A Fabric Defect Detection Method Based on Deep Learning

QIANG LIU, CHUAN WANG, YUSHENG LI, MINGWANG GAO, AND JINGAO LI

School of Mechanical Engineering, Shandong University of Technology, Zibo 255049, China

Corresponding author: Chuan Wang (17852030514@163.com)

ABSTRACT Fabric defect detection is a challenging task in the fabric industry because of the complex shapes and large variety of fabric defects. Many methods have been proposed to solve this problem, but their detection speed and accuracy were very low. As a classic deep learning method and end-to-end target detection algorithm, YOLOv4 has evolved rapidly and has been applied in many industries, showing good performance. This paper proposes an improved YOLOv4 algorithm with higher accuracy for fabric defect detection, in which a new SPP structure that uses SoftPool instead of MaxPool is adopted. The improved YOLOv4 algorithm with three SoftPools can process the feature map effectively, which has a significant advantage in reducing the negative side effects of the SPP structure and improving the detection accuracy. The improved SPP structure is used by the three outputs of Backbone, and in order to ensure that the output can be inputted into the subsequent PANet successfully, the network structure is improved that a series of convolution layers after the SPP structure is added for reducing the channel numbers of feature map to an appropriate value. In addition, contrast-limited adaptive histogram equalization is adopted in advance to improve the image quality, which results in strong anti-interference abilities and can slightly increase the mAP. Experimental results show that, compared with the original YOLOV4, the improved YOLOv4 increases the mAP effectively by 6%, while the FPS only decreases by 2. The improved YOLOv4 can identify the location of defects accurately and quickly, and can also be applied in other defect detection industries.

INDEX TERMS Fabric defect detection, deep learning, YOLOv4, SoftPool, detection accuracy.

I. INTRODUCTION

Fabric defect detection is an important step in the fabric production process. Human inspection with eyes for fabric defects is the traditional method used in the fabric industry [1], and visual inspections can identify and locate the defects. However, the human detection rate only reaches up to 12 meters per minute [2], and is a monotonous job with high repetition, a wasteful use of human resources and increasing costs, making it unsuitable for use in mass production. Although human detection is simple, the cloth production lines and cloth outputs are becoming more complex, which creates obstacles for workers in correctly identifying the location of defects. Additionally, cloth defects lead to a reduction in cloth prices, resulting in losses of 45%-65% [3]for the cloth manufacturer. Therefore, a new detection method, which has high detection accuracy and detection speed, is needed to replace the manual work currently used. With the advent of CNN (convolutional neural networks), and the

development of deep learning and machine vision [4], many detection methods, combining the advantages of deep learning and machine vision, have emerged, replacing traditional manual methods and image processing.

In this paper, first, the K-means clustering algorithm is analyzed to select the optimal anchor value according to the aims of this paper. Then, the structure of the SPP-net of the original YOLOv4 is analyzed, and its disadvantages are summarized. In addition, addressing the disadvantages, an improved YOLOv4 algorithm with improved SPP (Spatial Pyramid Pooling) and CLAHE (contrast-limited adaptive histogram equalization) is proposed. Finally, the improved YOLOv4 is compared with four classic algorithms: YOLOv4 with CLAHE, the original YOLOv4, Faster R-CNN and SSD (Single Shot MutiBox Detector), where the mAP is the main performance evaluation index. In addition, defect images and detection effect images are presented, which have clear contrasts.

The anchor of YOLO and SSD learns from Fast R-CNN [5]. The final prediction part of YOLOv4 is carried out in three feature maps, each of which has a grid point with

---

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

three anchors. YOLO obtains the width and height parameters of the BBox (bounding box) in advance. In the regression prediction phase, only the width and height parameters need to be adjusted instead of rebuilding the BBox. So the setting of good anchor parameters can accelerate the rate of convergence of the network training. The K-means clustering algorithm is used in YOLOv2 to process the Gt BBox (ground truth bounding box). The original K-means clustering generally uses a Euclidian distance to complete the clustering but a large target will result in a large Gt BBox that can lead to a large anchor, which can enlarge the Euclidian distance of a large box. The situation is obviously unreasonable, so the processed IOU is used in YOLOv2 instead of the Euclidian distance. The processed IOU represents the degree of overlap between each cluster center and other boxes. The original IOU indicates that the value of the IOU grows larger as the degree of overlap is higher, but the Euclidian distance that is expected in this paper is smaller when the degree of overlap is higher.
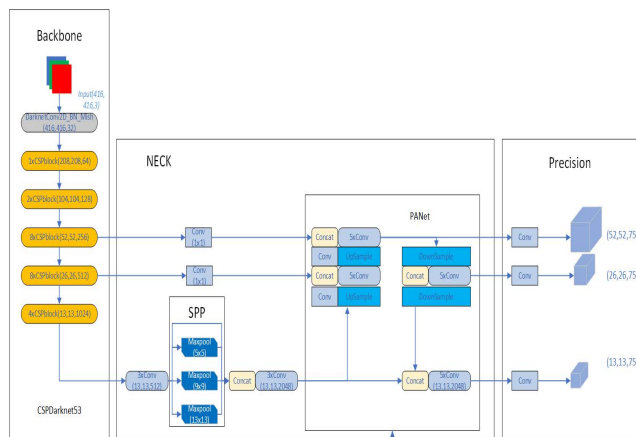

FIGURE 1. Structure of the original YOLOv4.

The network structure of YOLOv4 is shown in Figure 1, from which CSPdarknet53 is the backbone that references the residual blocks from ResNet [6], and combines CSPNet [7] with Darknet53 of YOLOv3. YOLOv4 also references the ideas of SPP-Net and FPN-Net, which are improved to be more suitable for defect detection.

The remainder of this paper is organized as follows: Section II reviews deep learning and the YOLOv4 algorithm. Section III describes the improved YOLOv4 algorithm and image enhancement. Section IV gives the experimental results of the improved YOLOv4 compared with three other algorithms: SSD, Faster R-CNN and the original YOLOv4. Section V draws the conclusions and summarizes some important points about this paper.

II. RELATED WORKS

YOLO (You only look once) [8] is an end-to-end neural network algorithm that has been continuously improved from YOLOv1 to YOLOv5. Numerous research papers have shown that YOLO has better speed and accuracy than the

other algorithms. FPS (Frames Per Second) and precision are considered comprehensively. An improved algorithm based on YOLOv4 is proposed in this paper, where images from the dataset are enhanced in advance, and image processing is combined with deep learning to improve detection results.

Bo *et al.* proposed the machine vision technique in which defects are detected by the Gabor filter, which is based on image processing, however, it has poor detection results for some types of defects [9]. Wiener filter is used to classify defective images by converting RGB images into binary images to improve the detection effect [10]. In addition, there are other methods to detect fabric defects. For example, Kazim *et al.* adopts a thermal-based defect classification method with K-nearest neighbor algorithm [11] and dimensionality reduction [12] to classify textile defects respectively. Image processing [13] and thermal images [14] are also used in defect detection. However, image processing and thermal image can only solve the classification problem. For these methods, the defects in the images are obvious, and defects can only be identified and can not be located correctly. Most of the traditional image processing algorithms have the shortcomings that only simple background and large object images can be processed effectively. So the neural network based methods are been studied by some researchers. Ouyang *et al.* [15] used an activation layer embedded convolutional neural network to detect defects. Liu *et al.* [16] combined image processing with deep learning and proposed a method in which image enhancement is implemented prior to using convolutional networks; accuracy was improved. Li [17] adopted focal loss [18] in ResNet50 [19] to solve the problem of uneven numbers of positive and negative samples. Although the above algorithms are feasible, there are some disadvantages; some have slow recognition speeds and others have low recognition precision. Faster R-CNN [20] is the most commonly used algorithm in fabric defect detection. Liu redesigned ROI (region of interest) pooling to consider the global features of images [21]. Li [22] used an improved multiscale detection algorithm, I-FPN, to improve the detection effect for small targets. Zhao [23] proposed an improved NMS (Non-Maximum Suppression) that considers interclass similarities in the detection process. Faster R-CNN, and its improvements, have been adopted by many researchers to increase the efficiency of detecting small targets. In general, the following are reasons for using Faster R-CNN. First, fabric defects have their own remarkable features compared to common defects. Among all kinds of fabric defects, there are some that account for a relatively large proportion in the images, such as WEFTS, WARPS, STAINS, FLOATS and CRACKYWEFTS, which is usually a very large spot, or the spot has the same width as the image. Other defects have small shapes (perhaps only a few pixels), such as NEPS, HOLE, SNAGS and KNOTS. Second, detection and recognition are relatively simple for large defects. However, detecting defects is very difficult for small targets, especially holes with only a few pixels, because the uneven number of positive and negative samples makes it difficult for a one-stage

network to learn the features of small targets. Although this problem can also exist in a two-stage network, a large number of negative samples has a small impact on detection because the region proposal network (RPN) eliminates many negative samples. To solve the imbalance of the number of positive and negative samples in a one-stage network, Kaiming He proposed the focal loss method, which can reduce the weight of large sample losses and increases the weight of small samples in total loss. However, focal loss is not effective in practical applications, and evenly reduces mAP (mean average precision) [24]. Therefore, the precision of a two-stage network is generally higher than its one-stage network counterpart. However, a two-stage network represented by Faster R-CNN is generally as lower than the others. Therefore, the two-stage network is not adopted in this paper.

As one of the typical first-order algorithms, YOLO has been improved over many generations. YOLO algorithms have developed rapidly, including YOLOv1 with various limiting accuracy problems, which was further improved in YOLOv2 [25] and YOLOv3. Compared with YOLOv1, YOLOv2 and YOLOv3, YOLOv4 has better performance and uses tricks [26] to improve the accuracy. For example, mosaic data augmentation, MISH activation function [27], K-mean clustering algorithm [28], FPN-net [29], PAN-net [30], SPP-net [31] and CSPdarknet53 are adopted as backbones. In addition, ordinary researchers can use a 1080Ti GPU to train a YOLOv4 model, which is beneficial to many scholars and is convenient for industry applications. Factories do not need to spend a large amount of time buying expensive hardware. Compared with the second-order algorithms, one-order algorithms represented by YOLO can meet the requirement for real-time detection [32].

For the original YOLOv4, if 416×416 images are used as the input of the backbone, the generated sizes of the feature maps are (52,52,256), (26,26,512) and (13,13,1024) [33], and will revert to the original size in the final prediction. The bottom feature map of the backbone of the original YOLOv4 is (13,13), which passes the SPP structure after three convolutions. The original SPP structure of YOLOv4 is different from SPP-net, which has only three maximum pooling branches, and the results are converted to a one-dimensional vector, which then carries out full connecting classification. The SPP structure in the original YOLOv4 uses four branches. In the final output, the results of the four channels are superimposed, and the final number of channels is quadrupled. After three convolutions, the number of channels is reduced and then output to the FPN structure.

The original SPP uses Maxpool that maxs pooling for each part, and solves the problem that the size of images inputted into CNN should be fixed. The SPP structure can enlarge the receptive field effectively, and the context features can be obtained more comprehensively by combining a pooling layer with different kernel sizes. The SPP structure has always been an excellent solution for classification and detection problems before the CNN appeared in [34]. The mAP of YOLOv3-SPP is higher than that of its counterpart without an

SPP structure [35]. The pooling layer is the key component of the CNN network, because the parameters required by the network are greatly reduced, which can increase the receptive field of subsequent convolutions [36], [37]. Most frameworks adopt Max Pooling or AVG Pooling (average pooling). For example, the SPP structure adopts three max pooling filters with different kernel sizes, but max pooling will cause some problems when selecting the maximum value from a specific range (such as 3×3). Although max pooling can reduce the number of parameters, a large amount of information will be lost in the selection process. In addition, you do not know whether the information of the background or the target is lost because it only selects the point with the most obvious features as the representative of the neighborhood. If the background is similar to the target, it is easy to lose useful information. In the detection of fabric defects, as the defects are very similar to the background, the use of max pooling in SPP will affect the detection performance and has the risk of losing important features. In contrast, although AVG pooling takes into account all the features of the neighborhood and retains more background information, the target feature intensity of the region will be reduced, and the obvious features will be ignored after the average is taken.

III. THE IMPROVED YOLOV4 AND IMAGE ENHANCEMENT
If a large amount of max pooling is used, such as an SPP structure, the accuracy of the defect location will be affected, which is very disruptive for the classification task of fabric defects in which the targets are similar to the backgrounds. It is likely, that owing to the use of maximum pooling, YOLOv4 with three SPP structures has a lower mAP than its counterpart with one SPP. The improved structure of SPP is shown in Figure 2.
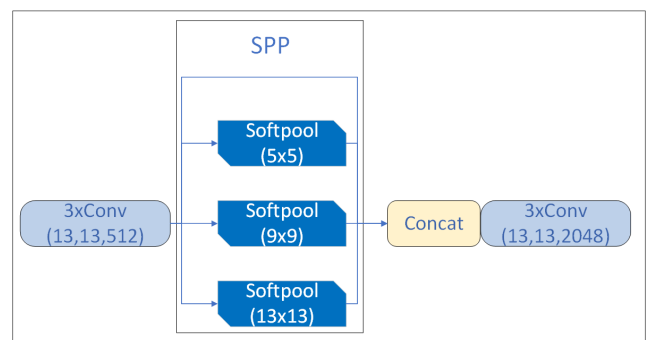

FIGURE 2. The structure of the improved SPP.

The improved SPP uses Softpool that selects the feature map in proportion to the corresponding value of the element, which avoids the advantage of Maxpool that misses some details. In addition, the improved SPP is adopted after the three feature maps are output by the backbone; the structure is shown in Figure 3.

The improved YOLOv4 adopts the improved SPP based on softmax, and three convolutions are added to the upper and middle of the improved SPP that uses gradient information to
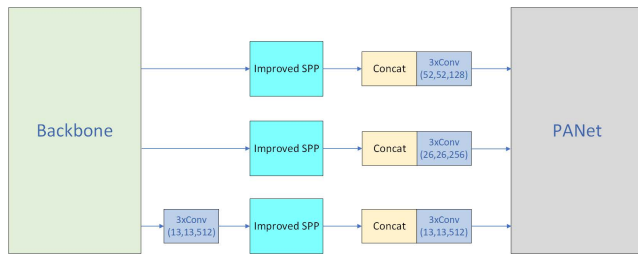
FIGURE 3. The structure of the improved YOLOv4.

improve the training effect, which increases the computation speed and reduces the running memory.

In order to input PANet (Path Aggregation Network) into all the feature maps with the improved SPP structure, three convolutions are added to the upper and middle of the improved SPP, and the results are adjusted to (52,52,128) and (26, 26,26,256). To reduce parameters and improve network speed, 1×1Conv from the two backbone branches is removed, which allows the data entered into the improved SPP structure to contain more detailed information.

Compared to the original YOLOv4, the improved network upgrades mAP by approximately 6%. Soft pooling [38], as proposed by Stergiou, Poppe and Kalliatakis, can effectively solve the problem of missing details in max pooling. Soft pooling is similar to stochastic pooling [39] and S3Pool [40], the latter two of which adopt the idea that the elements of the feature map are selected according to the probability value, namely, that the probability selected is larger with a larger value of the element, which is not like max pooling, in which only the element with the largest value is selected. This selection method can result in a random effect that is good at times and also bad at times. Therefore, a better and faster soft pooling based on softmax [41] is adopted to replace max pooling in this paper; the mathematical expression is as follows:

$$\hat{a} = \sum_{i \in R} w_i * a_i \tag{1}$$

where w_i is the weight: $w_i = \frac{e^{a_i}}{\sum_{j \in R} e^{a_j}}$ The significance of the weight is that the larger the value of this point is, the higher the influence on the result.

The advantage of soft pooling is that the method of filling in 0 in soft pooling is different from soft pooling in the back propagation phase, which differentiates it in having gradient information to improve the training effect. Furthermore, the running speed and memory occupied are less than those in max pooling.

In this paper, similar to the *IOU* loss, $1 - IOU$ is adopted to replace the Euclidean distance, which is mathematically expressed as follows:

$$distance\ (i, center) = 1 - IOU\ (i, center) \tag{2}$$

where *distance* $(i, center)$ represents the Euclidean distance, and *IOU* $(i, center)$ represents the degree of overlap between the cluster center and the box. In this paper, the width and

height parameters of the Gt BBox detected for fabric defects are special, which is very different from the default anchor of the COCO dataset in YOLOv4, because those boxes only represent the anchor values of 20 classes. Therefore, a K-means is executed several times and obtains many groups of parameters, among which the width and height parameters are averaged to obtain the anchor that is most appropriate for a regression. The width and height of the 9 anchors are (6.8,10.4), (8.7,355.3), (10.4,92.3), (14.6,19.5), (19.7,398.7), (38.8,42.0), (78.7,111.0), (88.5,13.0) and (410.0,34.0), respectively.

Part of the dataset selected in this paper is from Aliyun-FD-10500, and the remainder is from Kaggle and the actual photograph. The defect pictures and categories are shown in Figure 4.

In Figure 4, there are many serious defects, such as the WARP or WEFT defects. These defects, SPLICING, THICK PLACE, THIN PLACE, NEEDLE_LINE, COARSE END and COARSE PICK, have similar shapes, which are along the direction of the warp and weft, and are long, thin and slight. Therefore, the above defects are collectively called LINE defects in this paper. They often appear in cloth, and are considered a class in our classification task. FLOATS and LADDER defects are less common than LINE defects but are very destructive and almost impossible to repair or remove, and have distinct features. Viewed from this angle, FLOATS and LADDER defects are also collectively called FLOATS. HOLES, KNOTS, NEPS and SNAGS are also commonly found in fabric but they have small areas, perhaps only a few pixels, accounting for a small proportion of the entire cloth, which makes them difficult to detect. Because the features among these defects are similar, these defects are collectively called HOLE defects. In addition, the defects of COLOUR BLEEDING, DYE MASKS and OIL STAIN are similar to each other, and are collectively called STAINs. In general, the above classification method meets the practical requirement that a precise defect class need not be recognized, and only the general classification is enough; that is, detecting the defect is crucial and is differentiable from the defects crease and shade.

The reclassified dataset is flipped and blurred to increase the total number of samples. The number of defect classes is shown in Figure 5.

Figure 5 refers to the number of defects for the four classes, not the number of images, because some images may belong to different defects. The HOLE defects occupy a much smaller area of pixels than other defects and are not easy to detect. Therefore, HOLE defects should be given more attention, and many HOLE defects are adopted in this paper.

Images from the dataset are enhanced to highlight the contrast, which obviously improves the detection accuracy [42]. For image enhancements, many relative methods attempt to highlight the form of defects in the background; the Canny edge detection operator and Gabor filter are adopted to enhance images [43]. However, there is no optimal method for parameter selection. Finally, the method contrast-limited
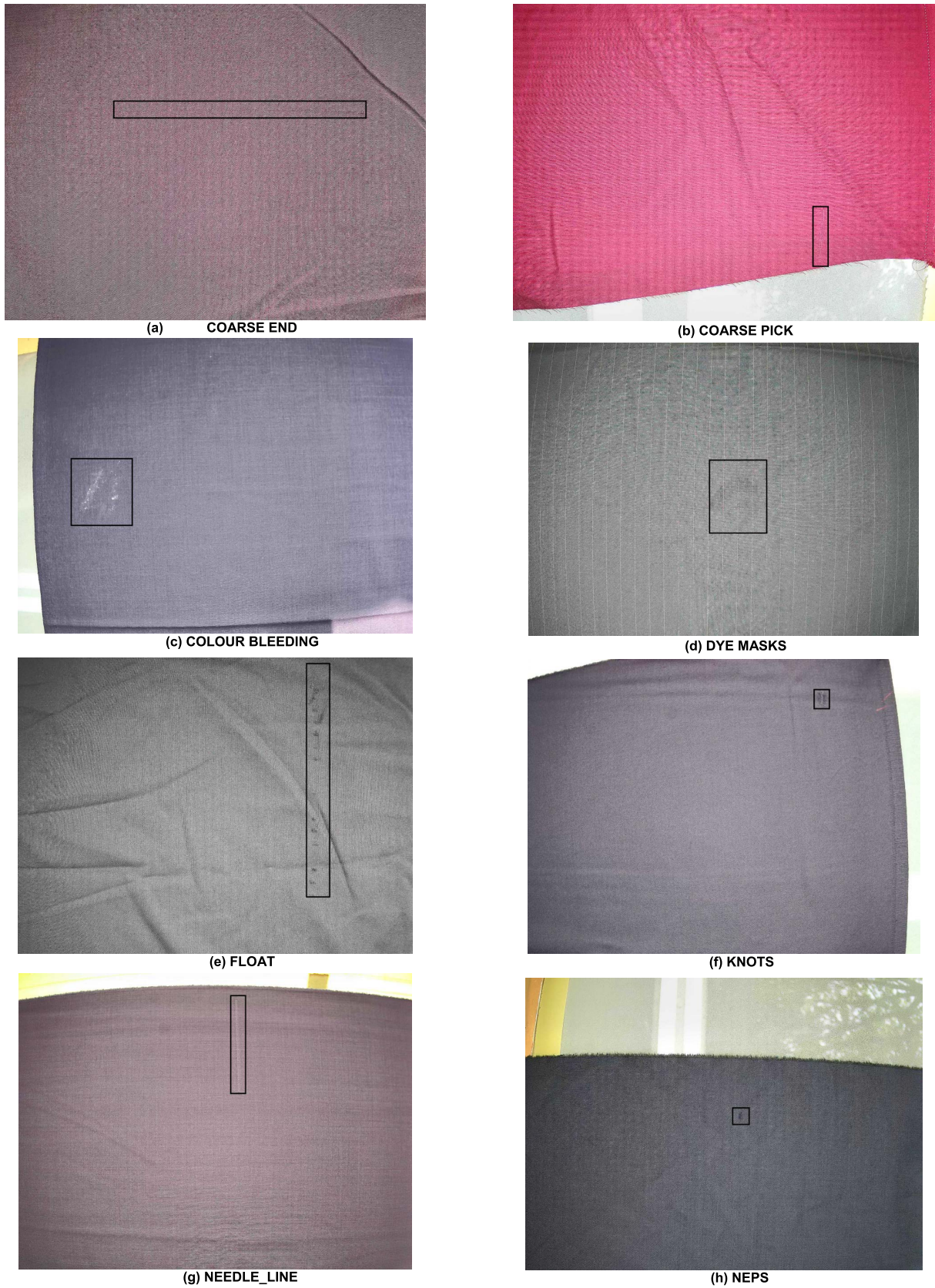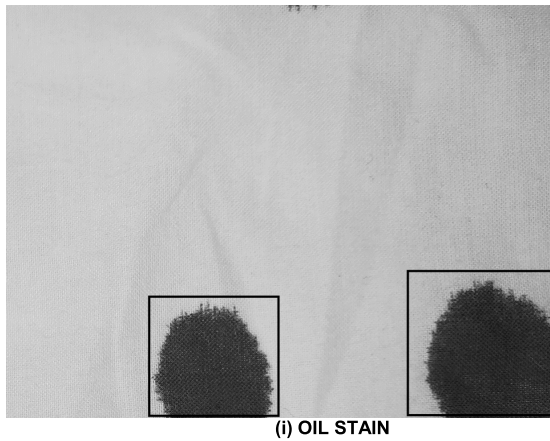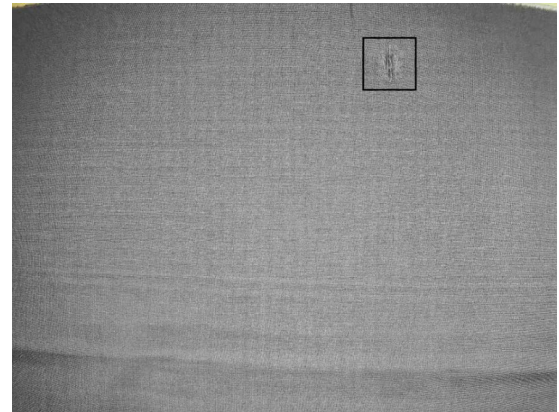
FIGURE 4. Classification of fabric defect.

(i) OIL STAIN

(j) SNAGS

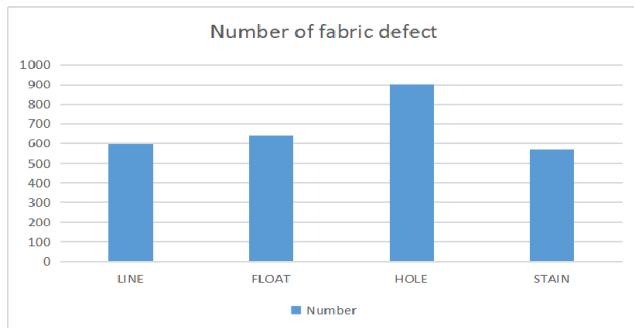**FIGURE 4.** *(Continued.)* Classification of fabric defect.

FIGURE 5. Numbers of fabric defect classes.

adaptive histogram equalization is adopted, which results in an improvement in the mAP of approximately 0.6%.

We carried out image enhancement on the images of the dataset to highlight the contrast, and the image enhancements can also slightly increase the detection accuracy. For image enhancements, we attempted numerous enhancement methods to highlight the shape of the defect in the background. We tried to use the Canny edge detection operator and the Gabor filter to enhance the image but it is very difficult to find an optimal method for parameter selection. Finally, we adopt contrast-limited adaptive histogram equalization. The image enhancement improves the mAP by approximately 0.6%.

CLAHE is an enhanced version of HE (histogram equalization), which is a simple and effective image enhancement method and can increase the contrast of images by adjusting the image's gray distribution. If the pixel value of the original image is relatively concentrated in the gray histogram, HE can increase the range of gray differences between pixels to obtain a clearer image [44]. The effect of HE is shown in Figure 6.

Although the images processed by HE are clearer and have stronger contrasts, some of the gray levels of the enhanced image will disappear, which leads to the loss of detail information. Moreover, HE can result in excessive enhancements [45], such as some regions where the brightness is very
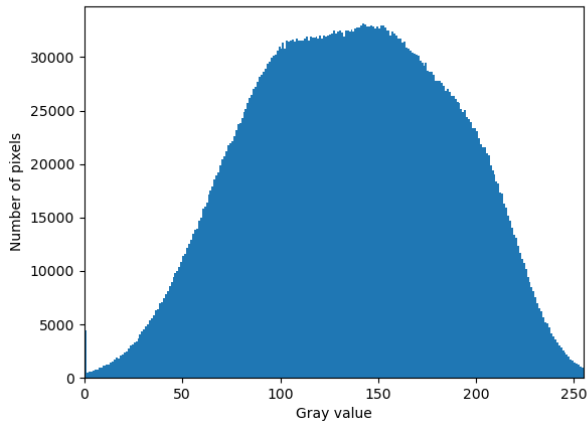

(a) The original image
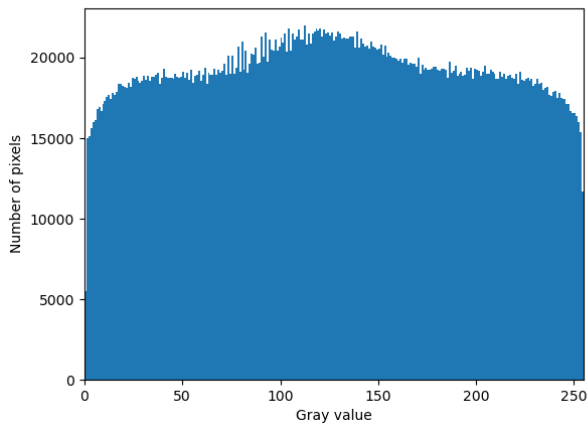

(b) The image intensification

FIGURE 6. Comparison of results of processed images.

low and the contrast is too large and become noise points after images are processed by HE.

In Figure 7, the abscissa represents the gray value with the range from 0 to 255, and the ordinate represents the number of pixels with the gray value. Figure 7 (a) is the gray

(a) The original gray level histogram

(b) The improved gray level histogram

FIGURE 7. Gray level histogram comparison.

(a) Histogram Equalization

(b) Contrast Limited Adaptive Histogram Equalization

FIGURE 8. The difference between HE and CLAHE.

histogram of the original image, similarly, Figure 7 (b) is the gray histogram after CLAHE. The comparison shows that the gray value of Figure 7 (b) is more average and the contrast is improved.
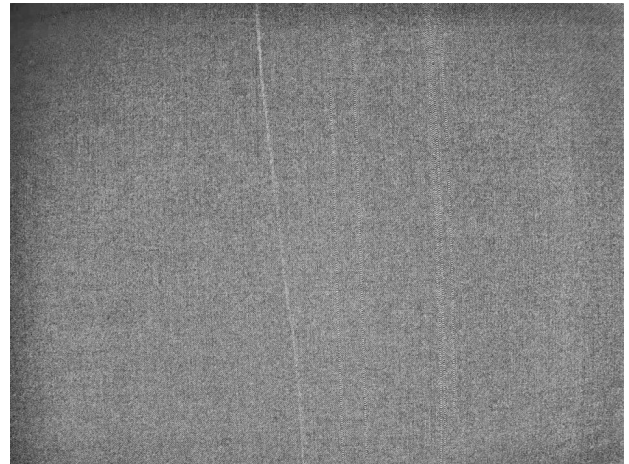
From the comparison, it can be concluded that the gray level histogram of image processed by HE has more evenly distribution in the entire range, namely that the values of pixels in the original image are redistributed and the numbers of pixels within certain grey bounds are roughly the same.

The effects of HE and CLAHE processing on images are shown in Figure 8. CLAHE has a positive effect, and is adopted in this paper. Most of the image pixels from the dataset in this paper are 2560 × 1920 or 1984 × 1488, which were processed by OpenCV. First, the input images are grayed, and the three channels are processed by CLAHE if the colorful images are used. However, the effect of the processed images is not good. For the lighting conditions where the original image was taken, improving the contrast of the three channels leads to the appearance of a colored aperture in the background, which can affect the training and recognition of STAIN defects.

Based on the above analysis, the following plan is adopted in which the grayscale image is used as input, which is then

processed by Gaussian filtering to eliminate noise, and then further processed by CLAHE to obtain the final images.

IV. EXPERIMENT

The deep learning framework adopted in this paper is PyTorch 1.2 based on the Windows 10 platform. The computer configuration for this experiment is as follows: CPU: Inter Xeon W-2245 with 3.90 GHz, GPU: NVIDIA Quadro P4000, video memory: 8 GB, RAM: 64 GB, SSD: 512G. The programming language is Python3.6, and OpenCV is adopted for image processing. The parameters of image size of the backbone, the learning_rate and the decay_rate are set to 416 × 416, 0.001 and 0.93, respectively. A learning rate of 0.93 indicates that the learning rate of each epoch becomes 0.93 times that of the previous epoch. One thousand epochs are trained, and the network trains the NECK part for 500 epochs by freezing, and the backbone epochs by unfreezing. The setting for the parameter batch_size has been adjusted several times based on the theory that the optimal batch_size is between 2 and 32 [46]. batch_size is set to 16 before freezing and then reset to 4 in consideration of the computer configuration after thawing. The improved SPP architecture requires testing to measure its improvement;

some of the tricks, such as mosaic data augmentation and cosine dependency schedulers [47], run well sometime and are sometime unstable. After a thousand iterations, the loss value fluctuates by approximately 0.4, which proves that the model proposed in this paper has good performance and has converged.

The detection results for fabric defects are displayed by the index mAP in the VOC dataset to evaluate the accuracy of the model; the visualization of each detection box is not evaluated. Two indices, precision and recall, are often used to evaluate the performance of target detection models, before which they are adopted, and the threshold value of IOU should be preset. When the degree of overlap between the prediction box and the real box is greater than the threshold value of IOU, the corresponding sample is called a positive sample; Alternatively, the corresponding sample is called a negative sample. The mathematical expression of the two indices, precision and recall, is as follows:

$$precision = \frac{T_P}{T_P + F_P} \times 100\% \qquad (3)$$

$$recall = \frac{T_P}{T_P + F_N} \times 100\% \qquad (4)$$

where T_P is the number of real positive samples, F_P is the number of false positive samples, F_N is the number of false negative samples.

Each of the two indices cannot evaluate the performance separately, so the index AP is introduced, and the mathematical expression is as follows:

$$AP = \int_R P(R)dR \qquad (5)$$

where P indicates precision, R is recall and AP is the curve integral of PR. AP represents the degree, which is the result of the model's prediction in the classification task. To obtain the overall prediction degree, the mAP index needs to be calculated. The formula is as follows:

$$mAP = \frac{\sum_{n-1}^{N} AP(n)}{N} \qquad (6)$$

where N is the total number of classes, and mAP is between 0 and 1. The larger the mAP, the better the model performance. The IOU threshold set in this paper is 0.5, which is equivalent to AP.5 and is called AP50 in the COCO dataset.

In this paper, three performance indices, mAP, precision and recall, are evaluated. In total, five algorithms, YOLOv4 with improved SPP and CLAHE, YOLOv4 with CLAHE, the original YOLOv4, Faster RCNN and SSD [48], are compared, and the results are shown in Table 1. For mAP and FPS of the original and improved YOLOv4 are emphasis, and theoretically, the FPS of SSD and Faster R-CNN must be less than the corresponding part of YOLOv4, so the FPS of these two are not tested.

Table 2 and Table 3 show that, compared with the original YOLOv4, the improved YOLOv4 proposed in this paper has improved greatly. For the small fabric defect HOLE, the improved YOLOv4 greatly increases the recall and precision.
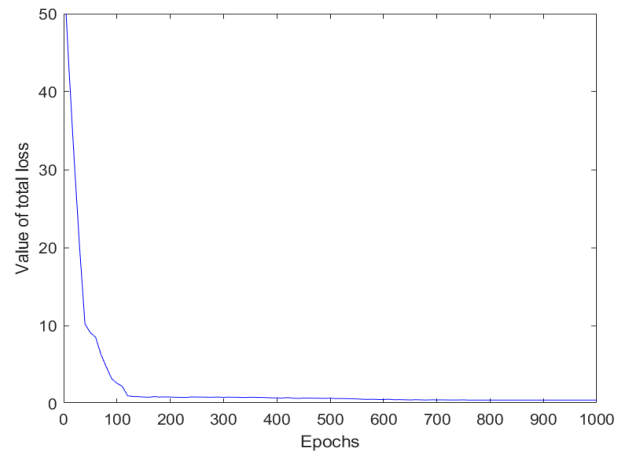
(a) The Value of Total Loss
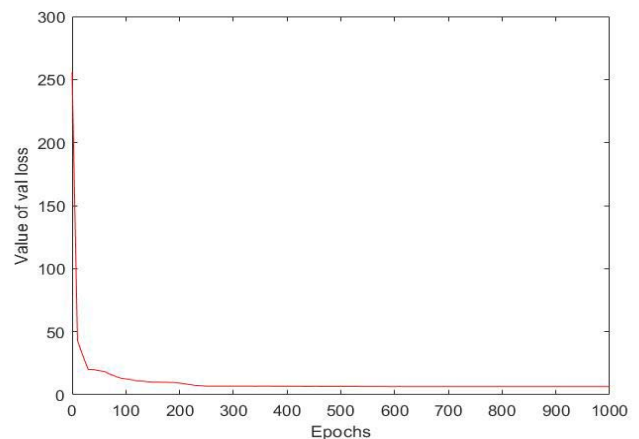
(b) The Value of Val Loss

FIGURE 9. The images of loss function.

TABLE 1. Experimental results of five algorithms for all classes.

Detection algorithm	mAP/%	FPS
The improved YOLOv4	86.5	21.2
The improved YOLOv4 (without CLAHE)	86	21.2
YOLOv4	80	23.7
SSD	70.4	x
Faster R-CNN	73.6	x

However, neither the original YOLOv4 nor the improved YOLOv4 effectively improved recall, which suggests that the defect LINE is similar to the background, which means that BBox has a low degree of overlap with GT BBox. Although it correctly detects the target, resulting in an excessively low recall. Most importantly, YOLOv3-SPP3 with three SPP structures has a mAP increase of only 0.9%, compared with YOLOv3-SPP1 with one SPP structure; however, the rate of reasoning decreases by 10 ms [49]. In striking contrast, the

(a) The original image

(b) The image processed by SSD

(c) The image processed by Faster R-CNN

(d) image processed by the original YOLOv4

(e) The image processed by the improved YOLOv4
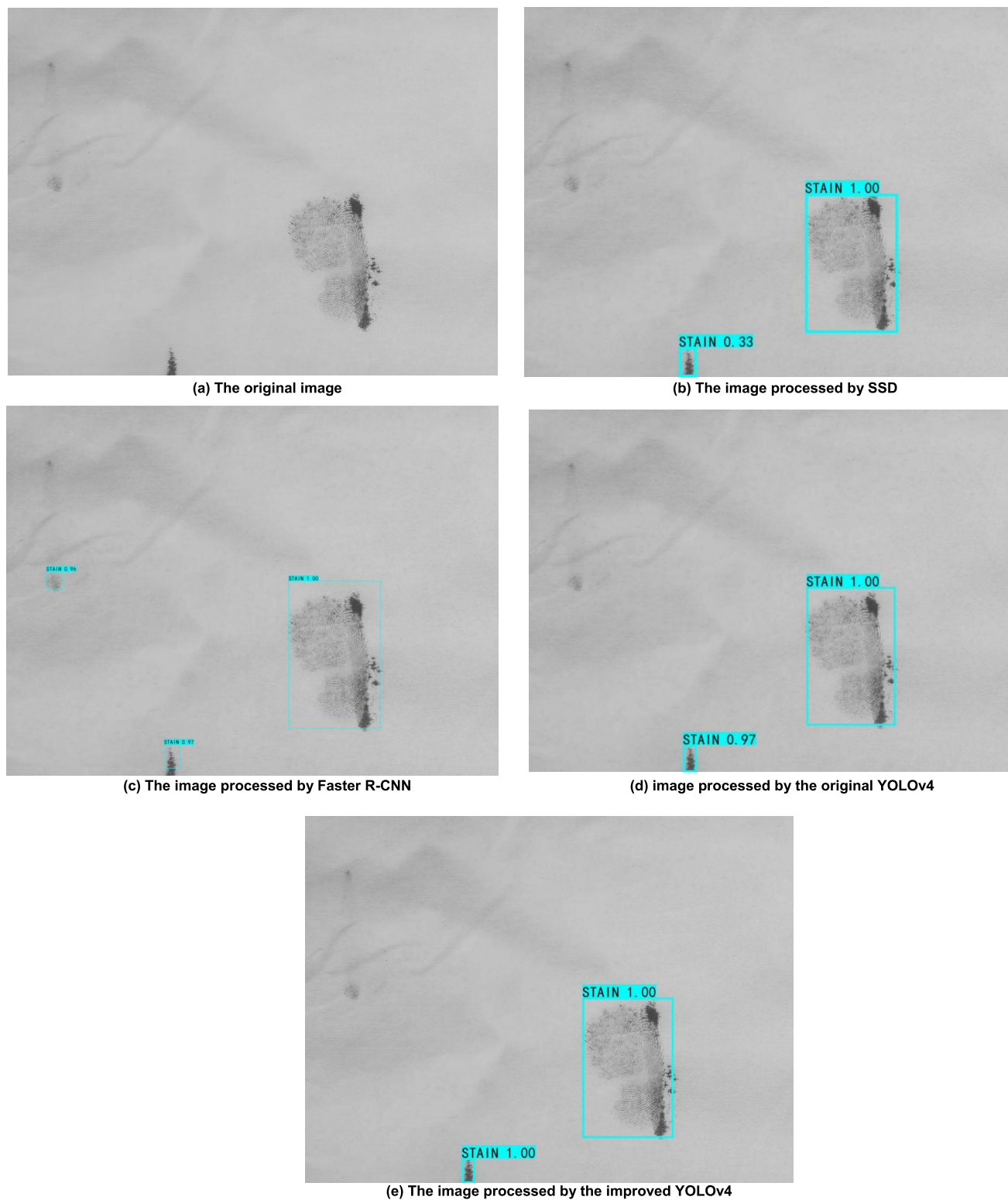
FIGURE 10. Detection results of each algorithm.

improved YOLOv4 with three SPP structures can improve mAP while decreasing FPS, with small changes in amplitude.

The loss function curve of the improved algorithm is shown in Figure 9.

(a) The image processed by the original YOLOv4

(b) The image processed by the original YOLOv4

(c) The image processed by the original YOLOv4

(d) The image processed by the original YOLOv4

(e) The image processed by the improved YOLOv4

(f) The image processed by the improved YOLOv4

(g) The image processed by the improved YOLOv4

(h) The image processed by the improved YOLOv4

FIGURE 11. Detection results of the original YOLOv4 and improved YOLOv4.

TABLE 2. Detection performance of the improved YOLOv4 for each class on the validation set of our dataset.

Class	AP/%	Precision/%	Recall/%	F1/%
HOLE	89.60	96.05	90.12	93.00
LINE	69.96	97.06	70.21	81.00
FLOAT	99.84	98.48	99.90	99.00
STAIN	86.38	98.11	86.67	92.00

TABLE 3. Detection performance of the original YOLOv4 for each class on the validation set of our dataset.

Class	AP/%	Precision/%	Recall/%	F1/%
HOLE	84.72	85.37	86.42	86.00
LINE	63.39	85.71	63.83	73.00
FLOAT	99.80	98.40	99.90	99.00
STAIN	83.17	94.34	83.33	88.00

After 1000 iterations, the final total loss of the model stabilized at approximately 0.4, and the loss of the value set stabilized at approximately 6. The loss of the model shows a downward trend, which proves that our model is learning. In the later stage, we used SGD to manually adjust the learning rate to further reduce the loss. The detection effects of all the algorithms are shown in Figure 10.

The above detection results show that whether the area of defect is big or small, the accuracy of the improved YOLOv4 has been improved a lot compared with the other algorithms. And it can be concluded that compared with the other first-order and second-order algorithms, YOLOv4 owns excellent performance.

The detection results of the original YOLOv4 and the improved YOLOv4 are shown in Figure 11.

For the detection time for the same image, the original YOLOv4 takes 42ms while the improved YOLOv4 takes 48ms, which are both short. So video streaming from PC computer can be detected easily by 20FPS.

The above detection results further proves that the improved YOLOv4 owns stable and higher accuracy, such as for (D) and (H), the original YOLOv4 and the improved YOLOv4 has the same detection accuracy 0.98. But for the other defects, the improved YOLOv4 to varying degrees increases the detection accuracy.

V. CONCLUSION

The model proposed in this paper has been improved for small cloth defect areas that are close to the background shape, and has overcome the low efficiency of the traditional manual detections. Compared with the original YOLOv4, the improved YOLOv4 upgrades mAP by 6%, while FPS only decreases 2. Based on extensive research and a detailed analysis of fabric defects, our network model is the most suitable for fabric defect detection. First, we redivided the anchor according to the characteristics of the defects and took average values to make the anchor more suitable for the application. Then, we used CLAHE to process the image to remove irrelevant color information and highlight the contrasts in the image. Finally, we improved the SPP structure by using soft pooling instead of max pooling. The improved SPP structure is used for each feature map. In fact, SPP structure greatly improves the performance of YOLO model. From YOLOv3-SPP to YOLOv4, YOLOv5 and YOLOX, these models all use SPP structure. In YOLOX, SPP structure is put into the backbone. The improvement for the SPP structure in this paper can also be applied in the other YOLO models and theoretically owns the similar effect. Defects from different materials often own many similar characteristics. Cracks on steel and road surface are similar to LINE and FLOAT in fabric defects, and shadows in medical images are similar to HOLE and STAIN in fabric defects. Therefore, the improved YOLOv4 proposed in this paper can be extended to other surface defect detection fields, among which batch detection of wood surface defects is one example. The model proposed in this paper has high precision and excellent real-time performance, and can be effectively applied for defect detections in industrial fabrics.

REFERENCES

[1] J. Wu, J. Le, Z. Xiao, F. Zhang, L. Geng, Y. Liu, and W. Wang, "Automatic fabric defect detection using a wide-and-light network," *Int. J. Speech Technol.*, vol. 51, no. 7, pp. 4945–4961, Jul. 2021.

[2] H. P. Liu, "Research on nonwoven fabric defect online detection system using machine vision," M.S. thesis, Dept. Control Theory Control Eng., Huazhong Univ. Sci. Technol., Wuhan, China, 2015.

[3] K. Srinivasan, P. H. Dastoor, P. Radhakrishnaiah, and S. Jayaraman, "FDAS: A knowledge-based framework for analysis of defects in woven textile structures," *J. Textile Inst.*, vol. 83, no. 3, pp. 431–448, 1992.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, Sep. 2015.

[5] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[7] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1571–1580.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[9] Q. U. Bo, L. U. Zhaoyang, L. I. Jing, L. Cui, and F. Zhou, "An improved multichannel Gabor filter algorithm for fabric defect detection," *J. Textile Res.*, vol. 30, pp. 37–40, Jan. 2009.

[10] K. Yıldız, Ö. Demir, and E. E. Ülkü, "Fault detection of fabrics using image processing methods," *Pamukkale Univ. J. Eng. Sci.*, vol. 23, no. 7, pp. 841–844, 2017.

[11] K. Yıldız, A. Buldu, and M. Demetgul, "A thermal-based defect classification method in textile fabrics with K-nearest neighbor algorithm," *J. Ind. Textiles*, vol. 45, no. 5, pp. 780–795, Mar. 2016.

[12] K. Yildiz, "Dimensionality reduction-based feature extraction and classification on fleece fabric images," *Signal, Image Video Process.*, vol. 11, no. 2, pp. 317–323, Feb. 2017.

[13] K. A. Yildiz and Buldu, "Wavelet transform and principal component analysis in fabric defect detection and classification," *Pamukkale Univ. J. Eng. Sci.-Pamukkale Universitesi Muhendislik Bilimleri Dergisi*, vol. 23, no. 5, pp. 622–627, 2017.

[14] K. Yıldız, A. Buldu, M. Demetgul, and Z. Yildiz, "A novel thermal-based fabric defect detection technique," *J. Textile Inst.*, vol. 106, no. 3, pp. 275–283, Mar. 2015.

[15] W. Ouyang, B. Xu, J. Hou, and X. Yuan, "Fabric defect detection using activation layer embedded convolutional neural network," *IEEE Access*, vol. 7, pp. 70130–70140, 2019.

[16] W. J. Liu, H. Liu, Z. R. Li, and D. Y. Lai, "A fabric defect detection algorithm based on image enhancement and CNN," *Comput. Technol. Develop.*, vol. 31, pp. 90–95, Jan. 2021.

[17] L. Li, "Research on fabric defect detection method based on convolutional neural network," M.S. thesis, Ind. Eng., Harbin Inst. Technol., Huazhong Univ. Sci. Technol., Wuhan, China, 2019.

[18] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN architectures for large-scale audio classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 131–135.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2015.

[21] Y. Y. Liu, "Research on fabric defect detection method based on deep learning," M.S. thesis, Dept. Control Sci. Eng., Harbin Inst. Technol., Harbin, China, 2020.

[22] J. J. Li, "Research on fabric defect detection algorithm based on deep learning," M.S. thesis, Dept. Electron. Commun. Eng., Univ. Electron. Sci. Technol. China, Chengdu, China, 2020.

[23] Z. Y. Zhao, "Research on recognition and detection of textile defects based on deep learning," M.S. thesis, Dept. Control Sci. Eng., Huazhong Univ. Sci. Technol., Wuhan, China, 2019.

[24] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[25] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[26] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[27] D. Misra, "Mish: A self regularized non-monotonic neural activation function," 2019, *arXiv:1908.08681*.

[28] A. M. Hafiz, "K-nearest neighbour and support vector machine hybrid classification," *Int. J. Imag. Robot.*, vol. 19, pp. 33–41, Jun. 2020.

[29] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[30] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1753–1761.

[31] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Jun. 2015.

[32] X. Li, Y. F. Zha, T. Z. Zhang, Z. Cui, W. M. Zuo, Z. Q. Hou, H. C. Lu, and H. Z. Wang, "Survey of visual object tracking algorithms based on deep learning," *J. Image Graph.*, vol. 24, no. 12, pp. 2057–2080, 2019.

[33] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, Jan. 2014, pp. 1988–1996.

[34] C. Dewi, R.-C. Chen, and S.-K. Tai, "Evaluation of robust spatial pyramid pooling based on convolutional neural network for traffic sign recognition system," *Electronics*, vol. 9, no. 6, p. 889, May 2020.

[35] S.-K. Tai, C. Dewi, R.-C. Chen, Y.-T. Liu, X. Jiang, and H. Yu, "Deep learning for traffic sign recognition based on spatial pyramid pooling with scale analysis," *Appl. Sci.*, vol. 10, no. 19, p. 6997, Oct. 2020.

[36] Y. P. Chen, J. N. Li, H. X. Xiao, X. J. Jin, S. C. Yan, and J. S. Feng, "Dual path networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 4467–4475.

[37] E. Real, A. Aggarwal, Y. P. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," in *Proc. AAAI*, Honolulu, HI, USA, 2019, pp. 4780–4789.

[38] A. Stergiou, R. Poppe, and G. Kalliatakis, "Refining activation downsampling with SoftPool," 2021, *arXiv:2101.00440*.

[39] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," in *Proc. ICLR*, Scottsdale, AZ, USA, 2013, pp. 1–9.

[40] S. Zhai, H. Wu, A. Kumar, Y. Cheng, Y. Lu, Z. Zhang, and R. Feris, "S3Pool: Pooling with stochastic spatial sampling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4003–4011.

[41] E. Jang, S. S. Gu, and B. Poole, "Categorical reparameterization with Gumbel–Softmax," in *Proc. ICLR*, Toulon, France, 2017, pp. 1–13.

[42] F. Wu, H. Zhao, and M. Wang, "Nighttime cattle detection based on YOLOv4," in *Proc. 12th Int. Conf. Graph. Image Process. (ICGIP)*, Jan. 2021, Art. no. 1172008.

[43] C. Low, A. B. J. Teoh, and C. Ng, "Multi-fold Gabor, PCA, and ICA filter convolution descriptor for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 115–129, Jan. 2019.

[44] X. Fu and X. Cao, "Underwater image enhancement with global–local networks and compressed-histogram equalization," *Signal Process., Image Commun.*, vol. 86, Aug. 2020, Art. no. 115892.

[45] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graph. Gems.*, pp. 474–478, Aug. 1994.

[46] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," 2018, *arXiv:1804.07612*.

[47] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. ICLR*, Toulon, France, 2017, pp. 1–16.

[48] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. ECCV*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

[49] P. Zhang, Y. Zhong, and X. Li, "SlimYOLOv3: Narrower, faster and better for real-time UAV applications," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 37–45.

QIANG LIU received the B.S. degree in automation from Weifang University, Weifang, Shandong, China, in 2007, the M.S. degree in control science and technology from the Shandong University of Science and Technology, Qingdao, Shandong, in 2010, and the Ph.D. degree in control science and engineering from the Harbin Institution of Technology, Harbin, Heilongjiang, China, in 2017. Since 2017, he has been a Lecturer with the School of Mechanical Engineering, Shandong University of Technology. He is the author of more than eight articles. His research interests include intelligent robot and intelligent control.

CHUAN WANG received the B.S. degree in mechatronic engineering from the Shandong University of Technology, Zibo, Shandong, China, where he is currently pursuing the M.S. degree in mechanical engineering. His research interest includes intelligent robot.

YUSHENG LI received the B.S. degree in mechanical manufacturing technology and equipment from the Shandong Institute of Engineering, Zibo, Shandong, China, in 2000, and the M.S. degree in mechatronic engineering from the Shandong University of Technology, Zibo, in 2003. From 2004 to 2005, he worked at Wuhan Kaimu Information Technology Company Ltd., Wuhan, as a Research Personnel. From January 2006 to May 2006, he worked at Shanghai Tongji Tongjie Digital Technology Company Ltd., Shanghai, as a Research Personnel. From June 2006 to December 2006, he was a Lecturer with the Engineering Training Center, Shandong University of Technology, where he was a Lecturer with the Office of Academic Affairs, from January 2006 to December 2007. Since January 2009, he has been a Lecturer with the School of Mechanical Engineering, Shandong University of Technology. He is the author of more than ten articles. His research interests include computer-aided design, computer-aided manufacturing, computer-aided education, and virtual reality.

MINGWANG GAO received the B.S. degree in mechatronic engineering and the M.S. degree in mechanical design and theory from Shandong University, Jinan, Shandong, China, in 2000 and 2004, respectively, and the Ph.D. degree in mechanical manufacturing and automation from the South China University of Technology, Guangzhou, Guangdong, China, in 2012. From 2004 to 2006, he was an Assistant with the School of Engineering and Technology, Shandong University of Technology, Zibo, where he was a Lecturer with the School of Engineering and Technology, from 2006 to 2008, and the School of Mechanical Engineering, from 2008 to 2018. Since 2018, he has been an Associate Professor with the School of Mechanical Engineering, Shandong University of Technology. He is the author of ten articles. His research interest includes mechanism design and control.

JINGAO LI is currently pursuing the bachelor's degree in mechatronic engineering from the Shandong University of Technology, Zibo, China. He is the author of one article. His research interests include machine vision and driverless.

● ● ●