}$$

where w is a normal vector of the hyperplane; b is the intercept, which is used to determine the specific location of the hyperplane. For a given training dataset Q and hyperplane (w,b), define the function interval of the hyperplane with respect to the sample points as

$$\xi_i = y_i(w \cdot x_i + b) \tag{9}$$

At last the problem of finding the optimal hyperplane is transformed into an optimisation problem:

$$\min_{w,b} \frac{1}{2} \| w \|^2 \tag{10}$$

Since it is strictly specified that all the sample points are correctly distributed on both sides, it is called hard interval classification. However, since the data may be nonlinear or there may be outliers interfering with it, hard partitioning is not easy to be achieved in general, so to avoid this problem, we introduce soft interval classification, which uses the hyper-parameter C (penalty coefficient) to control the complexity of the SVM model and the fault-tolerant ability of the SVM model, i.e., it is as large as possible and avoids the interval violation.

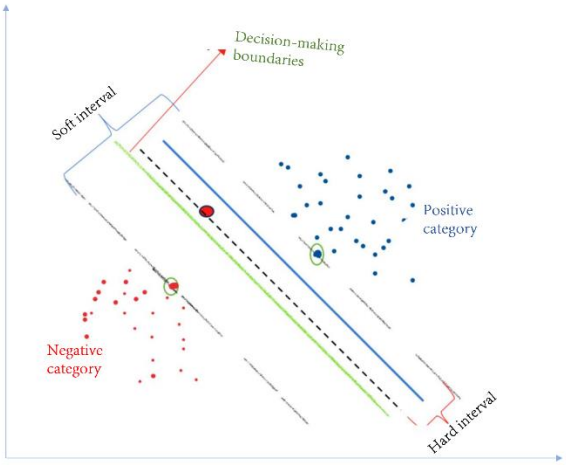

Figure 4: SVM-Soft Interval

The final mathematical expression of SVM can be expressed by the following formula, where x is the feature vector, y is the category label, w and b are the parameters of hyperplane, C is the penalty coefficient, ε is the slack variable, and φ is the kernel function:

$$\min_{w,b,\xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i \tag{11}$$

$$s.t. y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \ldots, n \tag{12}$$

2.2 Auto-Tagging

In order to automatically mark the location or area where surface defects occur, we adopt the Faster R-CNN algorithm model, use image segmentation techniques, and combine the defect features and their classification obtained in Section 2.1 to localize the defect location at the pixel level.

2.2.1 R-CNN model

Faster R-CNN uses Resnet network for feature extraction, followed by region generation by constructing a region proposal network RPN to extract defective candidate target rectangular regions, and then Fast R-CNN is used to detect defects based on the candidate regions.

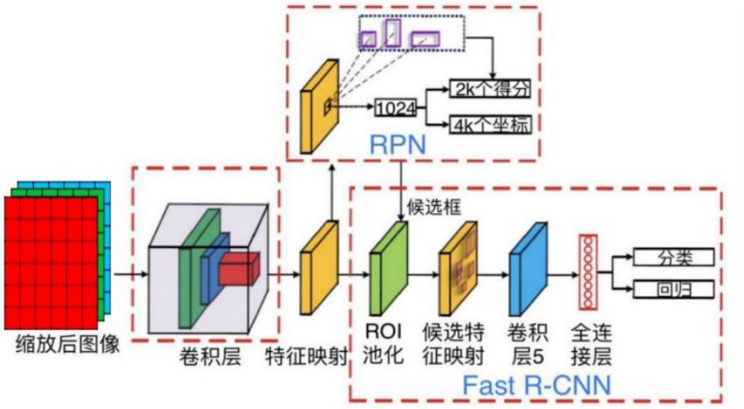

Figure 5: Faster R-CNN network structure

(1) Region Proposal Network

Since the feature extraction of defective regions has been solved in the previous problem, we directly use RPN to extract high-quality candidate frames.The role of RPN network is to input an image and output a batch of rectangular candidate regions.We invert the feature matrix of defective images obtained from Problem 1 and output it as a feature map, and use a small moving window (3×3 convolution kernel) for local scanning on the final convolutional feature map. A D-dimensional vector is obtained by this sliding convolution operation. Subsequently, this D-dimensional vector is fed into two fully connected layers including a location information regression layer and a classification layer to obtain detailed location P(x, y, w, h)[8] and classification information, where x and y are the coordinates of the centre of the candidate box, w is the width and h is the length.

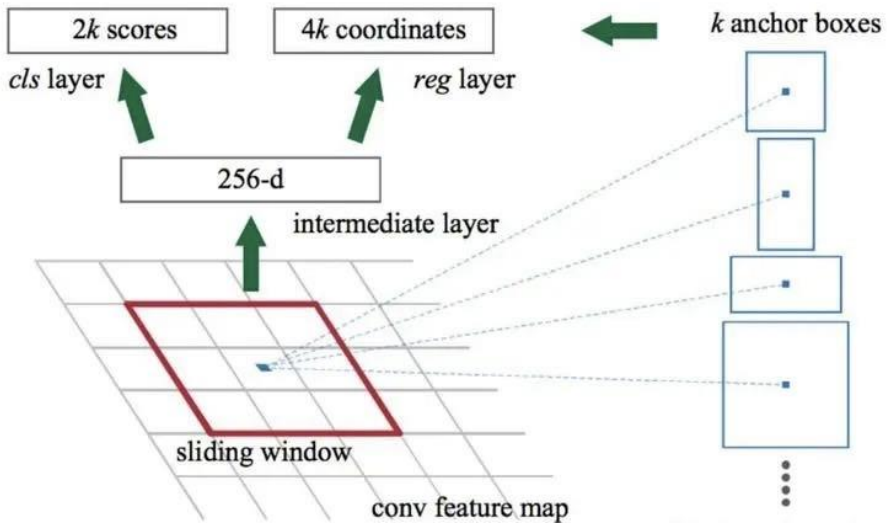

Figure 6: RPN network structure

(2) Faster R-CNN Defect Recognition

After the high-quality defect candidate region of RPN is obtained, Faster R-CNN starts target detection and identification to determine whether there is a target defect region in the candidate region, and we use the merging IOU to determine:

$$IOU = \frac{A \cap B}{A \cup B} \tag{13}$$

Based on this definition of interaction ratio, we define the following rule:

$$\begin{cases} ExistingObjectives, max_i IOU(A, B_j) or IOU(A, B_j) > 0.7 \\ contexts, IOU(A, B_j) < 0.3 \\ neutrality, 0.3 < IOU(A, B_j) < 0.7 \end{cases} \tag{14}$$

where A is the frame, B_i is the ith target region, IOU(A, B_j) is the concurrency ratio of the anchor frame A to the target region B_j, and $max_i IOU(A, B_j)$ is the maximum of the concurrency ratios of the anchor frame A to all target regions.

3 Experiments and Results

3.1 Classification

3.1.1 Image pre-processing

Image preprocessing is to eliminate the disturbing factors of the image and enhance the relevant information. Here, we uniformly reduce the size of the images in the kolektor defective dataset to 227×227 pixels to meet Alexnet's requirements for the input images, and at the same time, transform the background colours of all the images before training to achieve the data enhancement, and output several samples with different background brightness and darkness levels from the input images to cope with the actual recognition process of shadows or exposure situations[5].

The image format conversion is as follows, converting the original image format 500×1260 to 227×227 format:

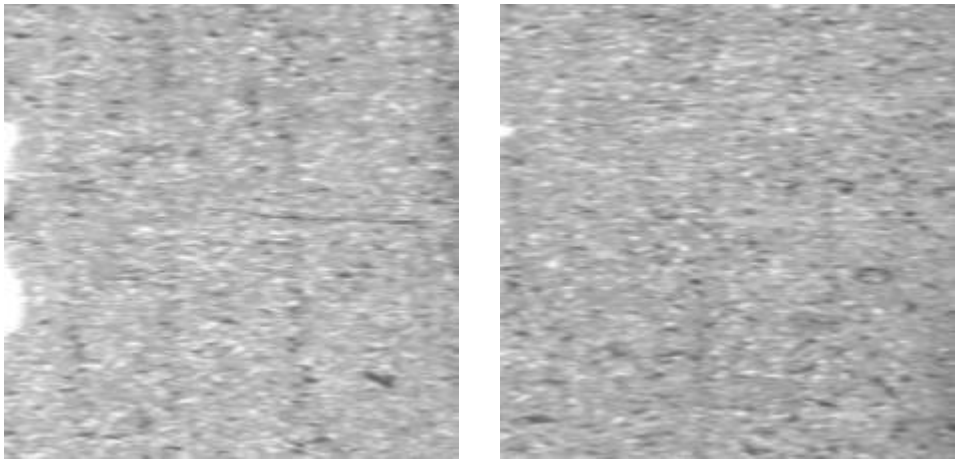

Figure 7: 227*227 format images

Data enhancement was then performed to transform the background environment of the original image to the scale [0.2, 0.5, 1.0, 1.2, 1.5] for ambient light and dark:

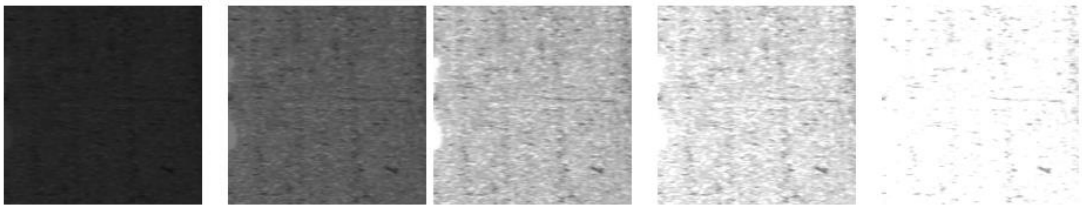

Figure 8: Background Enhancement

3.1.2 Feature extraction and classification

(1) Feature Extraction

In order to extract the features of the images, we used the Alexnet model, as using the last layer (fc7) directly would have output 4096 feature data for each image, which could easily lead to overfitting and computational difficulties in the subsequent SVM classification recognition. Therefore, we chose the penultimate layer of Alexnet (pool5) and performed a global average pooling of the feature map, averaging each 6x6 region into one value to obtain a 1x1x256 feature vector with a total of 256 values. The purpose of reducing the dimensionality of the features is achieved while retaining the important information.

(2) Evaluation indicators and Model training parameters

For the evaluation index of the model, the two evaluation indexes of training accuracy and Loss are chosen to evaluate the trained binary classification model for judging whether there are defects on the metal surface, and the evaluation index formula is:

$$\begin{aligned} Accuracy &= \frac{N}{M} \\ Loss &= -\sum_{i=1}^{M}\sum_{j=1}^{K} y_{ij}\log\, \hat{y_{ij}} \end{aligned} \tag{15}$$

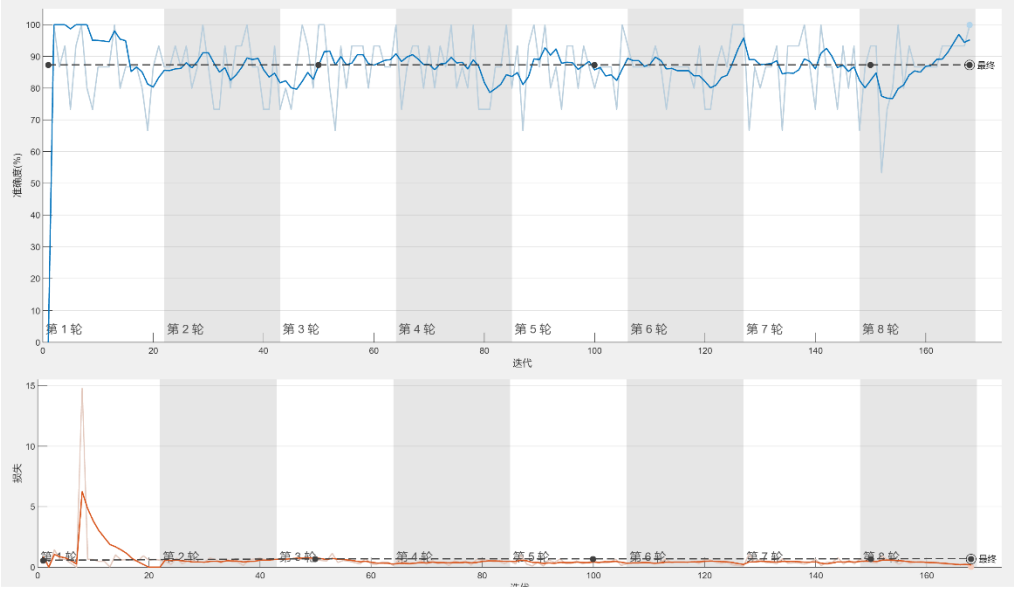

Figure 9: Accuracy and Loss

Finally, the feature matrix extracted by Alexnet, combined with the penalty parameter C obtained by SSA algorithm, is used for model training and validation. In the model training, the preprocessed image set is divided into defective and non-defective categories, while the validation set and training set are divided according to 3:7, the confidence level of the target category is set to 0.5, the batch iterative training method is adopted, and the batch is divided into 10 batches with a total of 200 iterations, the batch_size is set to 4, and the learning rate is set to 0.001. After repeated iterative training, the loss decreases rapidly in the first 50 iterations, and gradually smooths out after 100 iterations, in which the loss decreases in 50 iterations. After repeated iterations, the loss decreases rapidly in the first 50 iterations, and then the loss gradually smooths out after 100 iterations, in which the loss in the validation set is almost unchanged after 50 iterations, and the loss curve has been completely converged, so that we can accurately distinguish whether there is a defect on the metal surface or not.

batches	iteration	accuracy	loss
1	1	89.05%	5.4328%
4	50	89.05%	0.5325%
5	150	89.05%	0.5325%
10	200	89.05%	0.5325%

Table 1: Classification results

3.1.3 Computational volume and storage space of the model

The computational and storage space requirements depend on the structure of the model and the number of parameters. For the model Alexnet+SSA_SVM model, although the original model Alexnet has more convolutional and fully-connected layers[7], after our previous processing, such as dimensionality reduction of the feature matrix, global average pooling, etc., we also use the SSA fast parameter finding and the highly efficient binary classification model SVM[6], which makes the final defect detection model of the metal surface to be lightly deployed to the inexpensive mobile

devices while achieving a high recognition accuracy.

(1) Computational volume estimation:

We choose the FLOPs metric to measure the amount of computation.FLOPs is a metric that measures the amount of computation in a neural network by calculating floating-point operations.The estimation of model computation is based on the number of parameters in each layer and the size of the input feature maps, which can be accessed through the flops command in matlab.

$$FLOPs = parameters \times H \times W \tag{16}$$

(2) Storage space calculation

The storage space of the model mainly depends on the architecture and parameters of the model, for the original Alexnet model, the storage space is 61.1MB, which is reduced to 10.23MB after processing, considering that the SVM only needs to save the support vector information, which is very small compared to the Alexnet deep learning model for feature extraction.

FLOPs	Storage space
63.23M	10.23MB

Table 2: Storage space results

3.2 Auto-Tagging

The trained model is used to locate defects in the defective metal picture in the dataset kos05_Part5.jpg, it can be found that, for the identification of defects locating is very precise, the accuracy of the obvious determination of defects is 99.553%, at the same time, for the fuzzy defects locating is also with 55.154% of the probable estimation, the modelling accuracy of the marking can be found to be very high.

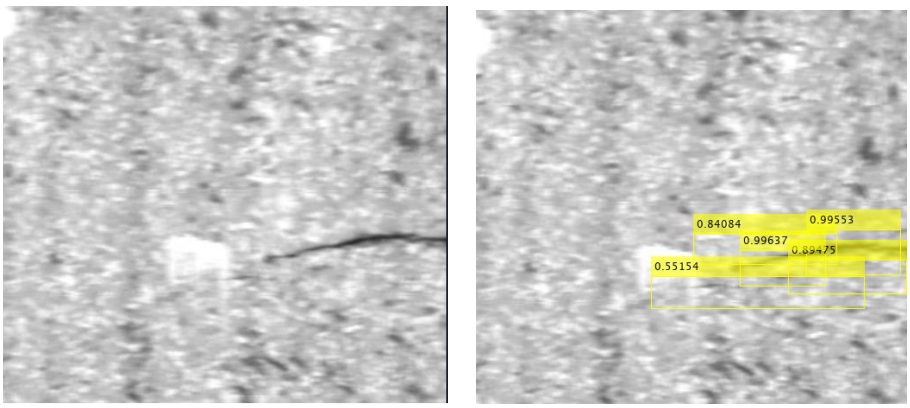

Figure 10: Automatically labeling results

We choose Root Mean Square Erro (RMSE) as the model evaluation index, which is a statistical index used to measure the difference between the predicted values and the actual observed values, the smaller the RMSE is, the smaller the difference between the predicted values and the actual observed values is, and the better the prediction performance of the model is. Meanwhile, the following results are obtained by combining the two evaluation indexes of accuracy and loss:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2]} \tag{17}$$

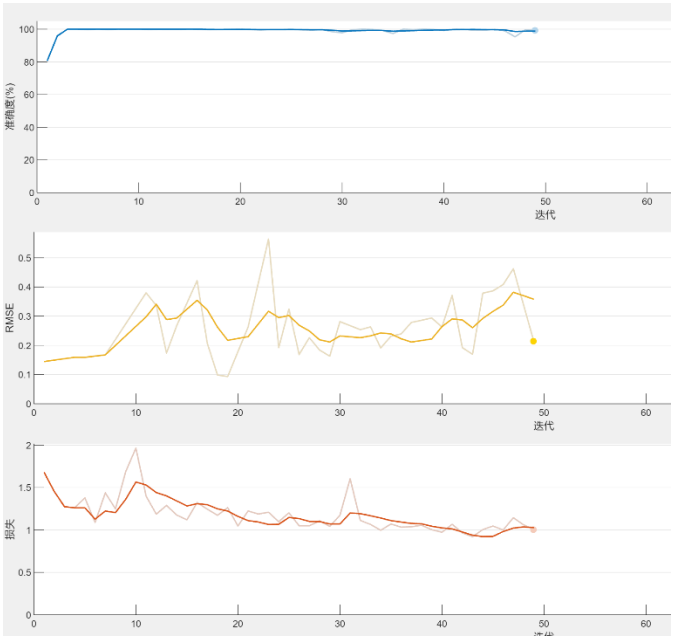

Figure 11: Evaluation of indicator graphs

It can be found that the RMSE of the model is gradually approaching to 1, and the loss of the model is decreasing, and the accuracy obtained is as high as 97.34%.

4 Conclusion

In this study, a metal surface defect recognition and localization model based on improved AlexNet-SVM and Fast R-CNN was successfully proposed. The proposed model pays special attention to lightweight design while ensuring high accuracy, so that it can be efficiently deployed on handheld devices with limited storage space and computing power. Through image preprocessing and data enhancement technology, combined with AlexNet network and SVM classifier optimized by SSA, a high recognition accuracy of 88.79% is achieved, and the storage space of the model is only 10.23MB. Further, Faster R-CNN and Region Proposal Network are used for pixel-level defect localization, and the accuracy is as high as 95.23%. Future work will focus on further optimization of the model, extended training of multiple data sets, integration with iot devices, and practical application testing of user feedback, in order to apply the model more widely in the field of intelligent manufacturing and product quality monitoring.

Reference

[1] Tabernik D, Šuc M, Skočaj D. Automated detection and segmentation of cracks in concrete surfaces using joined segmentation and classification deep neural network[J]. Construction and Building Materials, 2023, 408: 133582.

[2] Tabernik D, Šela S, Skvarč J, et al. Segmentation-based deep-learning approach for surface-defect detection[J]. Journal of Intelligent Manufacturing, 2020, 31(3): 759-776.

[3] Li J, Zhang L, Zheng W. Improved faster R-CNN and adaptive Canny algorithm for defect detection using eddy current thermography[J]. AIP Advances, 2024, 14(2).

[4] Jiang Q, Tan D, Li Y, et al. Object detection and classification of metal polishing shaft surface defects

based on convolutional neural network deep learning[J]. Applied Sciences, 2019, 10(1): 87.
[5] Wang Z, Zhu D. An accurate detection method for surface defects of complex components based on support vector machine and spreading algorithm[J]. Measurement, 2019, 147: 106886.
[6] Vishwanathan S V M, Murty M N. SSVM: a simple SVM algorithm[C]//Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290). IEEE, 2002, 3: 2393-2398.
[7] Alom M Z, Taha T M, Yakopcic C, et al. The history began from alexnet: A comprehensive survey on deep learning approaches[J]. arxiv preprint arxiv:1803.01164, 2018.
[8] Alippi C, Disabato S, Roveri M. Moving convolutional neural networks to embedded systems: the alexnet and VGG-16 case[C]//2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). IEEE, 2018: 212-223.
[9] Samir S, Emary E, El-Sayed K, et al. Optimization of a pre-trained AlexNet model for detecting and localizing image forgeries[J]. Information, 2020, 11(5): 275.