

Fabric defect detection using the improved YOLOv3 model

Junfeng Jing^{id}, Dong Zhuo, Huanhuan Zhang,
Yong Liang and Min Zheng

Abstract

To improve the detection rate of defect and the fabric product quality, a higher real-time performance fabric defect detection method based on the improved YOLOv3 model is proposed. There are two key steps: first, on the basis of YOLOv3, the dimension clustering of target frames is carried out by combining the fabric defect size and *k*-means algorithm to determine the number and size of prior frames. Second, the low-level features are combined with the high-level information, and the YOLO detection layer is added on to the feature maps of different sizes, so that it can be better applied to the defect detection of the gray cloth and the lattice fabric. The error detection rate of the improved network model is less than 5% for both gray cloth and checked cloth. Experimental results show that the proposed method can detect and mark fabric defects more effectively than YOLOv3, and effectively reduce the error detection rate.

Keywords

Fabric defect detection, YOLO, priori box, information fusion, error detection rate

Date received: 29 May 2019; accepted: 19 January 2020

Introduction

Traditional defect detection mainly relies on experienced professionals. It has a certain subjectivity and depends on the personal experience of the inspectors. In addition, the long-term labor would cause the detection rate to be greatly reduced. It is to notice that the manual detection speed is difficult to meet the needs of real-time online detection. The traditional automatic defect detection method is mainly based on the artificially designed feature set, including the statistical features,¹ structural features,² and spectral features³ of the image. These methods have achieved good results for specific fabric products. However for new fabric designs, or when the images capture environmental changes, these methods must be modified or even redesigned.

Compared with the previous artificial design features, the deep learning algorithm can automatically learn the multi-scale features of the image through the multi-layer

network, and can simultaneously acquire the local information of the image, as well as the abstract semantic information of the upper layer. The application of convolutional neural network (CNN) in fabric image defect detection can solve the multi-deformation and multi-scale problem of image, make it possible to construct deep and complex texture defect model, and realize intelligent detection and location of defects, which is of great significance to improve the product quality.

In 2006, Hinton et al.⁴ put forward the concept of deep learning for the first time, and mentioned that the deep

Xi'an Polytechnic University, Xi'an, China

Corresponding author:

Junfeng Jing, Xi'an Polytechnic University, 19 of the South Jinhua Road, Xi'an 710048, Shaanxi, China.

Email: jingjunfeng0718@sina.com



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

neural network model has strong feature learning ability. In 2014, Ross B. Girshick (RBG) and others used candidate regions to replace the sliding windows, CNNs instead of artificially designed features, and proposed R-CNN, which is a region-based convolutional neural network named.⁵ Based on this, spatial pyramid pooling (SPP)-NET⁶ transforms multiple convolutions into one convolution in R-CNN, which greatly reduces the computational complexity. Fast R-CNN⁷ combines the structural advantages of R-CNN and SPP-NET, and uses multi-task loss function to train the network to achieve the purpose of target detection; however the fast R-CNN still cannot meet the real-time application of target detection. Faster R-CNN⁸ combines candidate region and CNN classification into a complete network, which improves the detection speed of the network and realizes end-to-end training of the network.

In 2015, YOLO⁹ adopts an integrated detection scheme, which integrates candidate frame extraction, CNN learning features, and non-maximum suppression optimization¹⁰ to make the network structure simpler. The detection speed is nearly 10 times higher than that of the faster R-CNN. This makes the deep learning target detection algorithm to meet the requirement of real-time detection tasks under the computing power at that time; however, the detection performance on small targets is not good. YOLOv2¹¹ improves the network structure of YOLOv1 by adding batch normalization,¹² high-resolution classifier,¹³ convolution with anchor boxes,¹⁴ dimension clusters,¹⁵ and other optimization models to improve the accuracy of target regression and positioning. YOLOv3¹⁶ uses the residual network on the basis of YOLOv2 and combines the feature pyramid network (FPN)¹⁷ structure, using the binary cross loss function as the loss function. After extracting the features, the upper two layers of the feature map are up-sampled and merged with the corresponding feature maps of the network. After the convolutional network, the prediction results are obtained, and the accuracy and speed are achieved. Based on the good performance of YOLOv3, many researchers have introduced the network model into their own research field and achieved good results.^{18–20} Based on the above research, we apply the CNN to textile companies to solve the problem of fabric defect detection.

To improve the detection rate of fabric defects, the deep CNN YOLOv3 is used as the basic defect detection framework and is optimized to better detect fabric defects. The remainder of the article is organized as follows. First, we introduce the relevant part of YOLOv3. Then, the priori frame of YOLOv3 is modified using the *k*-means algorithm to cluster the defect data. Second, we combine the image pyramid to obtain the different scales of feature maps and add the detection layer to improve the network structure. Finally, the network is tested and analyzed using gray cloth and lattice fabric.

YOLOv3 network

YOLOv3 is an end-to-end target detection algorithm based on the regression theory. Combining CNN, non-maximum suppression algorithm and feature pyramid, it can predict the defect borders and categories. The bounding box is classified by independent logistic regression classifier instead of softmax, and the target class is predicted by binary cross entropy loss. Based on the above design ideas, YOLOv3 achieves good results in accuracy and speed.

Feature extraction

Based on the validity of CNN for feature extraction, YOLOv3 still uses CNN for feature extraction. YOLOv3 integrates YOLOv2, darknet-19, and ResNet to design the network structure for feature extraction. The convolution layers of 3×3 and 1×1 with better performance are used. The convolution step of convolution layer is set to 2 instead of pooling layer. Scale invariant features are transmitted to the next level of convolution, and a shortcut connection is added. Batch normalization and dropout operations are added after each level of convolution. The feature extraction network has 53 convolution layers, thus it becomes darknet-53. Comparing the performance of darknet-53 with others in the ImageNet, the test result of the Top-1 is to reach 77.2% mAP.

Usually, the detection targets have different scales. To detect objects of different sizes at the same time, the network must have the ability to detect objects of different sizes. However, as the network depth increases, the feature map decreases gradually; the smaller the size of the target is, the more difficult it is to detect. To detect objects of different sizes at the same time, YOLOv3 adopts the idea of FPN, uses up-sampling and feature fusion to detect objects of different sizes on feature maps, which improves the detection performance of small targets. The YOLOv3 network structure is shown in Figure 1. Detection layers are the 79, 91, and 103 layers that detect defects on multi-scale feature maps. The *k*-means method is used to cluster the objects in COCO data sets, and nine anchors with different sizes are obtained. Combined with the size of the predicted map, the anchors are equally divided into feature maps of different scales for detection.

Proposed network model

Prior frame determined

In YOLOv3, the idea of anchor boxes used in faster R-CNN is introduced. The *k*-means clustering algorithm is used to set three priori boxes for each scale, and a total of nine size priori boxes are clustered.

Applying a larger priori box on a smaller feature map can better detect larger objects. The size of some defective target boxes is shown in Figure 2. It is difficult to obtain accurate target information directly using the prior boxes in YOLOv3.

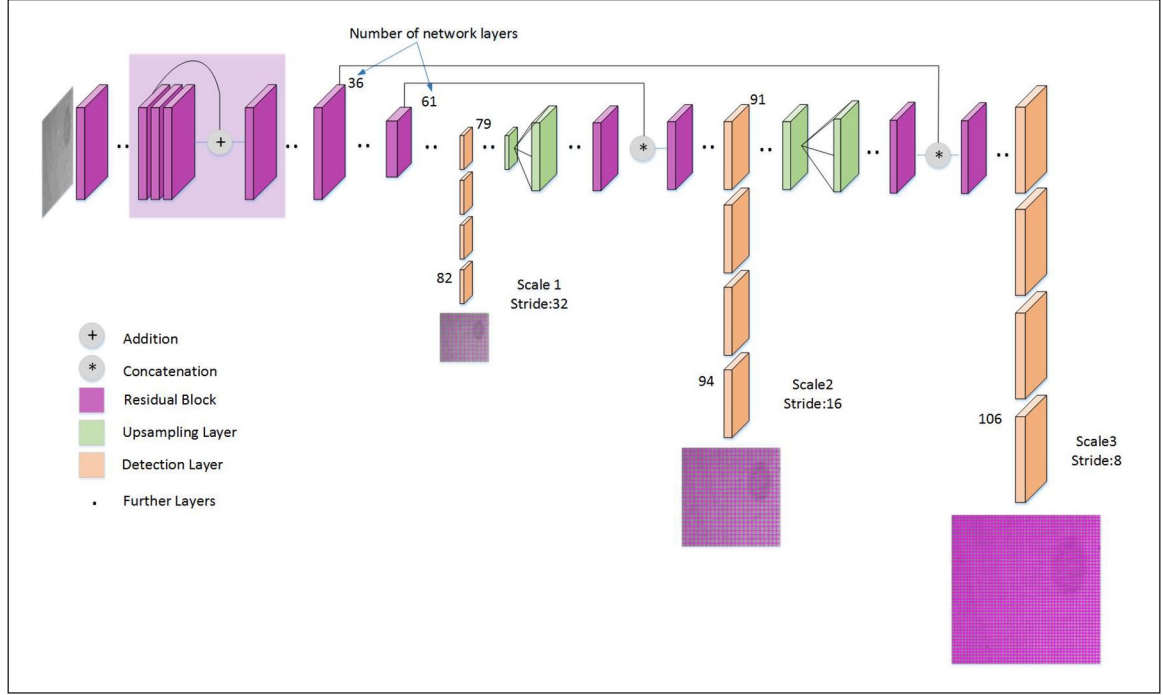


Figure 1. The structure of YOLOv3 network.

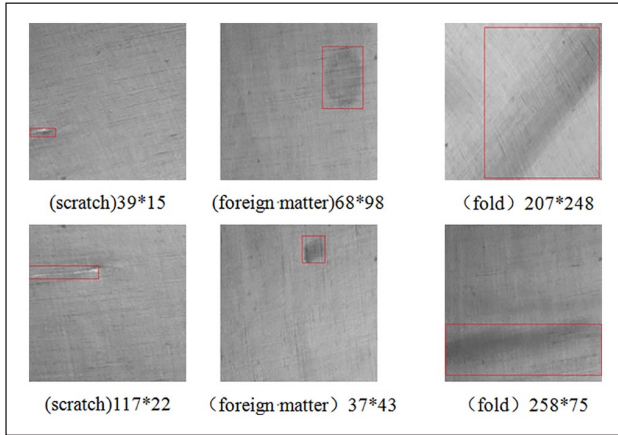


Figure 2. The size of partial sample defect.

There is an overlap between the predicted border and the actual border. The larger the overlap area, the better the model prediction result is. The overlap area size can be quantitatively analyzed by calculating the intersection-over-union (IoU).

After labeling the gray cloth data set, the cluster analysis is carried out, and the relationship between the number of clusters and *Avg IOU* is obtained by selecting $k=1-15$, as shown in Figure 3. When $k \leq 12$, as the value of k increases, the objective function also increases. When $k=12$, the maximum value is 81.38%. When $k \geq 12$, the objective function shows a decreasing trend. Therefore, when $k=12$ is selected, the number of anchors is the

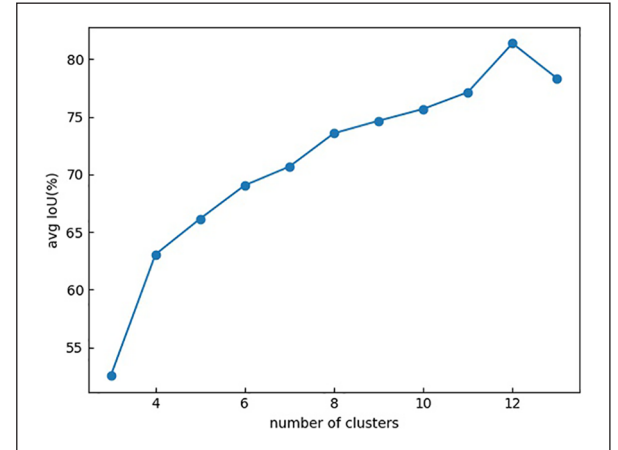


Figure 3. The clustering results of gray cloth.

optimal, and the size of the corresponding prediction frame is set to 12 clusters. The width and height of the corresponding cluster centers on the gray cloth data set, respectively, are given by (21, 93), (69, 65), (65, 217), (70, 24), (107, 148), (148, 103), (154, 405), (212, 65), (220, 210), (247, 387), (370, 367), (40, 159).

Network optimization

The CNN extracts the features of the target through layer-by-layer abstraction. One of the important concepts is the receptive field. If the field is too small, only local features can be observed. If the field is too large, too much invalid

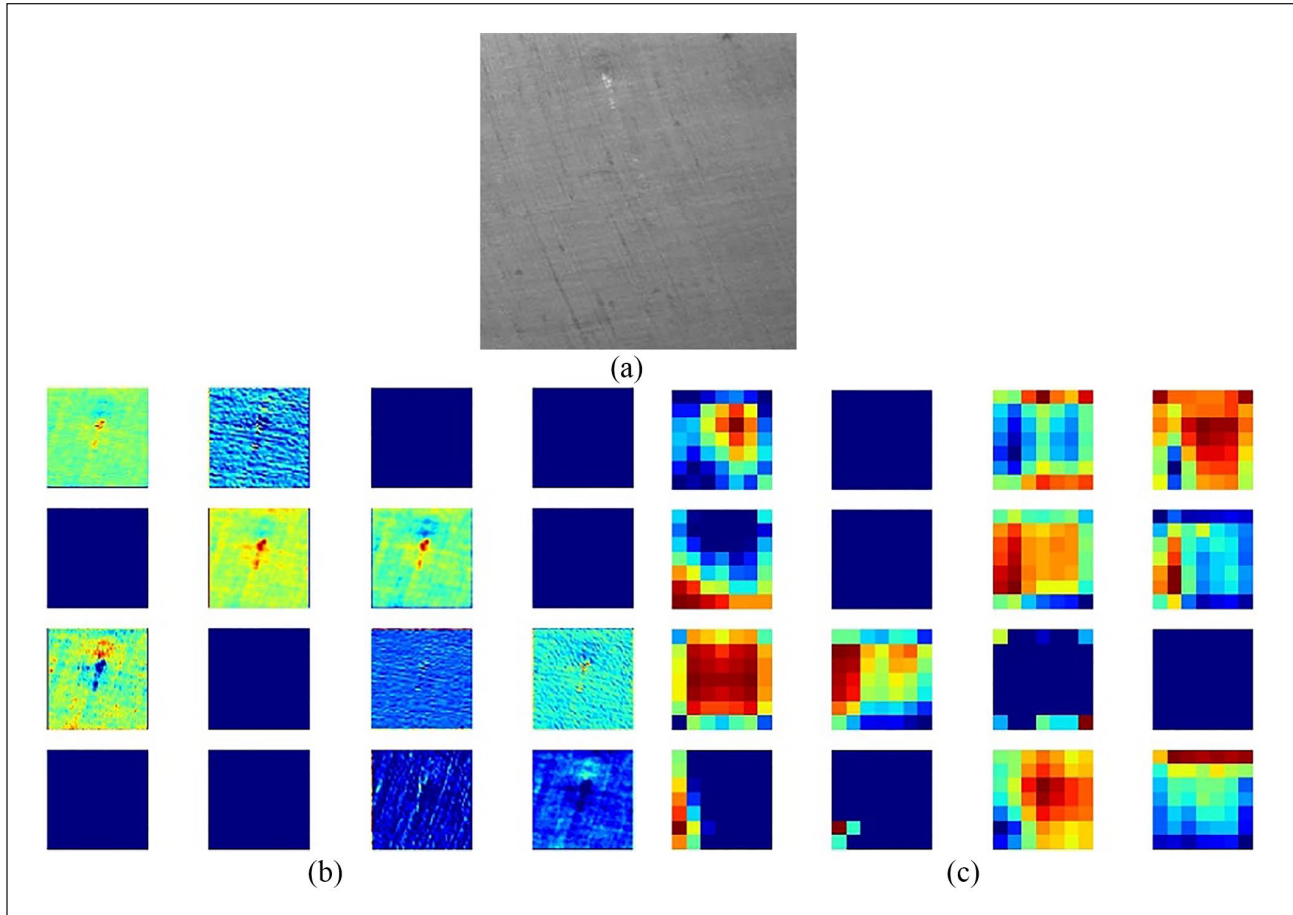


Figure 4. (a) Input data, (b) low-level features, and (c) high-level features.

information is obtained. Therefore, a various multi-scale model structures have been designed, mainly image pyramid and feature pyramid. The specific network architectures can be divided into the following: (1) multi-scale input, (2) multi-scale feature fusion, and (3) multi-scale features and predictive fusion.

In the YOLO model, the third structure is used for target detection, and the prediction is performed at different feature sizes. Finally, the results are fused. This structure is represented by the FPN in the target detection, which adds the high-level features to the adjacent low-level to form new features, and each layer separately forecasts. In the YOLOv3 model, when the input image size is 416×416 pixel, YOLOv3 performs target recognition, respectively, on the feature maps of sizes 13×13 , 26×26 , and 52×52 . Use smaller priori box for defect detection on larger feature maps.

The deeper features in CNN have a large receptive field and rich semantic information. The deeper features are robust to the attitude change, occlusion, and local deformation of the object, but due to the reduction of resolution, geometric details are lost. On the contrary, the shallow features have very small receptive fields and rich geometric details. The resolution is high but the semantic information

is relatively scarce. In a CNN, the semantic information of objects can appear in different layers, which is related to the size of the detected objects.

For small objects, shallow features contain some details. With the deepening of the number of layers, due to the large receptive field, the geometric details in the extracted features may disappear completely. It is difficult to detect small objects through deep features. For large objects, its semantic information will appear in deeper features.

In the process of network target recognition, the low-level features have rich details of the target and location information, while the high-level targets have rich semantic features. Through the multi-layer convolution and pooling process, the details and location information of the target are gradually reduced, whereas the semantic information is increasing. Figure 4(a) shows the input data, and the feature extraction is performed using a CNN. The low-level features are shown in Figure 4(b) and the high-level ones are shown in Figure 4(c).

In Figure 4, the defect type of the input data is scratch, and the defect area is small. The defect contour area is obviously different from the normal part. However, with the deepening of convolution and pooling, image texture features will become more and more blurred, which will

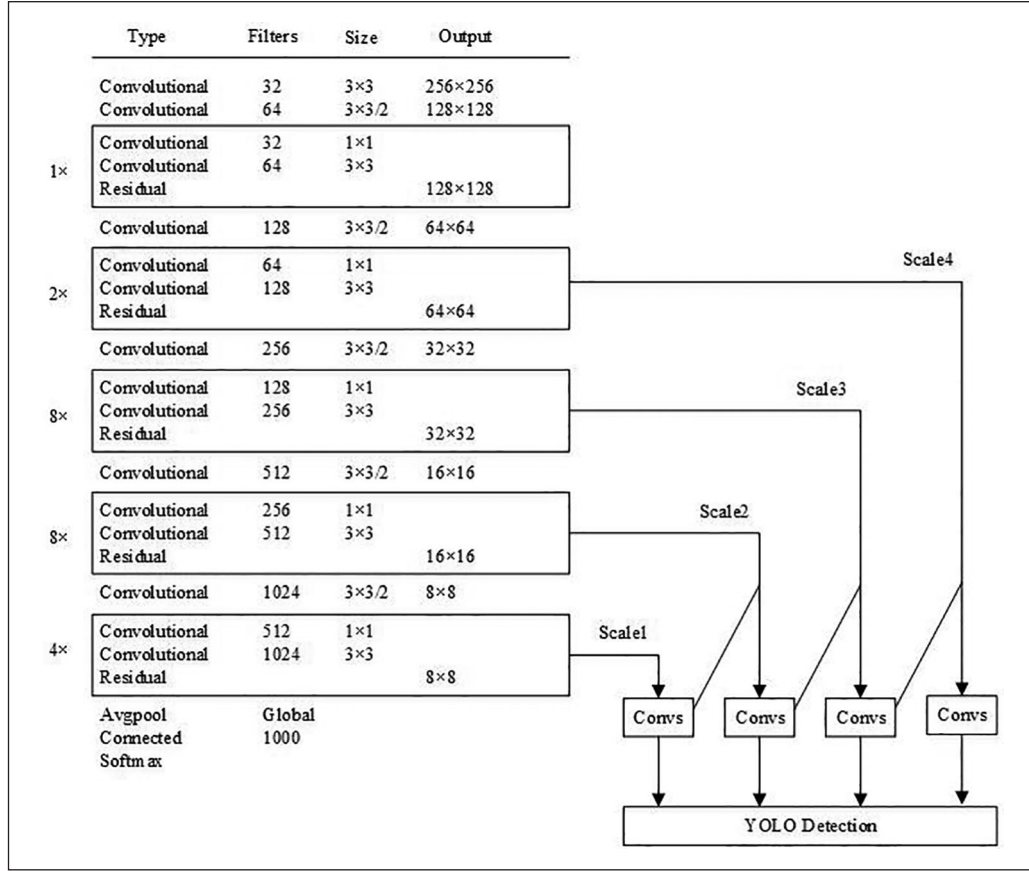


Figure 5. The framework of the target detection.

increase the difficulty of defect recognition. Therefore, feature fusion can be used to detect defects. Combined with the image pyramid, the high-level information obtained by the up-sampling is merged with the low-level features to obtain feature maps of different scales, and the detection layer is added to improve the network structure.

The target detection structure of the improved network model is shown in Figure 5, and the data on the left side indicate the number of repeating units. The detection layer is added when the feature map size is 104×104 . The dimensional clustering of target frames in data sets is carried out by the k -means method, and the detection layers with different scales are evenly distributed. The target detection is performed at four scales.

Results and analysis

Experimental environment and data

YOLOv3 is a representative multi-scale target detection algorithm, which can take into account both small and large targets, and has a good performance of detection for small targets. Ubuntu operating system is used in the experiment. The processor is Intel®CoreTMi7-6800K CPU@3.40GHz, the memory is 125.8GiB, and the

graphics card GTX1080 Ti. The darknet framework is configured in the experimental environment. It is a relatively lightweight open source deep learning framework based on C and CUDA. Its main features are easy to install, no dependencies, very good portability, and support for both CPU and GPU computing.

The test results of the YOLO model are greatly influenced by the samples, which need to be diverse and representative. Taking gray cloth and lattice as research objects, defect images were collected by industrial camera. The defect database is formed by enhancing and expanding the data set by means of rotation and contrast enhancement. The data set is divided into training set and test set. The number of defect samples is shown in Figure 6. The abscissa is the defect category and the ordinate is the number of data.

The main defect types in gray cloth include scratch, foreign matter, and fold. The scratches are mainly characterized by fine stripes. The foreign matter is expressed as a region with obvious contrast with the background color. The fold is characterized by partial protrusion or depression. The three types of defects have obvious differences in appearance, and each type of defect sample is similar.

The lattice fabric mainly contains three types of defects: ribbon yarn, broken ends, and hole. Compared with ribbon yarn, broken ends have more broken yarns, and the

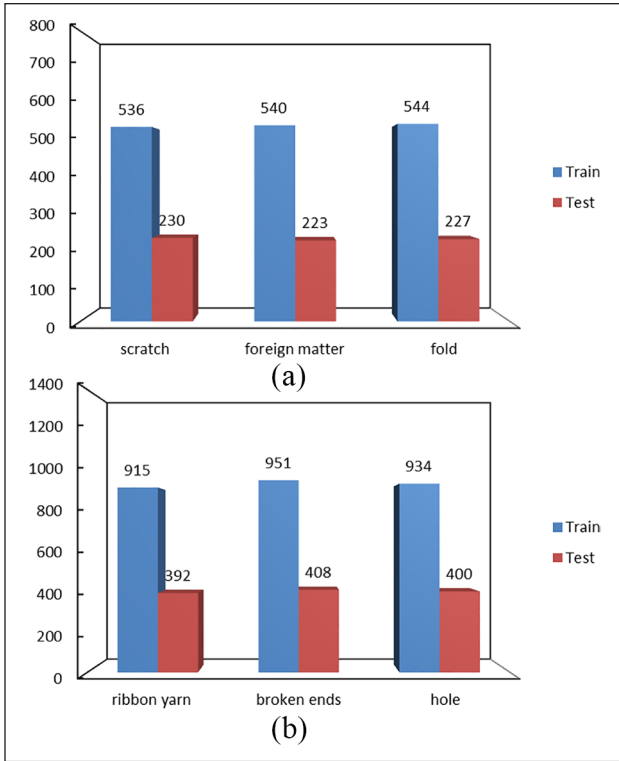


Figure 6. Defect sample number: (a) gray sample number and (b) lattice sample number.

appearance of broken ends is rectangular. Hole is the defect in some areas of the sample. Each type of defect has similarities, and there are significant differences between the three types of defects.

We name the image according to the same rule and set the image size to 416×416 pixel. The labelling software is used to mark the image according to the defect category and position. The defect is marked in the image, and the corresponding .xml file is generated, which contains the file name and ground truth of the corresponding image. To reduce the amount of calculation, the ground truth is normalized to a data range of 0–1.

Network training

Using the gray cloth data as the experimental object, the priori frame of YOLOv3 network is modified according to clustering results, and the experiment was compared with the YOLOv3 and the improved network model. The initial learning rate of the network is 0.001 and the total number of iterations is 8000 steps. The learning rate is reduced to 0.0001 and 0.00001 at 7000 steps and 7500 steps, respectively. Each 64 images are iterated in batches. The parameters are initialized using the source weights to accelerate the convergence of loss function. In the training process, the curve of loss and IoU are drawn as shown in Figure 7. The abscissa is the number of iterations, and the ordinate is the loss value and the mean IoU.

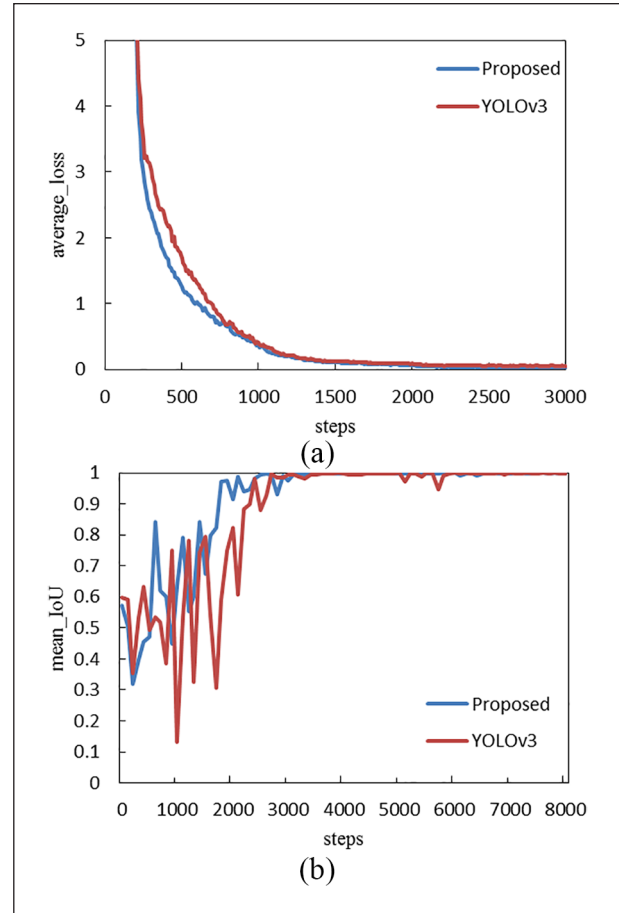


Figure 7. Curve of the training process: (a) loss value curve and (b) IoU curve.

From Figure 7, it can be seen that as the number of iterations increases, the curve of the average loss value tends to reach zero. When the number of iterations is about 2500, the loss value decreases to 0.01, and the improved method decreases rapidly. The merging ratio of the target box and the actual border is close to 1.

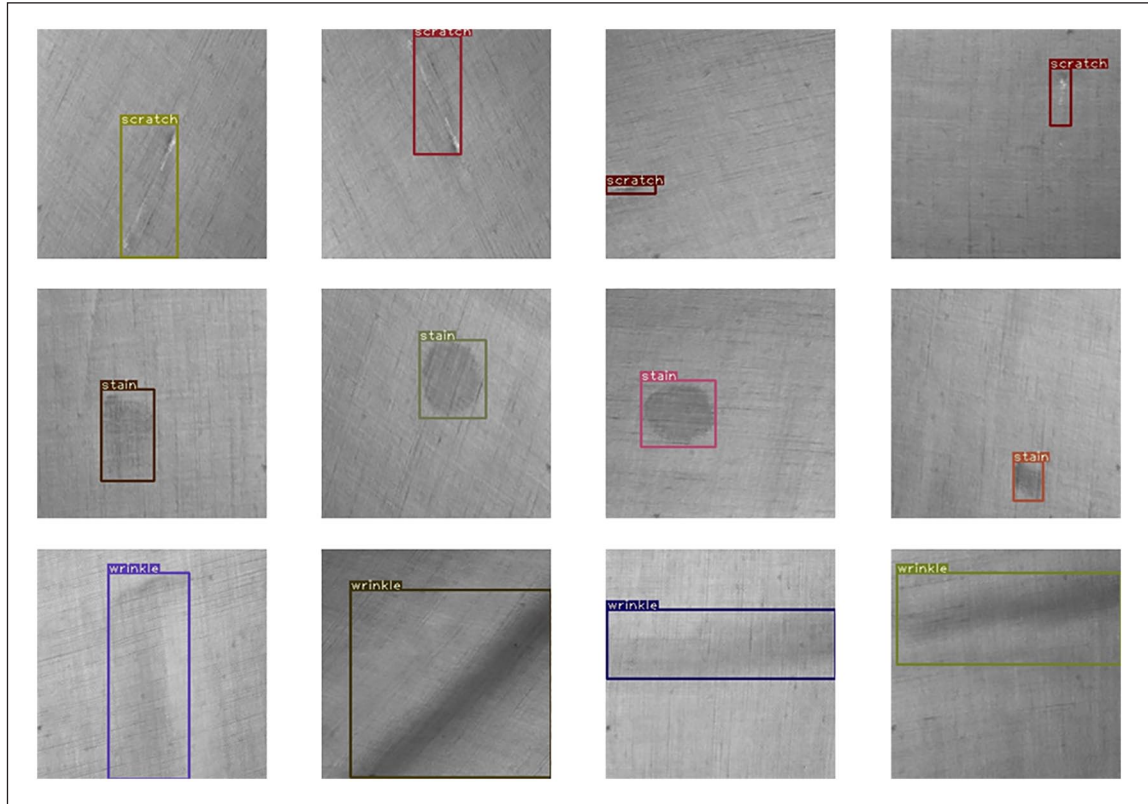
Network testing

To verify the accuracy of the model, the training model is tested with the gray cloth test set, and its detection result is shown in Table 1. The actual number of defects is counted, while the number of false detection and the rate of false detection are calculated. A portion of the test results are shown in Figure 8.

From Table 1, it can be seen that the total error detection rate of the improved network model is 2.19% and the original network model is 4.39%. The improved model is more accurate than the original model, and the error rate is reduced by 2.2%. Because the size of scratch defect in gray cloth is small and it is close to the background of gray cloth, the error detection rate of scratch is higher. Compared with foreign matter and fold, the error detection rate is lower.

Table 1. The test results of gray cloth.

Defect types	YOLOv3			Proposed		
	Test number	False number	Error rate (%)	Test number	False number	Error rate (%)
Scratch	230	5	2.17	230	3	1.3
Foreign matter	223	3	1.34	223	1	0.45
Fold	227	2	0.88	227	1	0.44

**Figure 8.** The inspection results of gray cloth.**Table 2.** The test results of lattice.

Defect types	YOLOv3			Proposed		
	Test number	False number	Error rate (%)	Test number	False number	Error rate (%)
Ribbon yarn	392	8	2.04	392	4	1.02
Broken ends	408	3	0.74	408	2	0.49
Hole	400	2	0.5	400	1	0.25

On this basis, the improved network model is used to detect the defects of lattices. Dimensional clustering analysis was carried out on the defect markers of samples. When the clustering number is 12, the cluster center is selected as the a priori box, which is (42, 8), (10, 56), (28, 30), (15, 83), (11, 128), (10, 169), (53, 31), (12, 146), (132, 21), (124, 23), (27, 108), (35, 102). The accuracy of the model is verified using test set. The test results are shown in Table 2, and the partial test is shown in Figure 9.

From Table 2, we can see that the total error detection rate of the improved network model is 1.76% and the original network model is 3.28% for the lattice data set. The improved model is more accurate than the original model, and the error rate is decreased by 1.52%. Because the size of the broken ends in the fabric is small and the color is close to the background of the checked fabric, the false detection rate of the ribbon is higher. Compared with the background of the test sample, the characteristics of the

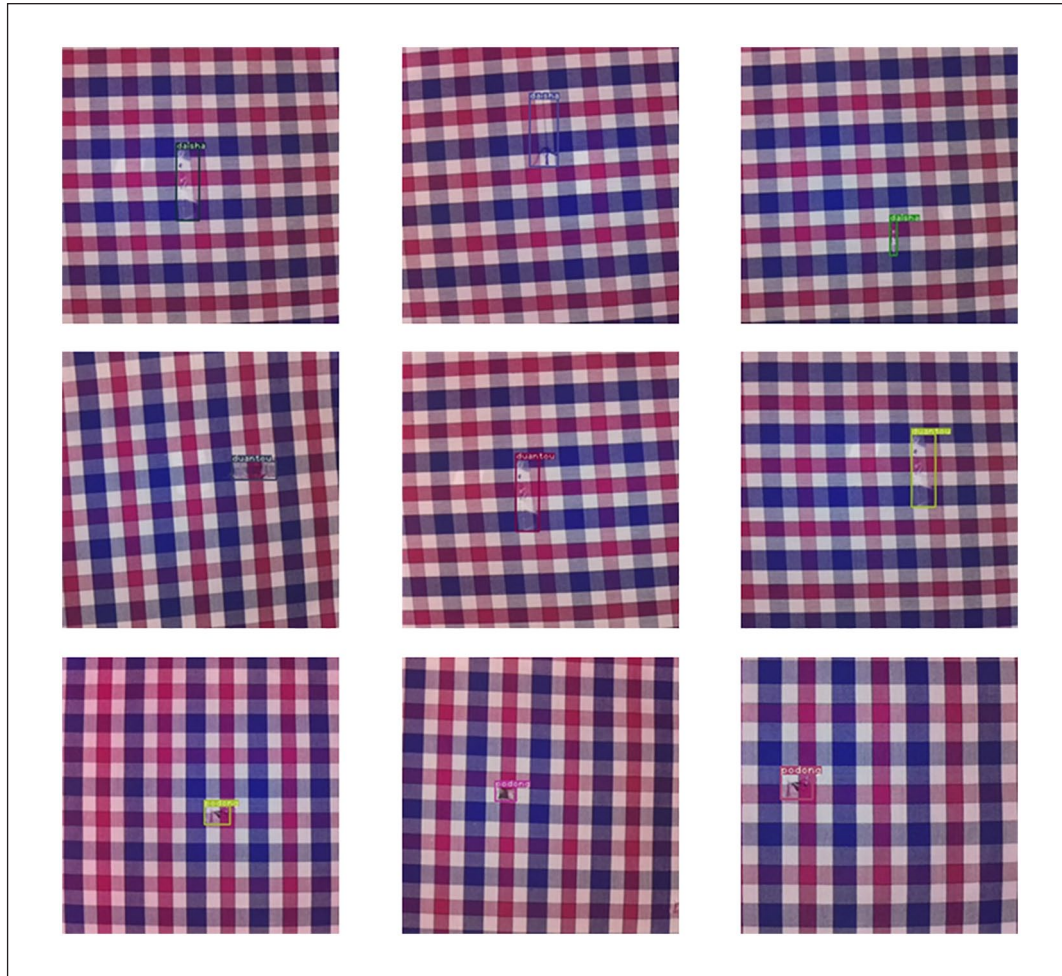


Figure 9. The test results of lattice.

holes are more obvious, the detection rate is higher and the error rate is 0.25%.

As each type of defect in the test set has similarity, and there are obvious differences between different defects, the false detection rate is low, and the error rate is mainly caused by missed detection.

In the network model testing, some of the missing samples are shown in Figure 10, and the arrow is the defect location. For the gray cloth samples, when the amplitude of the wrinkle defects is small, it is difficult to locate and classify the sample defects during the model detection; the defects such as holes and foreign objects are difficult to detect the defect location when they are located at the edge of the sample or the defects are small. There are few defects on the edge. The sample defect in the detection area is incomplete, and only the partial area of the defect is included, and the feature is incomplete and difficult to detect. For the lattice samples, the detection accuracy is improved, but some samples are also difficult to detect due to the low discrimination between the defect and the background color and the small defect in the total sample area.

To verify the performance of the network, the improved network model was compared with other networks on the test data set. Compared with the average accuracy of the network, under the same parameters, the experimental results are shown in Figure 11. The improved network model has a large improvement on the gray cloth and the lattice fabric, which can better detect the fabric defects.

During the experiment, calculate the average test time of the samples for comparison. FPS represents the number of picture frames processed per second, and we use FPS to evaluate the network detection speed. We test the samples on our experimental platform. When the size of the input samples is 416×416 pixel, the average test time in the original network model of YOLOv3 is 27.7 fps, while the average test time in the improved network model is 21.8 fps. In the improved network model, the detection time increased slightly, but the missed detection rate of the detection samples decreased. The total error detection rate of the gray fabric decreased from 4.39% to 2.19%, and the total error detection rate of the lattice fabric decreased from 3.28% to 1.76%.

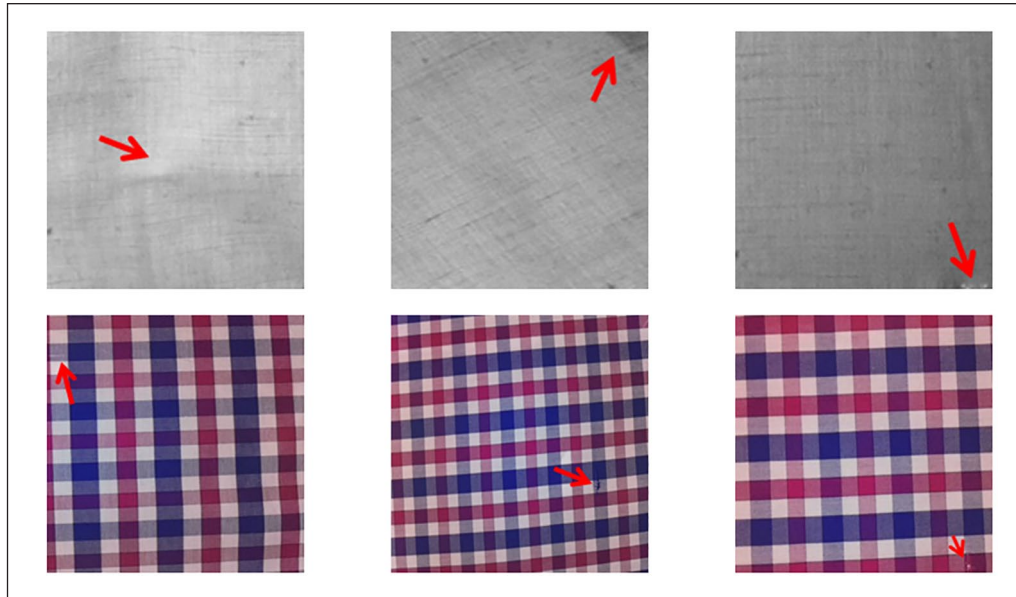


Figure 10. Partial missed sample.

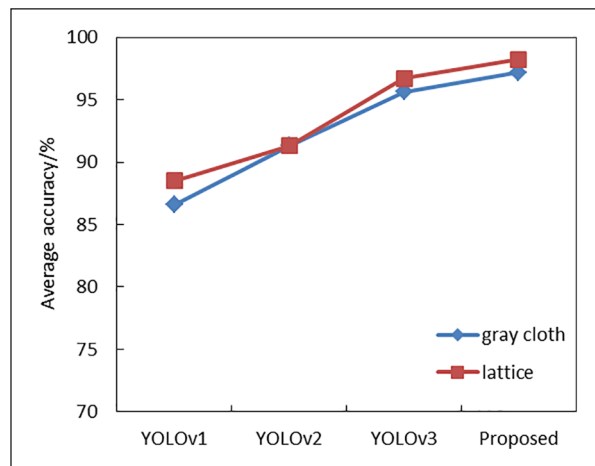


Figure 11. Test results of network models.

Conclusion

YOLOv3 network model is applied to fabric defect detection to realize fabric defect detection. To solve the problem that the initial anchor points in the YOLOv3 model are not suitable for fabric defect detection, the k -means algorithm is used to cluster the marker data, and the clustering center generated by clustering is used as the a priori frame for fabric defect detection. On this basis, the detection layer of YOLO is added to improve the detection rate of defects. The experimental results have shown that the proposed method can effectively reduce the error rate of the network and has good applications in both gray cloth and lattice. However, the real-time performance of the method needs to be improved, which will be our future research project.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Junfeng Jing  <https://orcid.org/0000-0001-6646-3698>

References

1. Jia L, Chen C, Liang J, et al. Fabric defect inspection based on lattice segmentation and Gabor filtering. *Neurocomputing* 2017; 238: 84–102.
2. Jing J-F, Chen S and Li P-F. Fabric defect detection based on golden image subtraction. *Color Technol* 2016; 133(1): 26–39.
3. Pan Q, Chen M, Zuo B, et al. The inspection of raw-silk defects using image vision. *J Eng Fiber Fabr* 2018; 13(3): 78–86.
4. Hinton GE and Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006; 313(5786): 504–507.
5. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Columbus, OH, 23–28 June 2014, vol. 1, pp. 580–587. New York: IEEE.
6. He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 2014; 37(9): 1904–1916.
7. Girshick R. Fast R-CNN. In: *Proceedings of the IEEE international conference on computer vision*, Santiago, Chile, 7–13 December 2015, pp. 11440–11448. New York: IEEE.

8. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE T Pattern Anal* 2017; 39(6): 1137–1149.
9. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. *Comp Vis Pattern Recognit* 2016; 1: 779–788.
10. Neubeck A and Van Gool L. Efficient non-maximum suppression. In: *Proceedings of the 18th international conference on pattern recognition*, Hong Kong, China, 20–24 August 2006, vol. 3, pp. 850–855. New York: IEEE.
11. Redmon J and Farhadi A. YOLO9000: better, faster, stronger. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, 21–26 July 2017, vol. 1, pp. 6517–6525. New York: IEEE.
12. Ioffe S and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *Int Conf Mach Learn* 2015; 37: 448–456.
13. Zhang Y, Zhang H and Huang TS. Collaborative and compressive high-resolution imaging. *Int C Patt Recog* 2012; 1: 3062–3065.
14. Ji H, Zhang X and Yang J. Discussion on the detection technology of fabric defects based on machine vision. *Adv Text Technol* 2011; 2: 11–15.
15. De G, Tapfer L, Catalano M, et al. Formation of copper and silver nanometer dimension clusters in silica by the sol-gel process. *Appl Phys Lett* 1996; 68(26): 3820–3822.
16. Kim KJ, Kim PK, Chung YS, et al. Performance enhancement of YOLOv3 by adding prediction layers with spatial pyramid pooling for vehicle detection. In: *Proceedings of the 15th IEEE international conference on advanced video and signal based surveillance*, Auckland, New Zealand, 27–30 November 2018, vol. 1, pp. 1–6. New York: IEEE.
17. Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, 21–26 July 2017, vol. 1, pp. 936–944. New York: IEEE.
18. Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput Electron Agr* 2019; 157: 417–426.
19. Zhang X, Yang W, Tang X, et al. A fast learning method for accurate and robust lane detection using two-stage feature extraction with YOLO v3. *Sensors* 2018; 18(12): E4308.
20. Tang C, Ling Y, Yang X, et al. Multi-view object detection based on deep learning. *Appl Sci* 2018; 8(9): 1423.