Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 12, No. 2, 2024

YOLOv5s fabric defect detection model with mixed attention mechanism

Zekai Kong

HENAN POLYTECHNIC UNIVERSITY, Jiaozuo, China
510045742@qq.com

Abstract: In the field of industrial quality inspection, fabric defect detection is regarded as an important research topic. The traditional inspection method mainly relies on manual visual inspection, which is time-consuming and easy to be affected by human factors, and has the problems of low accuracy, slow speed and high cost. Therefore, it is very important to use digital image processing technology and target detection technology to develop high precision, high efficiency and low-cost fabric defect detection model suitable for application deployment. Driven by deep learning-based neural network technology, many models with excellent performance have emerged in the field of object detection. This paper will improve on the basis of YOLOv5s model. ACmix module is introduced in Backbone layer of YOLOv5s model. This module overcomes the problem that convolution operation may cause local information loss after feature extraction, improves the model's limitation on feature extraction of complex background images, and enhances the model's processing ability in local feature fusion. Compared with the original YOLOv5s algorithm, mAP has improved by 0.6%. In the complex background, the detection accuracy of fabric defects targets has been significantly optimized and improved.

Keywords: Fabric defects; Deep learning; Object detection; YOLOv5s; ACmix.

1. Introduction

As a pillar industry of the national economy, the textile industry is also related to the life of the residents. It is widely used in People's Daily life, and the quality of cloth is closely related to the quality of people's life. Different quality of cloth will lead to its different prices, affecting the economic benefits of enterprises, and there are many problems in manual visual inspection. Therefore, the automatic fabric defect detection scheme is very important. Many scholars have made relevant research on fabric defect detection. Lin, S of Harbin Institute of Technology proposed an autonomous transfer learning network for multi-color fabric defect detection in order to solve the problem of insufficient training set samples. Zhang, J proposed a fabric defect system based on improved MobileNetV2-SSDLite cloud edge computing. ; Liu, Q. et al., adopting SoftPool instead of MaxPool's novel SPP structure ing, J. Et al proposed an improved yolov3 algorithm, combining the size of structural defects and k-means algorithm to carry out dimensional clustering of target frames, and added YOLO detection layer to feature maps of different sizes. The method was mainly applied to grey cloth and plaid fabric . Jun, X. Et al. proposed a two-stage strategy fabric defect detection method based on deep convolutional neural networks, which mainly includes two steps, namely local defect prediction and global defect recognition. The initiation-V1 model was used to predict whether there were defects in local areas, and the LeNet-5 model was used to identify the types of defects in the fabric. Rong-qiang, L. et al., proposed an improved convolutional neural network CU-Net for fabric defect detection, and improved the classical network U-Net Elemmi, M. C. Et al., use shallow and deep networks to classify structure images with and without defects; In order to compress model parameters, Wu Zhiyang et al., Xiamen University, proposed a dual-network parallel model training method based on convolutional neural networks for monochromatic fabric defects. Ouyang, W et al proposed a hybrid method, which uses statistical defect information and CNN to detect fabric defects. The improvements in this paper are as follows:

ACmix module is introduced in Backbone layer of YOLOv5s model. This module overcomes the problem that convolution operation may cause local information loss after feature extraction, improves the model's limitation on feature extraction of complex background images, and enhances the model's processing ability in local feature fusion.

2. YOLOv5

YOLOv5 algorithm is a target detection algorithm with high reliability and stability, and is easy to deploy and train, so it becomes one of the first-stage detection algorithms with the highest accuracy at present. In this paper, YOLOv5 version 6.0 is taken as the benchmark network model, which can be divided into three main parts: backbone network, neck network and head network. This model design makes the process of feature extraction and prediction more efficient, while the number of model parameters is small, the training speed is fast, and it is suitable for various scenarios with limited computing resources. In addition, the algorithm also has good generalization ability and adaptability, and shows excellent detection effect in various complex scenes. Model diagram of YOLOv5 algorithm.

(1) Input end

The input end of YOLOv5 mainly includes Mosaic data enhancement, adaptive computation anchor frame and adaptive scaling image.

The input terminal of YOLOv5 has been improved by adopting adaptive image scaling technology, which plays a crucial role in the improvement of reasoning speed. In the general object detection algorithm, in order to make different pictures have the same size, the input side often forces the picture to a standard size, which inevitably introduces some redundant black edges, thus affecting the inference speed to a

certain extent. However, due to the different aspect ratio of many images, YOLOv5 adopts a unique strategy, that is, according to the aspect ratio of the original image, the method with the least black edge is used to reduce the calculation amount, so as to achieve the purpose of greatly improving the reasoning speed.

(2) Backbone feature extraction network

The backbone network of YOLOv5 consists of Focus structure, C3 structure and SPP structure.

The innovation of YOLOv5 is the introduction of the Focus structure, the key step of which is the slicing operation of the picture. In this structure, the input image is first expanded by four times through interval sampling, so as to obtain a feature map with higher dimensions. Then, by convolution operation, the double down-sampled feature map can be realized without the loss of feature information. Compared with the common convolution operation, the Focus module is optimized to realize the downsampling with less computational cost and increase the channel dimension, thus reducing the number of parameters and improving the speed.

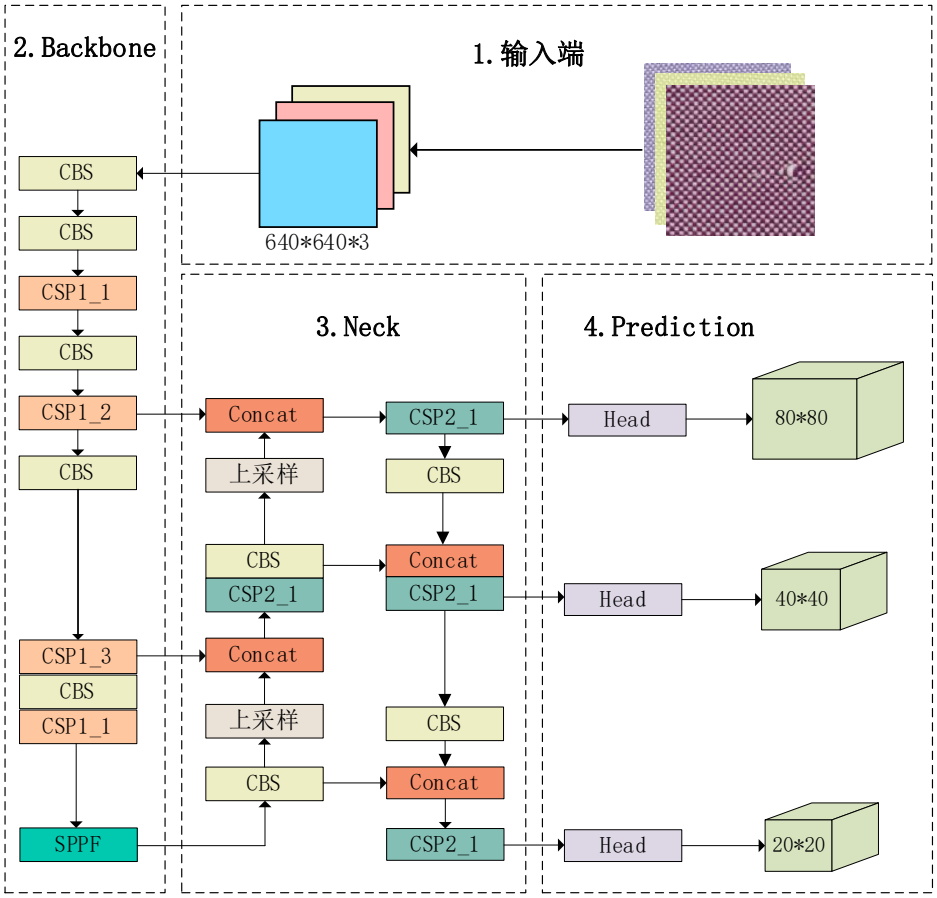

Fig.1 YOLOv5

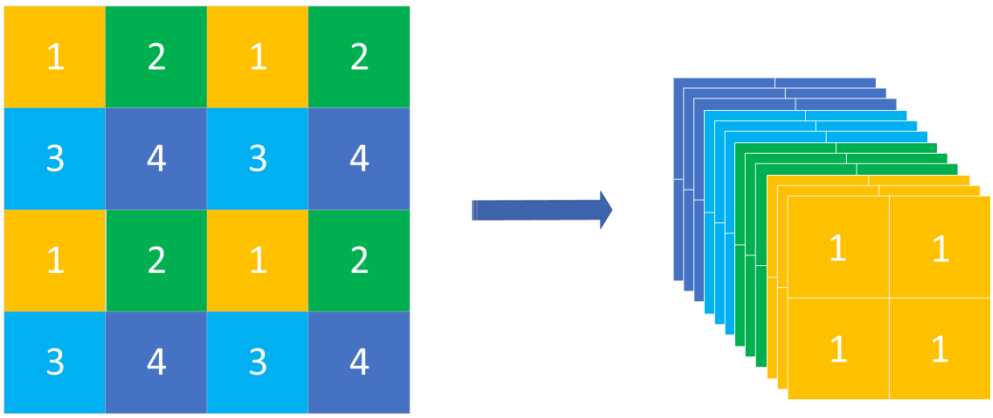

Fig.2 Focus structure

In YOLOv5, the cross-stage merge strategy in CSPNet is borrowed, and C3 module is designed, which is the main module for learning residual features. The C3 module divides the input features of the front end into two parts, one using a Bottleneck stack and standard convolution (CBL), the other using only a basic convolution module, and finally Concat the two parts. Due to the deep Backbone network of YOLOv5, the use of C3 module with residual structure can strengthen the gradient information in backpropagation, and effectively solve the problem of gradient disappearance and gradient repetition caused by deepening network. At the same time, C3 module can also make the obtained feature granularity finer, enhance the learning ability of the network, and reduce the computational load to a certain extent.

Based on the idea of spatial pyramid, SPP module uses maximum pooling to integrate the features of different receptive fields, which significantly improves the detection effect in complex multi-target scenarios in YOLOv5 algorithm. Its structure is shown in Figure 3. Firstly, the input channels are halved by a standard convolution (CBL) module, and then the input feature graphs are maximized with convolution kernel sizes of 5, 9, and 13 respectively, and the results of three maximized pooling are compared with those of non-pooling operations

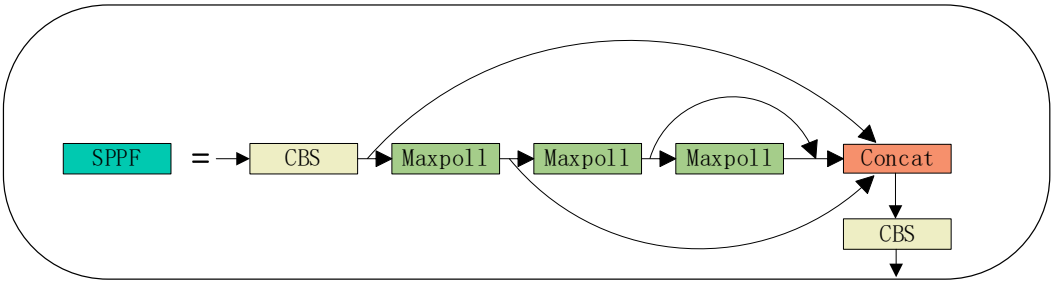

Fig.3 Internal structure of SPP module

Concat the data. Finally, the number of channels after concat is doubled. The advantage of SPP module lies in the fusion of local features and global features of the feature map, which enriches the information of the feature map, thus strengthening the detection effect in the scene with large size difference of the target.

(3) Neck strengthens the feature extraction network

The neck network of YOLOv5 adopts the methods of FPN and PANet. The basic idea of FPN is to upsample the output feature maps generated by multiple convolution downsampling operations of the feature extraction network to generate multiple new feature maps for detecting targets of different scales. The feature fusion path of FPN is top-down, that is, starting from the high-level feature map, and gradually generating multiple feature maps of different scales through upsampling and fusion operations. PANet, on the other hand, reintroduces a new bottom-up feature fusion path, which further improves the detection accuracy of objects of different scales by fusing low-level feature maps with high-level feature maps. This bottom-up feature fusion path enables the network to consider the feature information of different scales more comprehensively, thus improving the accuracy and robustness of target detection.

(4) Head Network

The head network of YOLOv5 consists of three detectors, corresponding to the detection process of the feature map of different scales. In the process of target detection, the regression loss function of bounding box is very important. YOLOv5 adopts CIoU as the bounding box regression loss function, which can effectively solve the problem of bounding box IoU and improve the accuracy and robustness of target detection. At the same time, YOLOv5 also adopts binary cross entropy loss as the classification loss function, which can effectively deal with the problem of target classification and improve the accuracy and stability of target detection. Such loss function design enables YOLOv5 to achieve excellent performance in target detection tasks.

3. YOLOv5 algorithm is improved

Human visual attention is a unique signal processing mechanism, and it is a survival and development mechanism formed after a long evolution. Studies have found that when processing large amounts of information, people automatically filter out secondary information and focus their attention on elements that need to be focused to quickly screen out high-value information. In recent years, scientists have begun to mimic human visual cognition by introducing attention mechanisms into neural network models, enabling networks to automatically focus on and learn important feature information in images. Studies have shown that the introduction of attention mechanism can effectively improve the network's ability to extract key information and improve the model's performance.

In the task of fabric defect feature extraction, convolutional neural networks usually use convolutional kernel for feature extraction. Traditional convolution operations apply aggregate functions on local receptive fields based on convolutional weights that are shared throughout the feature map, providing inductive bias for image processing. At present, self-attention mechanisms are widely used in the field of vision. Based on the context of input features, this mechanism improves the accuracy of feature representation by weighted average operation. The self-attention module usually determines the weights by calculating the similarity between related pixels, thus adaptively focusing different areas. This mechanism can effectively distinguish between objects (such as fabric defect points) and complex backgrounds, expand the network receptive field and capture more fabric defect point information features.

Studies have shown that early attention mechanisms, such as SENet and CBAM, can enhance the convolutional module. Based on this observation, Xuran Pan et al proposed the mixed attention mechanism ACmix[10] module in CVPR 2022. In this paper, ACmix module is introduced in Backbone layer of the model, which integrates the characteristics of ACmix and SPPF module, aiming at reducing information loss, aggregating more information, expanding receptive field, and improving the processing power of the model in local feature fusion. Through the introduction of AC-SPPF module, the optimization and improvement of the detection accuracy of fabric defect point targets under complex background are realized, and YOLOv5s has made remarkable progress in this aspect.

In this paper, a new model structure AC-SPPF is proposed to solve the problem that local feature information may be lost due to SPPF pooled downsampling after feature extraction by convolution operation. The model structure of AC-SPPF is shown in Figure 5. Before the fusion of the maximum pooled layers, the CBS Convolutional (1×1) pre-embedded convolutional self-attention module ACmix is introduced. The specific process is as follows: In Backbone, after a series of convolution operations, the input feature map first passes through the ACmix module to enhance information aggregation and weaken the interference of complex

background information. Later, the features at different scales are spliced together by reducing the number of channels in the convolution kernel. Then, the channel number is adjusted by the CBS convolution (1×1) module to restore the channel number to the original feature. Finally, the residual structure is combined with the original feature map to retain the rich local information in the original feature and output the feature.

The model integrates self-attention and convolution operations, and only ACmix module is added before SPPF to avoid excessive attention mechanism to increase the computational load. The reasoning speed of the model is improved with minimal computational overhead through ACmix module.

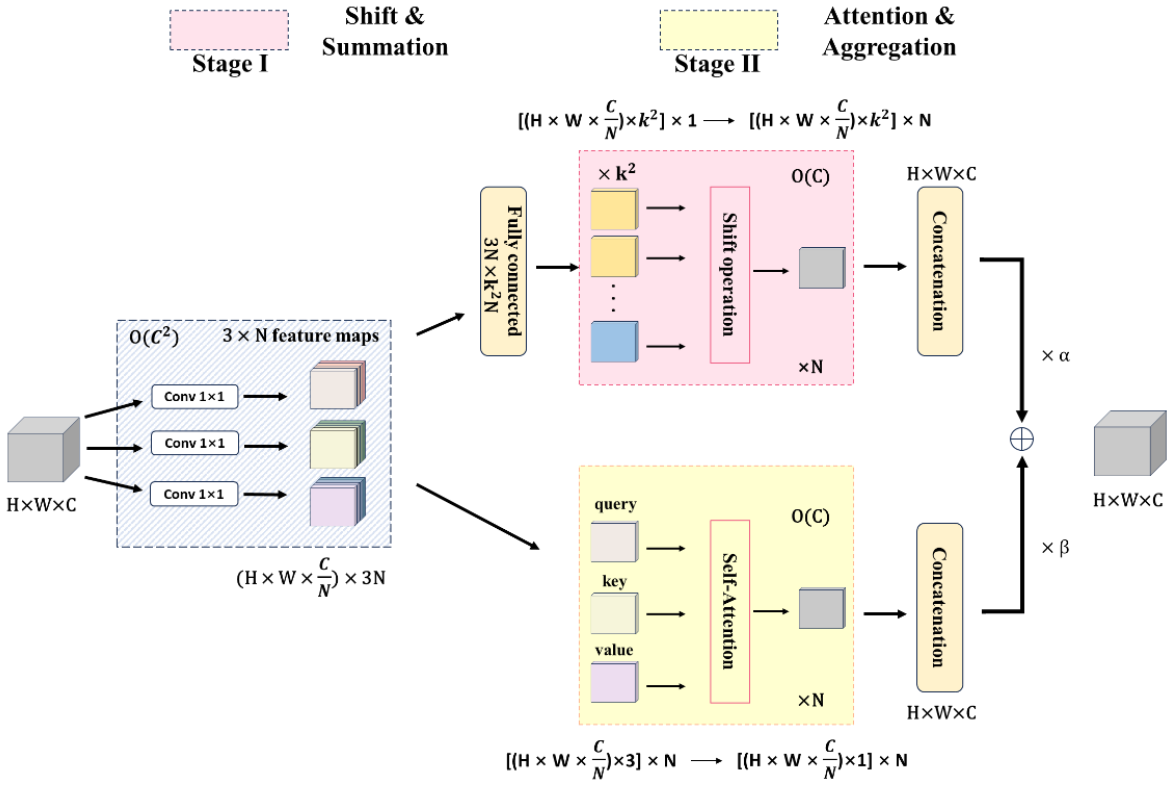

Fig.4 ACmix module structure3-

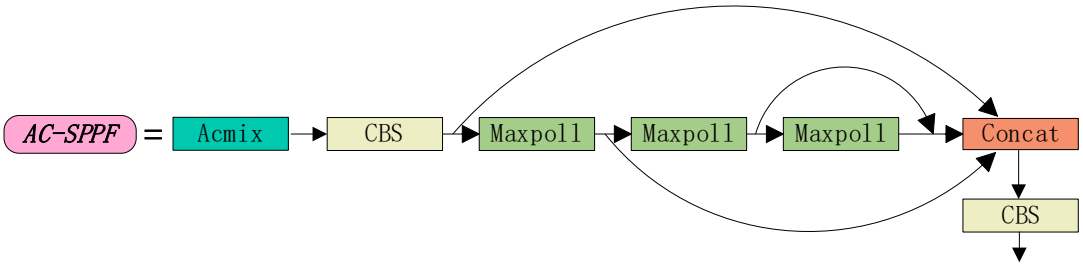

Fig.5 Structure diagram of AC-SPPF module3-

4. Experimental analysis

4.1. Evaluation index

Object detection is a complex task, which requires evaluation indexes to evaluate algorithm performance. The object detection results of the detection model are divided into true samples (there is a target in the image) and false samples (there is no target in the image). The real case can be further divided into two cases of the presence and absence of the target.

mAP is one of the commonly used evaluation indicators in the field of target detection, calculating the average of AP for all categories. AP (Average Precision) is the precision on a single class - the area under the recall curve, used to evaluate the performance of an algorithm on different classes. By calculating the accuracy and recall values under different confidence thresholds and plotting the AP curve, it is possible to observe how the algorithm's performance changes with the thresholds. The mAP is the average of all categories of aps and gives a complete picture of the overall performance level of the algorithm.

$$mAP(meanAveragePrecision) = \frac{\sum AP}{N(TomtalImages)} \quad (1)$$

The crossover ratio is one of the commonly used indicators in the evaluation of object detection algorithms, which is used to measure the accuracy of the algorithm. It determines the accuracy of the prediction result by calculating the ratio of the intersection area and the union area between the predicted

bounding box and the real bounding box. When the intersection ratio is greater than the set threshold, the prediction result of the model is considered correct; Otherwise, it is considered wrong. The threshold for the crossover ratio can be set according to the needs of the specific task.

$$IOU = \frac{A \cap B}{A \cup B} \tag{2}$$

Figure 6, where A represents the prediction box and B represents the target reality box:

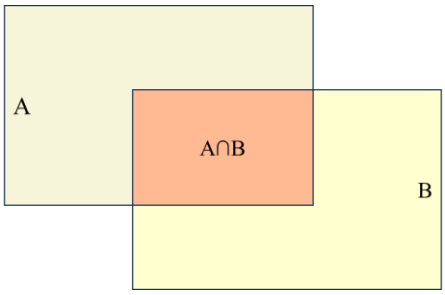

Fig.6 IOU schematic4-1

4.2. Data Set

The experimental data set in this paper comes from the actual textile production workshop. Due to the limited number of original samples, in order to expand the data set required for the experiment, the method of data enhancement is adopted. Specifically, the original data set was expanded to 2115 by image flipping, cropping, image mixing and other operations, and the number of various defect types was kept consistent. The data set contains several common fabric defect types, including broken yarn, stain, tear, cotton ball, deyarn, knot, warping knot, coarse dimension, etc. In order to reduce the influence of too many categories on the model, the three defect types of knot, warping knot and coarse dimension are uniformly classified as yarn defects. The final data set contains a total of 6 categories, and the 2115 samples are divided into the training set (1692 pieces) and the test set (423 pieces) according to the ratio of 8:2. The six categories were DuanSha (broken yarn), WuZi (stained), PoDong (broken hole), cotton ball (MianQiu), TuoSha (removed yarn), and DaiSha (with yarn). Meanwhile, labelimg annotation tool was used to manually label the data set, and the annotation file was saved in text format according to the VOC data set format.

4.3. Experimental analysis

By comparing the data of the proposed YOLOV5-ACMIX algorithm and YOLOv5 algorithm, it can be observed that mAP has a 0.6% improvement, which is very significant compared with the 0.5% improvement of YOLOv5 algorithm. This indicates that Acmix hybrid attention module can effectively enhance the network's attention to small target defect points, and make the backbone network more focused on extracting global information of images. Although the parameters have been increased, they are still within a reasonable range.

Table 1. comparative experiment

Method	Precision (%)	Recall (%)	mAp@0.5 (%)	mAp@.5:.95(%)	Parameters(M)
YOLOv5	94.2	93.3	95.9	51	7.02
YOLOv5-ACmix	95.2	92.2	96.5	49.9	7.85

5. Conclusion

Aiming at the challenges brought by interference factors such as image noise and background, Acmix module is added to Backbone layer of YOLOv5s network to form a new AC-SPPF module, which reduces the information loss in feature extraction of the original model, aggregates more information and expands receptive field, and improves the processing ability of the model in local feature fusion. Compared with the original YOLOv5s algorithm, mAP has improved by 0.6%, and the detection accuracy of fabric defect targets under complex background has been significantly optimized and improved.

References

[1] LIN S, HE Z, SUN L. Self-Transfer Learning Network for Multicolor Fabric Defect Detection[J]. Neural Processing Letters, 2022.

[2] ZHANG J, JING J, LU P, et al. Improved MobileNetV2-SSDLite for automatic fabric defect detection system based on cloud-edge computing[J]. Measurement, 2022, 201:111665.

[3] LIU Q, WANG C, LI Y, et al. A Fabric Defect Detection Method Based on Deep Learning[J]. IEEE Access, 2022,10({}): 4284-4296.

[4] JING J, ZHUO D, ZHANG H, et al. Fabric defect detection using the improved YOLOv3 model[J]. Journal of Engineered Fibers and Fabrics, 20,15({}): 2078651668.

[5] JUN X, WANG J, ZHOU J, et al. Fabric defect detection based on a deep convolutional neural network using a two-stage strategy[J]. Textile Research Journal, 2021,91(1-2): 130-142.

[6] RONG-QIANG L, MING-HUI L, JIA-CHEN S, et al. Fabric Defect Detection Method Based on Improved U-Net[J]. Journal of Physics: Conference Series, 2021,1948(1): 12160.

[7] ELEMMI M C, ANAMI B S, MALVADE N N. Defective and nondefective classif ication of fabric images using shallow and deep networks[J]. International Journal of Intelligent Systems, 2022,37(3): 2293-2318.

[8] Wu Zhiyang, Zhuo Yong, Li Jun, et al. Convolutional Neural Network based Fast Defect Detection Algorithm for Monochromatic fabric [J]. Journal of Computer-Aided Design and Graphics, 2018,30(12): 2262-2270.

[9] OUYANG W, XU B, HOU J, et al. Fabric Defect Detection Using Activation Layer Embedded Convolutional Neural Network[J]. IEEE Access, 2019,7({}): 70130-70140.

[10] PAN X, GE C, LU R, et al. On the Integration of Self-Attention and Convolution[Z]. 2022.