Improved YOLOv8 garment sewing defect detection method based on attention mechanism

Zengbo Xu[1], Yuchen Bao[2], Bingqiang Tian[3]
College of Textile and Clothing, Shanghai University of Engineering Science,
Songjiang, Shanghai, 201620, China
[1]Corresponding author
E-mail: [1]*xuzengbo@aliyun.com*, [2]*2269435134@qq.com*, [3]*1324379205@qq.com*

Check for updates

Abstract. For the garment sewing defect detection method, this paper proposes an improved YOLOv8-FPCA scheme based on the YOLOv8 algorithm, which improves the YOLOv8 target detection head to enhance the information extraction of small target defects, then introduces Focal Loss to optimize the loss function to guide the network to better handle target data sets with different difficulties and imbalances. Finally, the attention mechanism CA is added to the YOLOv8 network structure to achieve multi-scale feature fusion extraction, and the attention mechanisms CABM and SENet are added at the same locations for experimental comparison. The results show that after increasing the attention mechanisms of CBAM and SENet, the mAP@0.5 model increased by 1.7 % and 1.9 % respectively. The CA attention mechanism emphasizes the importance of location information, and the model has better accuracy and recall after adding the CA attention mechanism, with a 3.7 % increase in mAP@0.5, indicating that YOLOv8-CA has better performance in sewing defect detection.

Keywords: target detection, YOLOv8, attention mechanism, sewing defect detection.

1. Introduction

Economic development promotes the expansion of the garment production scale. Various clothing brands flow around the globe through international trade and other forms. To enhance the competitiveness of garments in the consumer market, it is particularly important to achieve garment quality inspection in the process of garment factory inspection [1]. Since the process of garment production requires manual sewing, it inevitably leads to garment sewing defects. Common sewing defects can be divided into: broken threads, uneven stitching, threaded ends, poorly turned threads, and other problems. Traditional garment quality inspection is mainly conducted manually through sampling, but a manual inspection can be disturbed by external factors to produce visual fatigue, resulting in reduced judgment and low reliability of detection [2]. In the actual export trade process, sewing defects are likely to affect the overseas sales of garments and cause economic losses to enterprises, so it is necessary to carry out defect detection of garment sewing defects to ensure the quality of exports.

In recent years, deep learning techniques have been widely used in the field of defect detection. Rezazadeh N. [3] used convolutional neural networks and persistence spectra to identify shallow cracks in rotating systems, Nima R. [4] used wavelet transform and deep learning methods to classify cracks in rotor bearing systems. At the present time, the traditional detection scheme mainly adopts manual detection, and the detection of faults is also mainly focused on the faults of fabrics. To make up for the shortage of manual detection, manufacturers have tried to achieve the detection of defects by using image processing methods, which are traditionally performed mainly by statistical analysis, structural analysis, and spectral analysis [5-6]. Histogram statistics [7] uses the distance matching function to find the location of defects in the fabric, and the grayscale co-occurrence matrix [8] achieves the detection of defects by calculating different features based on binary patterns. Fourier transforms [9] are used to monitor the spatial information of the fabric and determine the presence of defect information by the irregular structure of the fabric defects.

Wavelet transformers [10] are used to obtain effective fabric blemish information by extracting fabric feature information. As well as other methods such as the Gabor filter [11] and the Gaussian mixture model [12] are used to detect and identify blemishes using image processing. These methods have some advantages over manual detection, but frequency domain methods such as wavelet transform and Gabor filter are very sensitive to noise and interference. In the presence of noise or interference, the model may produce false positives or missed positives. After AlexNet [13] was proposed, deep convolutional neural networks have been widely used. The mainstream one-stage and two-stage deep learning algorithm frameworks have also been applied to defect detection. Two-stage algorithms such as R-CNN [14], Fast R-CNN [15], Faster R-CNN [16], MaskR-CNN [17], etc., divide the detection problem into two parts: candidate region generation and target detection, which can effectively handle targets of different scales with high detection accuracy. The single-stage target detection algorithms You Only Look Once [18] (YOLO) and SSD [19]. The target detection problem is transformed into a regression problem, and the class and bounding box of the object are predicted simultaneously by convolutional neural networks. Single-stage detection has real-time performance and a simple network structure with fast inference speed. Compared with structure-based, spectral analysis-based, and statistical-based image recognition methods, deep learning-based target detection not only has higher accuracy but can be applied to large-scale data training. Bao Y. C. [20] the YOLOv8-MBRGA is an improved algorithm based on YOLOv8, which introduces the BiFPN pyramid to replace the concatenation in the head layer, and transmits the semantic information to different feature scales so as to enhance the integration of the features.

To solve the problem of garment sewing defects, the improved YOLOv8 algorithm can not only achieve the target detection task in multiple scenes but also support multi-scale information feature fusion in feature information extraction with higher accuracy and recall. In this paper, we propose research on an improved YOLOv8 garment sewing defect detection method based on the attention mechanism. The Coordinate Attention [21] (CA) mechanism is chosen to encode horizontal and vertical location information into channel attention, which enables the mobile network to give attention to a larger range of location information without bringing too much computational effort. Compared with the original YOLOv8 network structure and the two-stage algorithm, the improved model improved mAP@0.5, recall, and accuracy.

The main contributions of this paper are as follows.

1. Construct the dataset with sewing defects as the object of model training.

2. Adding two small target detection heads to the YOLOv8 algorithm to focus on the extraction of feature information of small targets with defects.

3. Add a CA attention mechanism to enhance the multi-scale feature extraction capability of YOLOv8 and enhance the attention to small targets with defects.

4. Optimize CLOU to introduce Focalloss [22] to solve the influence of the background texture of garments during target detection.

2. Related work

There are many methods to achieve the task of fabric defect detection, which are mainly divided into structure-based, spectrum analysis-based, and statistical image processing schemes, and another is the target detection task based on deep learning. Zahra Pourkaramdel [23] proposed a new defect detection scheme for local binary patterns, which can detect 97.6 % of fabric defects, However, the disadvantage is that the detection effect is poor. Andong [24] proposed a double sparse low-order decomposition method for fabric printing defects with a template fabric graph as the defect prior to achieve defect detection for printing class defects. Yudi Zhao [25] proposed an integrated CNN model based on visual long- and short-term memory to enhance feature information extraction from images with the help of convolutional neural networks in complex texture backgrounds, which is effective for fabric defect detection.

Target detection algorithms usually need to process large-scale image and video data. The

high-performance computing power of GPU enables target detection algorithms to process large-scale data quickly, accelerating the process of data loading, pre-processing, feature extraction, and result output. Compared with detection models such as local binary patterns and double sparse low-order decomposition, target detection algorithms require large datasets as training data for the models, from which data feature information is extracted for the identification and classification of defects. The high computational power of GPUs has inspired researchers to optimize and innovate target detection algorithms to design more efficient, accurate, and complex target detection algorithms. Jing [26] et al. used the K-means algorithm to improve YOLOv3 [27] by adding detection layers to different feature maps to achieve a fusion of information in the higher and lower layers of feature maps, effectively reducing the false detection of fabrics. Qiang Liul [28] used SoftPool to improve the SPP structure to reduce the negative side effects of the SPP structure and process the feature maps effectively. Shengbao Xu [29] proposed a deformable convolution instead of the traditional square convolution to improve the R-CNN. It can automatically learn feature representation, has better robustness for details such as small objects and edges, and achieves better results on small target detection. Chen Jiang [30] implemented an improvement to Faster-RCNN by incorporating an attention mechanism to achieve defect detection and enhanced feature fusion for Pennsylvania fabrics. Two-stage detection algorithm networks are slow and complex and usually require more computational resources and time compared to single-stage target detection algorithms. This makes the two-stage detection algorithm perform poorly in real-time applications and demanding speed scenarios.

The above is the use of target detection for fabric defects. The current main process of defect detection is through feature extraction of defect information, and then determining the type of defect. In the past, the work was mainly done on the fabric for defect detection, and the fabric was flattened for detection. Due to the integrity of the garment, it is not possible to perform a flattening inspection of the fabric. In the process of sewing fabrics together to make garments, various sewing defects are created. Compared to fabric inspection, garment sewing defects detection is more challenging and complex. To solve the problem of sewing defects, we improved the YOLOv8 algorithm to enhance the extraction of more complex background features for multi-scale garments.

3. Experimental method

In this section, we introduce YOLOv8, the attention mechanism, and the improved scheme. Details are shown below.

3.1. YOLOv8 network structure

In this paper, the YOLOv8 single-stage target detection algorithm is proposed to implement the task of detecting defects in garment sewing. Based on the success of previous YOLO algorithms, YOLOv8 introduces new features and improvements. Compared to traditional image pre-processing and two-stage methods, YOLOv8 has high speed and accuracy. YOLOv8 detects and fuses feature maps at different levels, enabling the algorithm to simultaneously perceive targets of different scales, and has better robustness and accuracy in detecting small targets such as sewing defects. YOLOv8 is currently used as the state-of-the-art target detection algorithm, and this paper improves YOLOv8 so that the improved algorithm can be better applied to garment sewing defects detection. Compared with previous target detection algorithms, YOLOv8 combines the advantages of previous algorithms and designs a new backbone network, C2F, with improved detection heads and loss functions to improve the performance and flexibility of the algorithm. the YOLOv8 network structure is shown in Fig. 1. YOLOv8 is shown in:

1. The backbone network draws on the C3 module in YOLOv5 and the ELAN design idea in YOLOv7[30] and adopts the C2F module to obtain richer gradient flow information while ensuring lightweight.

2. Change the coupling head to an uncoupling head to separate the closely connected parts for higher flexibility and maintainability.

3. Abandon the Anchor-Base method using Anchor-Free anchorless model can directly predict the center of the object, which reduces the number of prediction boxes and improves the detection speed.
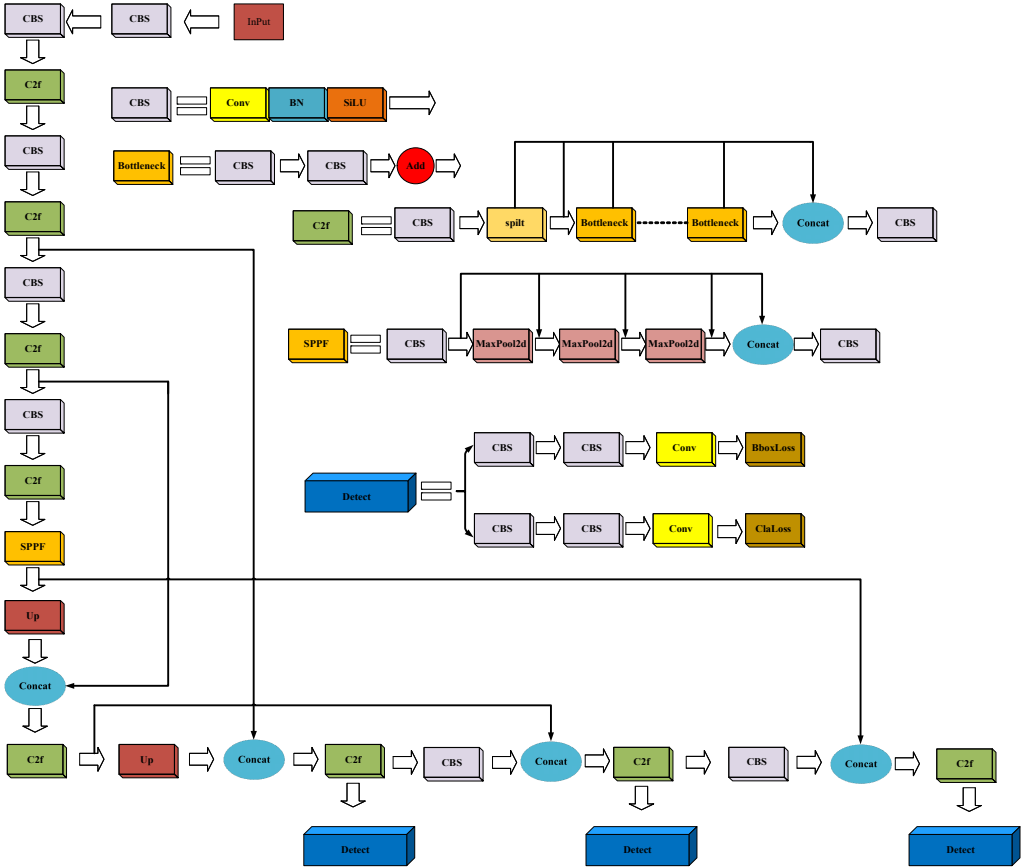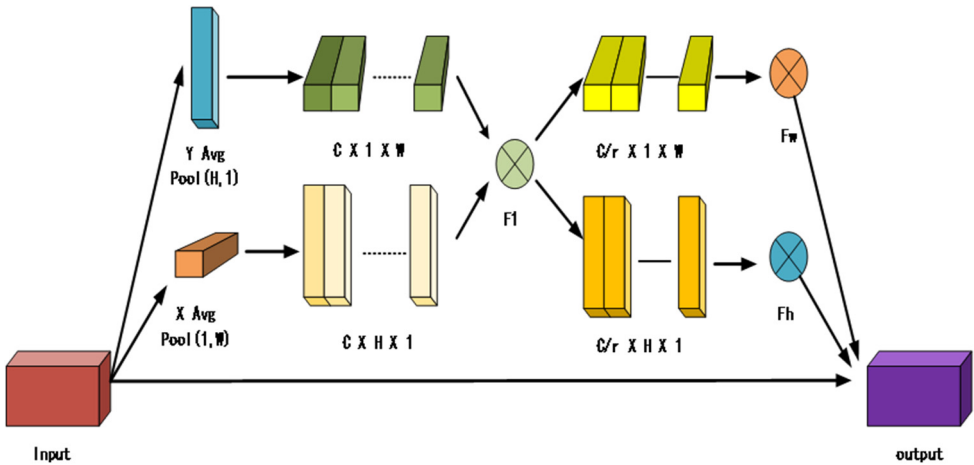

Fig. 1. YOLOv8 network structure diagram

Fig. 2. CA attention mechanism diagram

3.2. CA attention mechanism

The attention mechanism is a technique widely used in deep learning to help models focus on important parts of the input data and thus improve the performance of the model. The garment sewing defect studied in this paper belongs to small target information, and their features are not easily observed on a global scale. Although YOLOv8 has the mechanism of multi-scale fusion, it can detect different feature maps and fuse these feature maps. The main purpose of adding the attention mechanism is to improve the feature extraction of the convolutional neural network, and since adding the attention mechanism is likely to destroy the network structure of the original model, in this YOLOv8 improvement, this paper adds the coordination attention (CA) attention mechanism after the convolutional layer and does not change the number of channels of the original network. The CA attention mechanism is shown in Fig. 2.

In the CA attention mechanism, the input feature map ($C \times H \times W$), in the height (H) and width (W) directions, is pooled to generate two feature maps of size $C \times H \times 1$ and $C \times 1 \times W$, respectively, whose height and width on the cth channel are and Z_c^w, respectively, as shown in Eq. (1):

$$
\begin{aligned}
Z_c^h(h) &= \frac{1}{W} \sum_{0 \leq i \leq W} X_c(h, i), \\
Z_c^w(w) &= \frac{1}{H} \sum_{0 \leq i \leq H} X_c(j, w).
\end{aligned}
\tag{1}
$$

Eq. (1) is the pooling operation on the height H and width W, respectively.

Next, an intermediate feature map of size $C/r \times 1 \times (H+W)$ is obtained by concatenating these two feature maps in the channel dimension and applying a 1×1 convolution operation (denoted as F1), where r denotes the down sampling ratio and is calculated as shown in Eq. (2):

$$
f = \delta\big(F_1([Z^h, Z^w])\big),
\tag{2}
$$

where δ is a nonlinear activation function that maps the horizontal and horizontal feature information:

Subsequently, the intermediate feature map is split into two feature vector sums f_w and f_h. Finally, these two feature vectors are convolved by 1×1 (denoted as F_w and F_h) and processed by the activation function to obtain a feature map with the same number of channels as the input feature map, and the variation formula is shown in Eq. (3):

$$
\begin{aligned}
g^h &= \sigma\big(F_h(f^h)\big), \\
g^w &= \sigma\big(F_w(f^w)\big),
\end{aligned}
\tag{3}
$$

where σ is the activation function $sigmoid(x)$ that activates the features that have been convolved.

The whole process from the input features to the output features can be expressed by Eq. (4):

$$
Y_c(i, j) = X_c(i, j) \times g_c^h(i) \times g_c^w(j).
\tag{4}
$$

3.3. The small target detection head

The YOLOv8 network model incorporates multiscale features in the top-down path, after five downsampling to get the feature expression (P1, P2, P3, P4, P5) but, it does not change the size of the feature map of the final detection head. The feature map sizes of YOLOv8 target detection heads p3, p4, and p5 are 80×80, 40×40, and 20×20 used to detect targets with sizes of 8×8, 16×16, and 32×32, respectively, which results in YOLOv8's ability to detect large targets better than that

of small targets. In this paper, most of the self-built garment sewing defects in the dataset are small target defects that are difficult to find with the naked eye, and the pixels of these small targets are smaller than 8×8, which poses a challenge to the detection and recognition of the original YOLOv8 algorithm. Many tiny defects will lose most of their characteristic information even after five layers of downward adoption, resulting in p3, a high-resolution inspection head, being unable to successfully detect the defects, leading to the leakage of defect information.

To solve the problem of detecting small targets for garment sewing defects, we add the detection head of tiny objects corresponding to a 160×160 detection feature map for detecting targets above 4×4, as shown in Fig. 3. We elicit feature information in the P2 layer, which contains rich underlying feature information after two downward adoptions in the backbone. The newly added P2 detection head in the Neck network obtains both top-down and bottom-up feature information in the network structure and uses Caoncat's feature fusion to make the P2 layer detection head fast and effective when dealing with tiny targets. The newly added P2 layer and the original three detection heads can effectively mitigate the negative impact of scale variance. The added detection is designed for the underlying features and is generated using low-level, high-resolution feature maps. It is more sensitive to tiny targets and improves defect detection. Although the introduction of these detection heads increases the number of parameters and the number of computers, it allows the model to achieve better accuracy in blemish detection.

3.4. Introducing focal loss to improve loss function

CIoU is improved based on DIoU by increasing the loss of detection frame scale, which makes the difference between the predicted box and the real box smaller. Introducing Focal Loss in the garment sewing defect detection task reduces the influence of background complex information on the defect information, increase attention to samples that are difficult to classify, and improves the overall performance of the model by giving higher weights to these samples. The combination of CIoU and Focal Loss to obtain Focal CIOU improves the localization accuracy of the target boxes and handles the category imbalance problem. The network is guided during training to better handle target boxes and categories with different difficulties and imbalances. It has a good effect improvement under the background images with complex sewing defects. The Focal Loss formula is shown in Eq. (5):

$$FL(p_i) = -\alpha_s(1 - p_i)^r Ibp_i, \tag{5}$$

where α_s takes values between 0 and 1 to control positive and negative samples; p_i is the modulation factor of prediction, the closer to 1 is easy to classify and the loss is reduced; r is the index modulation factor, which controls the weight adjustment magnitude of difficult and easy samples by adjusting the size of the index.
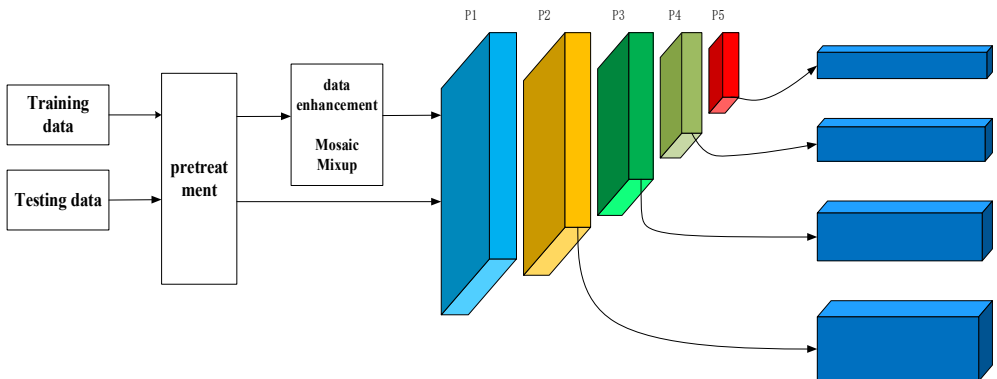

Fig. 3. Adding target detection header diagram

3.5. Improved YOLOv8 algorithm

To achieve the detection of small targets such as sewing defects, we first add small target detection heads to the Neck network of the YOLOv8 algorithm, perform upsampling, connect multi-scale feature layers, and reuse the C2F module to increase the depth of the network to achieve the extraction of small target feature information. CA is a more efficient attention mechanism that makes up for the shortcomings of the CBAM [31] (Convolutional Block Attention Module) attention mechanism, and SE [32] (Squeeze-and-Excitation Networks) attention mechanism. Convolutional Block Attention Module) attention mechanism, SE(Squeeze-and-Excitation Networks) attention mechanism.CA attention mechanism not only has dependencies in long-distance channels, but also can retain location information more accurately, so that without destroying the existing YOLOv8 network structure, this paper plans to add the attention mechanism CA before SPPF, and add the attention mechanism CA in the p2 and p3 detection head part of Neck network to enhance the capability of YOLOv8's multi-scale feature extraction and improve the detection efficiency, and finally use the Focal Loss function to deal with the sample imbalance phenomenon and suppress the complex image background. The improved scheme is called YOLOV8-FPCA, and the network structure of the improved YOLOv8-FPCA algorithm is shown in Fig. 4.
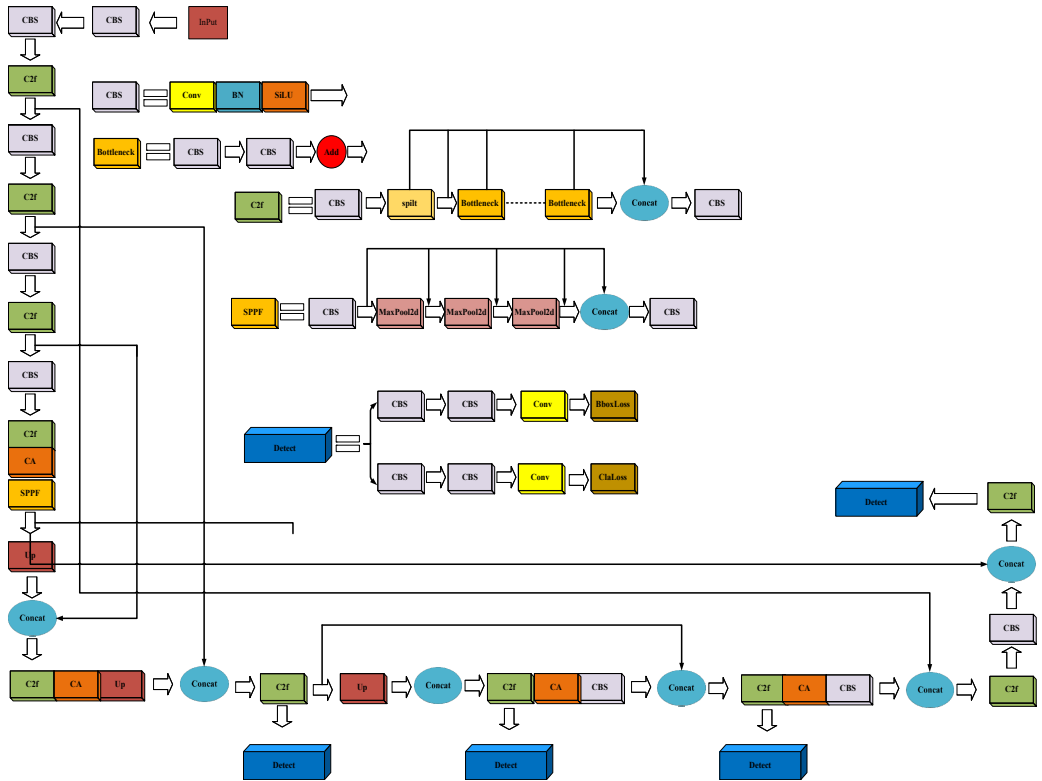
Fig. 4. YOLOv8 improved network structure diagram

4. Construction and pre-processing of the dataset

4.1. Construction of the dataset

The garments in the market have been qualified by the factory and there is no defect data. In order to obtain effective clothing sewing defect data, this article selects data through the offline

defect database of clothing manufacturing plants. The clothing fabrics collected this time are mainly made of cotton, linen and chemical synthetic fabrics through woven or knitted production, with high weaving density. The clothing fabrics collected are regular summer garment fabrics without embroidery or prints. For photographing the defects, a data set of 5000 garment sewing defects was obtained by placing the garments by spreading them flat, hanging them upright, etc., and using an AF-focused, 44-megapixel, 16x zoom camera for sampling. Finally, use the annotation tool to annotate the defect according to its type. The training set, validation set, and test set are divided according to 7:2:1.

The size of the rectangle box should meet the requirement that the rectangle box should be completely selected for the defects, and a part of the normal information should be taken into it. For the places where defects gather, defects can be selected through multiple rectangle boxes. In actual production and life, there may be multiple sewing defect information on the same piece of clothing. When labeling defects, all defect information needs to be comprehensively labeled. If the range of defects exists is relatively large, the range of frame selection is too large and is not conducive to algorithm model learning. Therefore, large defects can be divided into multiple local defects and then use multiple rectangular boxes for annotation. The defect annotation is shown in Fig. 5.
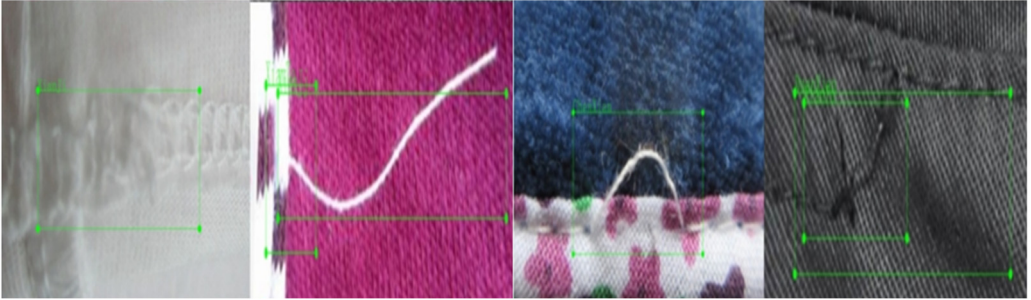

Fig. 5. Data labeling diagram

4.2. Pre-processing of data sets

The types of defects in the dataset of this paper are broken threads, uneven stitching, threads, car threads, to ensure the balance of the dataset in the sampling process, the number of each type of defects collected is 1250, and the total number of datasets in this paper is 5000. However, in the actual training process, Gaussian blur, Random noise, random rotation and other methods were used to enhance the original 5,000 data at a ratio of 1:6 and uniformly transform the pixel size of the data set images to 640×640. Finally, the total number of enhanced data sets was 30,000. Each category the total number of defect data is 7500, including 5250 training sets, 1500 verification sets, and 750 test sets.

YOLOv8 can perform Mosaic and mix-up data enhancement by setting parameters. mosaic enhancement, four randomly selected images are stitched together in proportion to increase the diversity of the data. In the mixup, two randomly selected training samples are linearly combined to mitigate the overfitting of the model on the decision boundary. The mixup formula:

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j,$$
$$\tilde{y} = \lambda y_i + (1 - \lambda)y_j, \tag{6}$$

where λ is the number of random samples of beat distribution.

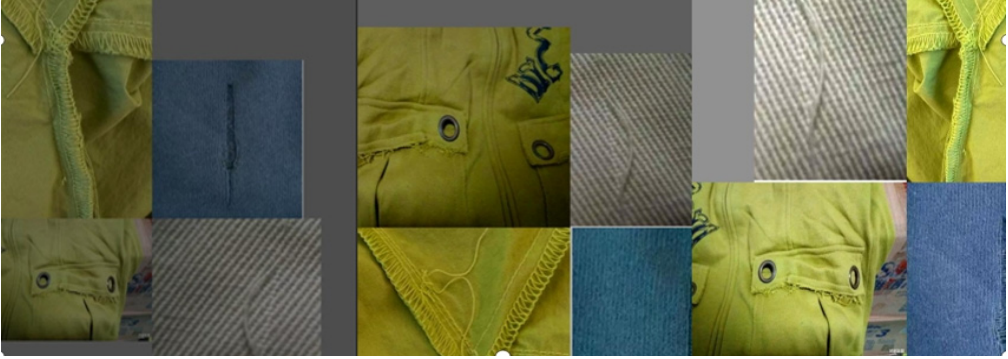The effect of Mosaic and mixup enhancement is shown in Fig. 6.

Fig. 6. Data enhancement effect diagram

4.3. Experimental environment and parameter settings

The environment of the experiments is trained in AutoDL using cloud GPU with RTX3090 24G graphics card. The environment settings are shown in Table 1. The model training parameters are set as shown in Table 2.

Table 1. Environment configuration information table

Name		Name	
Video cards	RTX3090	Video memory	24G
PyTorch	1.8.1	Python	3.8
Ubuntu	18.04	Cuda	11.1
CPU	15V 8338C	Memory	80G

Table 2. Training parameters

Parameters		Parameters	
Learning rate	0.01	Weight decay	0.0005
Batch size	128	Momentum	0.937
Image size	640×640	Number of iterations	300
Workers	20	Patience	50
Dropout	0.4	Nms	True

4.4. Experimental model evaluation criteria

All training models will be evaluated using uniform metrics, including the model's Precision, Recall, and Average Detection Precision (mAP), F1 scores. The evaluation criteria for this model are as follows:

$$precision = \frac{TP}{TP + FP}, \tag{7}$$

where TP denotes true positives and FP denotes false positives. Precision higher values in the assessment process are better. It is used to measure the percentage of positive samples with correct predictions among all predicted positive samples.

Verify how many pairs of data were recovered in the correct dataset:

$$Recall = \frac{TP}{TP + FN}. \tag{8}$$

where FN represents a false negative.

The F1-value is the sum of precision and recall, and the higher the value of F1, the better the

robustness of the model:

$$F_1 = \frac{2 * precision * Recall}{precision + Recall}. \tag{9}$$

Eq. (10) represents the mean of each class, while Eq. (11) measures the mean of the model as a whole, and only one of the most commonly used data for comparing models:

$$AP = \sum_{k=0}^{k=n-1} [R(k) - R(k+1)] * P(k), \tag{10}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} (AP)_i. \tag{11}$$

5. Experimental results

In the experimental results section, we will experimentally verify the performance of the YOLOv8-CA model.

5.1. Comparison of data enhancement methods

YOLOv8 can set parameters for Mosaic and mix up data enhancement. YOLOv8 turns on Mosaic enhancement by default to improve the generalization ability of the model and its adaptability to complex scenes. Improved model detection for small targets, occluded targets, and complex backgrounds. In this section, the models are compared using the default versions of YOLOV8 with Mosaic, mixup, and without data enhancement turned on, as shown in Table 3. By comparing the three approaches, it is found that using the default Mosaic data enhancement approach obtains better reflective correlation on P, R, F1, and mAP data, and the experiment will be continued next using Mosaic data enhancement.

Table 3. Comparison experiments of different data enhancement methods

Serial number	Mosaic	Mixup	P	R	F1	mAP@0.5
1	×	×	80.6 %	72.5 %	0.763	82.6 %
2	×	√	81.4 %	71.2 %	0.759	81.9 %
3	√	×	84.6 %	74.9 %	0.789	83.5 %

5.2. Comparative analysis of target detection algorithms

This article uses a total of six algorithm models from the literature and different versions of YOLO to conduct verification on self-made data sets. Table 4 shows that YOLOv8 has a precision of 84.6 %, a recall of 74.9 %, an F1 score of 0.789, a value of 83.5% for mAP@0.5, and a value of 58.9 % for mAP@0.5-0.95. Compared with the second-ranked YOLOv7 algorithm, YOLOv8 is 4.77 % higher than YOLOv7 in terms of mAP@0.5 results. YOLOv8 is comprehensively higher than YOLOv7 as well as other types of algorithmic models in terms of precision and recall. The experimental results show that the YOLOv8 algorithm can achieve classification and identification detection of sewing defects, and the YOLOv8 algorithm enhances the feature information of the network model and the ability of multi-scale fusion extraction, which can focus on the extraction of feature information of small targets of defects.

5.3. Comparison of ablation experiments

The improvement scheme of the YOLOv8 algorithm in this paper, after adding a small target detection head, using Focal Loss to optimize the loss function, and introducing a CA attention mechanism to enhance the extraction of feature information, finally obtained the improved

algorithm YOLOv8-FPCA as shown in Table 5. The addition of the minutiae table detection head mAP@0.5 to the original YOLOv8 algorithm increased by 2.6 %, and the accuracy, recall, and F1 score were significantly improved.

**Table 4.** Comparison of target detection algorithms

Experimental algorithms	P	R	F1	mAP@0.5	maAP@0.5-0.95
SSD [Doc. 17]	83.55 %	25.22 %	0.39	68.88 %	40.2 %
R-CNN [Doc. 12]	32.8 %	86.11 %	0.46	73.45 %	51.24 %
Faster-RCNN [Doc. 14]	60.4 %	59.38 %	0.59	71.94 %	52.31 %
YOLOv5	79.16 %	47.83 %	0.60	61.68 %	48.17 %
YOLOv7	80.31 %	52.17 %	0.63	78.73 %	54.48 %
YOLOv8	84.6 %	74.9 %	0.789	83.5 %	58.9 %

Table 5. Comparison experiments of improvement schemes

Improvement program			P	R	F1	mAP@0.5	maAP@0.5-0.95	Parameters	GFLOPS	FPS
Add-p2	Focal	CA								
×	×	×	84.6 %	74.9 %	0.789	83.5 %	58.9 %	3157200	8.9	48
√	×	×	87.4 %	76.6 %	0.816	86.1 %	60.2 %	3354144	17.4	42
√	√	×	87.7 %	76.9 %	0.819	86.5 %	60.9 %	3355433	17.4	43
√	√	√	90.2 %	81.2 %	0.84	90.2 %	65.6 %	3360824	17.4	43

Immediately afterward, we continued to add Focal and CIoU combined into Focal CIoU to guide the network to better handle target boxes and categories with different difficulty and imbalance during the training process, compared with adding only small target detection head mAP@0.5 up 0.4 %, no significant improvement after adding Focal due to the high quality of the dataset constructed in this paper after mosaic processing. Finally, the attention mechanism CA is added to focus on the important part of the input data to improve the model's performance in detecting small targets of sewing defects, and mAP@0.5 continues to rise by 3.7 % with a significant effect.

The experiments show that adding a small target detection head, Focal, and CA attention mechanism all achieve the rise of model accuracy, recall, F1 score and mAP@0.5. Compared with the original YOLOv8 algorithm, the improved algorithm YOLOv8-PFCA has a 6.7 % increase in mAP@0.5 and a 90 % increase in model calculations. However, the final parameters of the model only increase by 6.4 %, and the actual detection speed FPS drops from 48 to 43. There was no significant decline. It is shown that the improved algorithm YOLO-PFCA is effective and can be better utilized for garment sewing defect detection. The ablation experiment mAP@0.5 curve and loss curve are shown in Fig. 7.
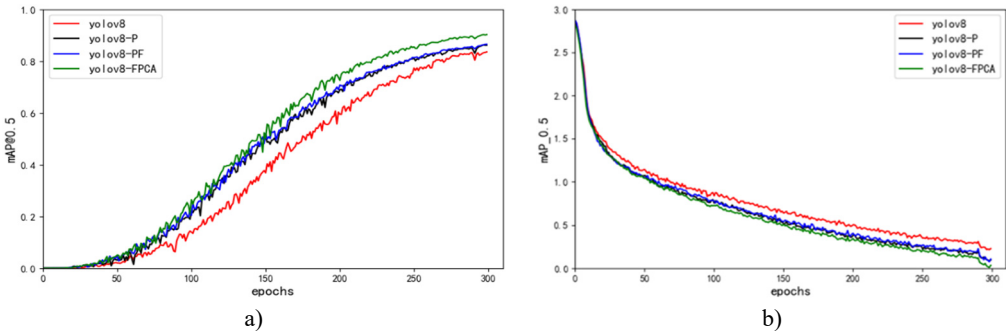
Fig. 7. a) The loss function of the ablation experiment mAP@0.5 plot b)

## 5.4. Attentional mechanism comparison experiment

In order to highlight the enhancement advantage of CA attention mechanism to YOLOv8

algorithm, this paper adds the attention mechanism CABM at the same position of CA attention mechanism in YOLOv8-PFCA for comparison. Using YOLOv8-FPCA and the algorithms YOLOv8-FPCBAM with the addition of CBAM and YOLOv8-FPSE with the addition of SE for comparison, CBAM combines spatial attention and adaptively adjusts the different spatial locations in the feature maps and the importance of the channels. However, CBAM needs to compute the channels and spatial attention, and it needs to align feature maps at different scales, and therefore the computational complexity is high, which can affect the performance of the model. The main advantage of SENet is that it can dynamically adjust the weight of each channel by learning the relationship between channels without introducing too many additional parameters and computational overhead, but the parameters of the SE module need to be optimized through a complex training process, which leads to an increase in training time and computational cost and reduces the generalization ability of the model.

Table 6. Experiments comparing attentional mechanisms

Add-p2	Focal	SE	CA	CBAM	P	R	F1	mAP@0.5	maAP@0.5-0.95	Parameters	GFLOPS	FPS
√	√	×	×	×	87.6 %	76.9 %	0.81	86.5 %	60.9 %	3355433	17.4	43
√	√	√	×	×	88.6 %	79.2 %	0.83	88.4 %	63.5 %	3362336	17.4	45
√	√	×	√	×	90.2 %	81.2 %	0.84	90.2 %	65.6 %	3360824	17.4	43
√	√	×	×	√	88.9 %	79.6 %	0.83	88.2 %	64.3 %	3420034	17.5	41

According to the model evaluation standards, the algorithm models after adding attention mechanisms CA, CABM, and SE have all been improved. As shown in Table 6, when the CABM attention mechanism is added, mAP@0.5 increases by 1.7 % in the data set. After adding the attention mechanism SENet, mAP@0.5 increases by 1.9 %. The CA attention mechanism has the advantages of both CABM and SENet attention mechanisms. Therefore, after adding the CA attention mechanism, the mAP@0.5 of the YOLOv8-FPCA algorithm increases by 3.7 %, and is the best in precision, recall, and F1 score. items, and after adding the CA attention mechanism, the detection speed (FPS), parameter amount, and calculation amount of the algorithm did not decrease in performance. From the experimental results, it is clear that adding the attention mechanism is more beneficial to the model enhancement, and the detection effect of YOLOv8-FPCA is shown in Fig. 8. Coordinate Attention emphasizes the importance of location information and enhances the expression of the model by weighting the location coordinates with attention, which has a better adaptation effect on sewing defect detection applications. This shows that the advantage of Coordinate Attention is its ability to enhance the model's ability to model different locations, which is especially suitable for tasks that require accurate modeling of local details. The heat map of the improved algorithm YOLOv8-FPCA is shown in Fig. 9.
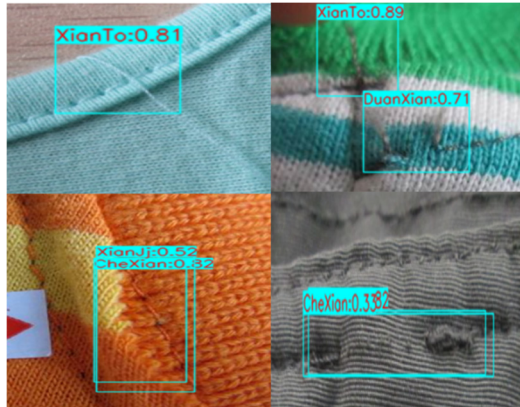
Fig. 8. YOLOv8-CA detection effect
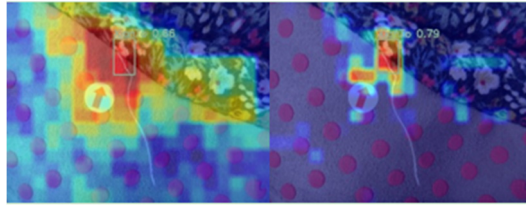
Fig. 9. Improved algorithm heat map display

6. Discussion and analysis

6.1. Contribution of improved YOLOv8 algorithm for garment stitch fault detection

The quality rating of garments is determined by their sewing defects. The traditional method of manual detection of garment sewing defects is inefficient and inaccurate, and the missed defects can cause economic losses to the sales of garments when they reach the market. Research on defect detection based on image pre-processing or a two-stage algorithm cannot meet the needs of industrial inspection. Therefore, this paper proposes an improved YOLOv8 algorithm, which enhances the multi-scale feature fusion of defect information and increases the small target detection head of defect information. It is of great significance to enhance garment quality inspection.

6.2. Deficiencies in work

The homemade dataset in this paper lacks data collection for complex defects in textured backgrounds. This leads to a poor generalization of the improved algorithm to complex backgrounds. In conclusion, the exploration of this paper centers on the improvement and optimization of the algorithm model, and if it is put into practical use in the future, we need to collect more complex data information and improve the generalization ability of the model.

7. Conclusions

In this paper, we improve the basic version of YOLOv8, deepen the depth of the YOLOv8 network model, add small target detection, enhance the ability of multi-scale extraction of feature information of the network model, and focus on the extraction of feature information of small targets of defects. Add attention mechanism in the backbone and head. Enhance the multi-scale feature extraction capability of YOLOv8 to improve detection efficiency. By comparing the attention mechanisms of CA, CABM, and SE, the improved YOLOv8-FPCA algorithm is finally found to be the best for detecting garment stitch faults.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Zengbo Xu: ideas; formulation or evolution of overarching research goals and aims, acquisition of the financial support for the project leading to this publication, development of

methodology; creation of models. management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse, conducting a research and investigation process, specifically performing the experiments, or data/evidence collection. application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data, management and coordination responsibility for the research activity planning and execution, provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.

Yu Chen Bao: management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse, conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.

BingQiang Tian: application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data, management and coordination responsibility for the research activity planning and execution, provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools..

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] W. Y. Li and L. D. Cheng, "New progress in fabric defect detection based on computer vision and image processing," *Journal of Textile Research*, Vol. 3, pp. 158–164, 2014, https://doi.org/10.13475/j.fzxb.2014.03.029

[2] Yao M. J. et al., "Research on clothing ironing target detection algorithm based on YOLOv5," *Journal of Qingdao University (Engineering Technology Edition)*, Vol. 38, No. 1, pp. 24–33, 2023, https://doi.org/10.13306/j.1006-9798.2023.01.003

[3] N. Rezazadeh, M.-R. Ashory, and S. Fallahy, "Identification of shallow cracks in rotating systems by utilizing convolutional neural networks and persistence spectrum under constant speed condition," *Journal of Mechanical Engineering, Automation and Control Systems*, Vol. 2, No. 2, pp. 135–147, Dec. 2021, https://doi.org/10.21595/jmeacs.2021.22221

[4] R. Nima and F. Shila, "Crack classification in rotor-bearing system by means of wavelet transform and deep learning methods: an experimental investigation," *Journal of Mechanical Engineering, Automation and Control Systems*, Vol. 1, No. 2, pp. 102–113, Dec. 2020, https://doi.org/10.21595/jmeacs.2020.21799

[5] R. Ren, T. Hung, and K. C. Tan, "A generic deep-learning-based approach for automated surface inspection," *IEEE Transactions on Cybernetics*, Vol. 48, No. 3, pp. 929–940, Mar. 2018, https://doi.org/10.1109/tcyb.2017.2668395

[6] J.F. Jing, H. Ma, and H.H. Zhang, "Automatic fabric defect detection using a deep convolutional neural network," *Coloration Technology*, Vol. 135, No. 3, pp. 213–223, Mar. 2019, https://doi.org/10.1111/cote.12394

[7] M. S. Biradar, B. G. Sheeparmatti, P. M. Patil, and S. Ganapati Naik, "Patterned fabric defect detection using regular band and distance matching function," in *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, Aug. 2017, https://doi.org/10.1109/iccubea.2017.8463904

[8] R. O. Konda Reddy, B. Eswara Reddy, and E. Keshava Reddy, "Classifying similarity and defect fabric textures based on GLCM and binary pattern schemes," *International Journal of Information Engineering and Electronic Business*, Vol. 5, No. 5, pp. 25–33, Nov. 2013, https://doi.org/10.5815/ijieeb.2013.05.04

[9] Chi-Ho Chan and G. K. H. Pang, "Fabric defect detection by Fourier analysis," in *IEEE Transactions on Industry Applications*, Vol. 36, No. 5, pp. 1267–1276, Jan. 2000, https://doi.org/10.1109/28.871274

[10] X. Yang, G. Pang, and N. Yung, "Robust fabric defect detection and classification using multiple adaptive wavelets," *IEE Proceedings – Vision, Image, and Signal Processing*, Vol. 152, No. 6, p. 715, Jan. 2005, https://doi.org/10.1049/ip-vis:20045131

[11] L. Tong, W. K. Wong, and C. K. Kwong, "Differential evolution-based optimal Gabor filter model for fabric inspection," *Neurocomputing*, Vol. 173, pp. 1386–1401, Jan. 2016, https://doi.org/10.1016/j.neucom.2015.09.011

[12] S. Susan and M. Sharma, "Automatic texture defect detection using Gaussian mixture entropy modeling," *Neurocomputing*, Vol. 239, pp. 232–237, May 2017, https://doi.org/10.1016/j.neucom.2017.02.021

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, Vol. 60, No. 6, pp. 84–90, May 2017, https://doi.org/10.1145/3065386

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014, https://doi.org/10.1109/cvpr.2014.81

[15] R. Girshick, "Fast R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, https://doi.org/10.1109/iccv.2015.169

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137–1149, Jun. 2017, https://doi.org/10.1109/tpami.2016.2577031

[17] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 386–397, Oct. 2017, https://doi.org/10.1109/iccv.2017.322

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, https://doi.org/10.1109/cvpr.2016.91

[19] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2016, pp. 21–37, https://doi.org/10.1007/978-3-319-46448-0_2

[20] Y. C. Bao et al., "Improved garment defect detection algorithm based onYOLOv8 On-line detection system of cloth color difference based on machine vision," *Journal of Donghua University (Natural Science)*, Vol. 50, No. 4, pp. 49–56, 2024, https://doi.org/10.19886/j.cnki.dhdz.2023.0296

[21] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13713–13722, Jun. 2021, https://doi.org/10.1109/cvpr46437.2021.01350

[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42, No. 2, pp. 318–327, Feb. 2020, https://doi.org/10.1109/tpami.2018.2858826

[23] Z. Pourkaramdel, S. Fekri-Ershad, and L. Nanni, "Fabric defect detection based on completed local quartet patterns and majority decision algorithm," *Expert Systems with Applications*, Vol. 198, p. 116827, Jul. 2022, https://doi.org/10.1016/j.eswa.2022.116827

[24] A. Liu, E. Yang, J. Wu, Y. Teng, and L. Yu, "Double sparse low rank decomposition for irregular printed fabric defect detection," *Neurocomputing*, Vol. 482, pp. 287–297, Apr. 2022, https://doi.org/10.1016/j.neucom.2021.11.078

[25] Y. Zhao, K. Hao, H. He, X. Tang, and B. Wei, "A visual long-short-term memory based integrated CNN model for fabric defect image classification," *Neurocomputing*, Vol. 380, pp. 259–270, Mar. 2020, https://doi.org/10.1016/j.neucom.2019.10.067

[26] J. Jing, D. Zhuo, and H. Zhang, "Fabric defect detection using the improved YOLOv3 model," *Journal of Engineered Fibers and Fabrics*, 2020, https://doi.org/10.1177/155892502090826

[27] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," *arXiv: Computer Vision and Pattern Recognition*, 2018.

[28] Q. Liu, C. Wang, Y. Li, M. Gao, and J. Li, "A fabric defect detection method based on deep learning," *IEEE Access*, Vol. 10, pp. 4284–4296, Jan. 2022, https://doi.org/10.1109/access.2021.3140118

[29] X. Shengbao, Z. Liaomo, and Y. Decheng, "A method for fabric defect detection based on improved cascade R-CNN," *Advanced Textile Technology*, pp. 48–56, 2022.

[30] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, https://doi.org/10.1109/cvpr52729.2023.00721

[31] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2018, pp. 3–19, https://doi.org/10.1007/978-3-030-01234-2_1

[32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2011–2023, Jun. 2018, https://doi.org/10.1109/cvpr.2018.00745

Zengbo Xu, Shanghai University of Engineering Science, Associate Professor. Research interests: computer vision and image analysis, 3D body measurement and parametric reconstruction, clothing virtual simulation and personalised customisation, mainly engaged in computer vision and 3D virtual modelling, with a certain research foundation in image analysis, pattern recognition, non-contact 3D body measurement and personalised reconstruction, clothing virtual simulation and haute couture. He has a certain research foundation in image analysis, pattern recognition, non-contact 3D human body measurement and personalised reconstruction, clothing virtual simulation and haute couture.

Yuchen Bao Shanghai University of Engineering Science, Graduate Student. Research interests: computer vision and image analysis.

Bingqiang Tian Shanghai University of Engineering Science, Senior Laboratory Technician.