An Improved-YOLOX Model for Detection of Fabric Defects

Changzheng Chen

Abstract—Fabric defect detection is a crucial stage in fabric production. In the past, defect detection relied heavily on manual inspection, which was inefficient, time-consuming, and expensive. This paper proposes an Improved-YOLOX network for fabric defect detection based on computer vision. First, we design an attention module by combining channel attention and spatial attention to extract useful information for improving the accuracy of the network. Second, Dynamic convolution is introduced in the detection head to enhance the detection of small targets. Third, the VariFocal loss function and the GIOU loss function are used to improve the performance of the network. The experiments show that our approach on three fabric datasets can reach an average of over 96.47% mAP, and the speed exceeds 30 FPS, which can meet the requirements of industrial production.

Index Terms—Improved-YOLOX, attention module, defect detection.

I. INTRODUCTION

Fabrics are utilized in many different fields, including the clothing, medical and aerospace [1-4]. During production, broken yarn, overlapping yarn, and machine faults can result in defects that impact the fabric's quality, cost, and sales. As shown in Figure 1, common fabric defects are present. For a long time, manual inspection was the primary method for detecting fabric defects. However, this approach had many disadvantages. First, manual detection was slow; Secondly, it was extremely susceptible by subjective factors. This may lead to false detection and missed detection. Thus, there is an urgent need for automatic detection methods based on computer vision.

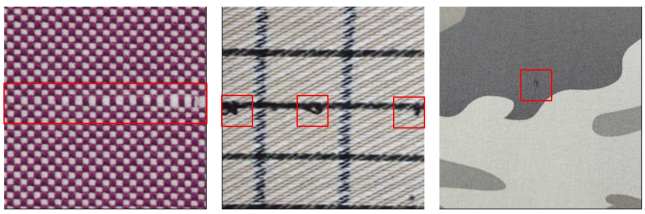

Figure. 1. Three common defects. The red box is the location of the defects.

Convolutional neural networks have become widespread in image recognition, detection, classification, and segmentation as a result of the advancement of computer vision. And convolutional neural networks have achieved significant success in defect detection [5]. Therefore, many researchers have proposed various networks for industrial fabric defect detection [6]. Liu et al. proposed an improved YOLOv4 algorithm and designed a new SPP structure that provided more precise detection of fabric defects [7]. Fan et al. proposed a method for fabric defect detection based on a combination of convolutional neural networks (CNNs) and variable autoencoders (VAEs) [8]. K. Gopalakrishnan et al. applied a deep learning technique to detect structural defects using convolutional neural networks and multiple optimizers [9]. However, in complex industrial environments, many factors can cause a variety of defect types and sizes, resulting in a poor detection result. However, none of the above studies discuss the detection of small defects or the generalizability of the networks. To address the above issue, we propose the Improved-YOLOX network, which can achieve a better balance between detection accuracy and speed, and improve detection accuracy for small defects. The experiments demonstrated the network's performance on three datasets. The contributions of this paper can therefore be summarized as follows:

ⅰ. We propose Improved-YOLOX for fabric defect detection to face complex industrial environments.

ⅱ. To enhance detection accuracy, we design an attention module that combines channel attention and spatial attention.

ⅲ. We introduce Dynamic convolution in the detection head to improve the detection of small defects. The VariFocal loss function are used to balance positive and negative samples. And the GIOU loss function solves the problem of prediction frames and actual frames overlapping.

The rest of the paper is organized as follows. Section 2 presents related work, including traditional methods and deep learning methods. Section 3 provides a detailed description of the proposed method. Section 4 shows the experimental and discussion of the results. Finally, the whole paper is summarized, and the focus of future work is presented.

II. RELATED WORK

This section reviews defect detection methods based on computer vision, including traditional methods and deep learning methods.

A. *Traditional method*

Traditional defect detection methods are divided into three methods based on statistic, structure, and filter method. The statistical method employs the difference between the gray value distribution characteristics of defects and the background to identify defects. Li et al. extracted and selected salient histogram feature to obtain feature vectors that could effectively distinguish defective and defect-free fabric images [10]. For the model method, Liu et al. proposed a low-rank decomposition for fabric defect detection in their method [11]. A new low-rank decomposition model was

Manuscript received March 14, 2023; revised November 3, 2023.

Changzheng Chen is a lecturer of Computer School, Xi'an Aeronautical Institute, Xi'an, China. (e-mail:22005105@qq.com).

developed in order to extract the sparse regions containing defective pixels and to perform thresholding. For the filtering method, the image is converted into the frequency domain, and the defect location is determined by analyzing the spectral characteristics. Qin et al. proposed a novel algorithm for fabric defect detection that combines residual energy distribution and Gabor features [12]. However, the above traditional methods necessitate painstakingly designed templates to extract features, as well as classifier and threshold segmentation to identify defects. Meanwhile, this method is time-consuming and are not suitable for defect detection in industry.

B. *Deep learning method*

Traditional detection methods have many disadvantages when applied to actual industrial production. So deep learning based on convolutional neural networks is widely used for defects detection due to its high accuracy and adaptability to the scene [13]. At present, defect detection based on neural networks is divided into the following two categories: one-stage networks, such as the YOLO series and SSD series. And two-stage networks, such as the R-CNN series. The one-stage network focuses on improving the speed of the network [14]. Zhou et al. could effectively detect fabric defect by improving YOLOv5s [15]. In addition, through a series of optimization measures, the accuracy and speed of the detection was enhanced, which had a certain industrial value. Jing et al. proposed a fabric defect detection method based on the improved YOLOv3 model and used a k-means algorithm for the dimensional clustering of the target frame to achieve better locate target box position [16]. Meanwhile, the combination of low-level features and high-level information could be used in fabric detection to improve the accuracy of the network. Xie et al. proposed an improved structural defect detection network based on SSD and added full convolution extrusion excitation (FCSE) module to increase the model's detection accuracy [17]. The two-stage network primarily improves the detection accuracy of the network [18]. Li et al. proposed a Cascade R-CNN fabric defect detection, which used a single input image to scale into multiple images with different resolutions for training, carried out dimensional clustering on the size of defects in datasets and adopted soft non-maximum suppression instead of traditional non-maximum suppression [19]. You et al. proposed a feature pyramid network based on Faster R-CNN [20]. This network combined features from multiple layers and directions to enhance the detection and localization of targets. When faced with complex industrial scenarios, despite the fact that the deep learning approach yielded superior results, there are still numerous aspects that must be enhanced. Simultaneously, the balance between detection accuracy and speed is an important challenge.

III. METHOD

YOLOX, as a new generation of the YOLO network, achieves an excellent speed-to-performance balance [21]. The YOLOX structure is comprised of three components: the backbone network, the neck network, and the detection head. In the backbone network, the CSPDarknet53 structure is used to extract the feature of the input image. In the neck network, different levels of information are fused to enhance feature extraction to improve detection capability of the model. In the detection head section, it enables the classification and location of defects. In sophisticated industrial environments, however, YOLOX's detection capabilities perform poorly. Therefore, we propose Improved-YOLOX network based YOLOX.

Our proposed Improved-YOLOX network is shown in Figure 2. First, to enhance the model's robustness, the CASA attention module is embedded within the YOLOX backbone section. Secondly, to accomplish the detection of small defects, we implement Dynamic convolution [22] in the detection head section. Finally, the VariFocal Loss function [23] and the GIOU Loss functions [24] are utilized to enhance the effectiveness of the model.

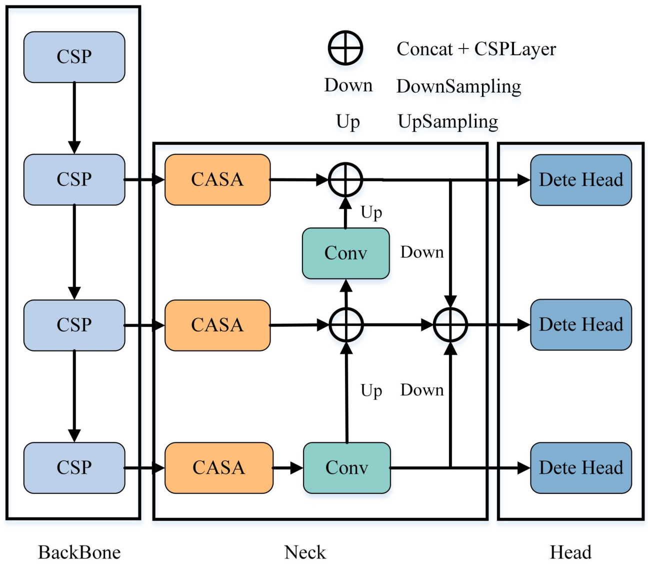

Figure. 2. Improved-YOLOX structure. The network is divided into three main sections: the backbone section, the neck section, and the detection head section.

A. *CASA Attention Module*

Researchers have increased the depth and width of the network continuously to enhance the network's detection capabilities [25-26]. This method, however, increased model size and computational effort and was susceptible to information redundancy. In the current study, the attention module improved the detection capability of the network through adaptive feature refinement of the input features [27]. However, most of the studies employ channel attention or spatial attention, which cannot extract features in a comprehensive way [28-30]. In this paper, we propose an attention module combining channel attention and spatial attention (CASA) to improve the detection of industrial defects in networks. Our proposed CASA module is shown in Figure 3. CASA module uses channel attention and spatial attention to emphasize meaningful features, which can effectively help information flow in the network and advise it to concentrate on "where" and "what".

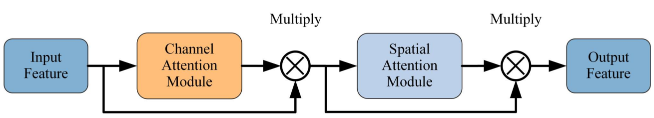

Figure. 3. CASA Attention Module.

The CASA module first performs channel attention, which can be viewed as a computational unit that enhances the network's ability to express features and concentrate on the "what" is the most essential information part of the input. The channel attention module is shown in Figure 4. First, the global average pooling is performed to extract the precise location information as follows: for the input feature map C×H×W, pooling is performed according to the X and Y directions to generate feature maps of size C×H×1 and C×1×W respectively. Then, the C×1×W feature map is transformed to generate a C×W×1 feature layer, followed by a concatenation operation, and a dimensionality reduction and activation operation using 1×1 convolution. Finally, the generated feature map is split along the spatial dimension, and the 1×1 convolution is used to perform upgrading and activation operations, and the output is obtained by multiplying with the input.

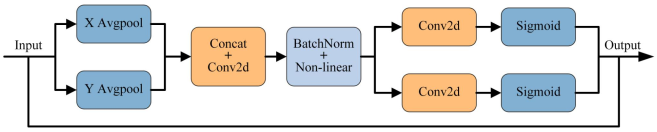

Figure. 4. Channel Attention Module.

Next, the CASA module conducts spatial attention, focusing on the "where" of crucial input information. The spatial attention module is shown in Figure 5. First, the average pooling features and maximum pooling features across channels are generated, which aggregates all channel information. This information is then concatenated and convolved through a standard convolutional layer to generate a spatial attention feature layer. Finally, the weights are obtained by the sigmoid function and are multiplied with the input to obtain the final output.

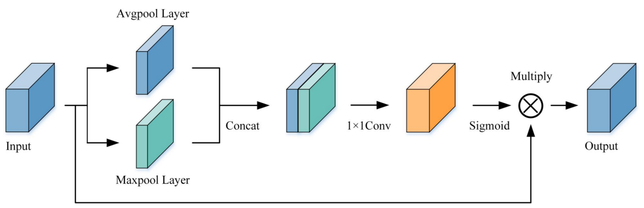

Figure. 5. Spatial Attention Module.

We embed the proposed attention module into the backbone network, which can better extract the channel and spatial information in the network to improve the detection capability of the network.

B. *Optimize Detection Head*

Small defects are frequently generated in fabric production for a variety of factors, resulting in poor detection of the network. Therefore, the detection head must be optimized to improve the detection of small targets. Dynamic convolution is introduced in the detection head. The Dynamic convolution is shown in Figure 6. It does not increase the depth and width of the network. At the same time, it has stronger representation capability, which can achieve more accurate prediction of targets and improve network performance. Dynamic convolution employs different convolution kernels, and each convolution kernel also has different weight coefficients. The weight coefficients are generated through using the attention module. The softmax layer can generate unique weights for each convolution kernel. After getting the weight coefficients, they are multiplied by the corresponding convolution and added together to produce the final convolution weights. Then, it performs matrix multiplication with the input to get the output and finally complete the Dynamic convolution operation.

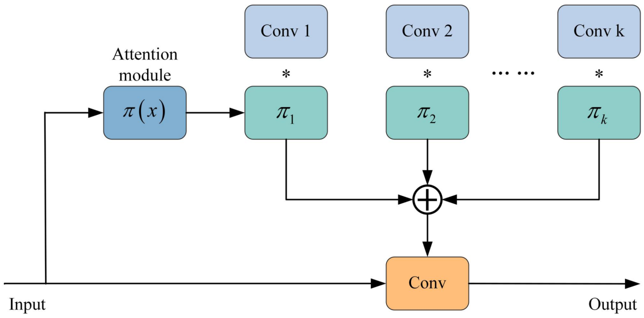

Figure. 6. Dynamic convolution.

C. *Loss function*

The loss function in YOLOX is composed of three parts: classification loss, confidence loss, and localization loss. In classification loss, the BCE loss function is used to calculate the loss based on the true frame and the prediction frame. In the confidence loss, the Varifocal loss function is utilized to address the sample imbalance issue more effectively. In the localization loss, the IOU loss is substituted with the GIOU loss function, which optimizes the situation in which the true frame does not intersect with the predicted frame. Our approach is explained in detail below.

1) Classification loss

In the classification task, the BCE loss function is calculated based on the predicted results of the real frame and the true frame. The BCE loss function is defined as follows:

$$L_{\mathrm{Cls}} = -[y \log \hat{y} + (1-y)\log(1-\hat{y})] \quad (1)$$

where L_{Cls} represents the BCE loss function, y is the value of the true frame, and $\hat{y}$ is the output of the prediction head.

2) Confidence loss

In the real industry, datasets frequently contain unbalanced positive and negative samples, which can have an effect on accuracy. Therefore, we introduce the Varifocal loss function to solve this problem. In order to address the positive and negative sample imbalance problem, an α factor is introduced in the negative loss term to achieve balance while simultaneously weighting the positive sample q. If a positive sample has a higher IOU, its contribution to the loss will be larger. This focuses the training on those high-quality positive examples that get higher APs than those of lower quality. The Varifocal loss function is defined as follows:

$$L_{\mathrm{Obj}} = \begin{cases} -q(q\log(p)+(1-q)\log(1-p)) & q>0 \\ -\alpha p^{\gamma}\log(1-p) & q=0 \end{cases} \quad (2)$$

where q is the true sample and p is the prediction score (q=0 is the detected negative sample and q>0 represents the detected positive sample).

3) Localization *loss*

After the feature points have been determined, the real frame and the predicted frame are used to calculate the localization loss. The overlap between the two frames, however, makes it impossible to identify the defects. Therefore, we use the GIOU loss function, as it considers both overlapping and non-overlapping regions. In the meantime, it can more accurately reflect the overlap between the real frame and the predicted frame. And the greater the overlap, the higher the robustness of the model. GIOU loss function is defined as follows:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

$$L_{\mathrm{Reg}} = IoU - \frac{|C \setminus (A \cup B)|}{|C|} \quad (4)$$

where A is the area of the real frame, B is the area of the predicted frame and C is the area of the minimum external frame.

Therefore, the overall loss function of YOLOX in this paper is as follows.

$$L_{Tatal} = L_{Cls} + L_{Obj} + L_{\mathrm{Reg}} \quad (5)$$

IV. EXPERIMENT

In this section, we will evaluate our model using a variety of experiments. The trained data and the detected data are presented on a local workstation. The configuration is GeForce RTX3090, 24GB memory, and the software runs on Windows 10.0, Python 3.7, and the PyTorch 1.7 deep learning framework. In the experiments, the batch size was set to 16, SGD was the optimizer, the learning rate was set to 1e-2, and a cosine learning rate schedule was used in training model.

A. Datasets

In our experiments, we will use three fabric defects datasets from actual industrial scenarios, namely Grid Fabric, Red Dotted Fabric, and Camouflage Fabric. Each dataset contains various types of defects, including line breaks, overlaps, etc. The number of datasets is shown in Table 1. Each dataset is split into a training set and a test set, with 90% of the dataset being used for network training and 10% of the dataset for network testing.

TABLE I: Data set

Dataset	Train Set (number)	Test Set (number)	Total Set (number)	Image Size
Grid Fabric	334	37	371	256×256
Red Dotted Fabric	378	42	420	256×256
Camouflage Fabric	312	35	347	256×256

B. Evaluation indicators

Mean Average Precision (mAP), an evaluation metric for measuring the performance of a target detection algorithm is obtained by taking a combined weighted average of the average precision (AP) for all classes. The formula for calculating mAP is as follows:

$$mAP = \frac{\sum_{i=1}^{N} AP_N}{N} \quad (6)$$

where N represents all classes, AP represents average precision for all class.

FPS, an evaluation metric for the speed of a target detection algorithm, refers to frames per second transmission. The more frames per second, the smoother the action displayed. The formula for calculating FPS is as follows:

$$FPS = \frac{Total\ Frame}{Total\ Time} \quad (7)$$

where *Total Frame* represents the number of transferred frames and *Total Time* represents the time it takes to transmit the number of frames.

GFLOPs, G floating point operations, is used to evaluate the amount of model computation. The formula for calculating GFLOPs is as follows:

$$FLOPs = 2HW\left(CiK^2 + 1\right)Co \quad (8)$$

$$GFLOPs = 10^9 FLOPs \quad (9)$$

where C_i is the number of channels input, C_o is the number of channels output and K is the size of the convolutional kernel.

C. Experiments

1) CASA attention module comparative experiments

We embedded the CASA attention module into the YOLOX backbone network and evaluated and validated the performance on the Grid Fabric dataset. We compared it with four attentional modules: ECA, CA, SE, and CBAM. Five attention modules were embedded in the backbone network in the same way to demonstrate the effectiveness of our proposed module. The experimental results are shown in Table 2.

TABLE II: mAP of different attention modules on Grid Fabric Set.

Network	YOLOX	+ECA	+SE-Net	+CBAM	+CA	**+CASA**
mAP(%)	88.78%	89.38%	92.05%	90.82%	92.64%	**92.98%**

According to the results, the accuracy of our attention module on the Grid Fabric dataset was improved by 4.2%. In comparison with the other four attention modules, our module had the highest accuracy. We discover that CASA can effectively improve the robustness of the network. Although the information obtained at the backbone network is simple, it still contains rich texture profile information. Consequently, the attention module is embedded behind the backbone network to improve the channel and spatial features.

2) Loss function fixed reference experiment

The Varifocal loss function was introduced as confidence loss. Since a variable factor α was incorporated into the equation, we experimented with its various value using the Grid Fabric dataset in order to ensure higher accuracy on our fabric defect dataset and to guarantee the validity of the detection. In the experiments, we took 0.1 to 0.4 for α and 2 for γ. Four experiments were conducted with varying values (0.1, 0.2, 0.3, 0.4) to verify the contribution of different values for the positive and negative samples. The experimental results are shown in Table 3. mAP was the highest when α = 0.4. and compared with the original YOLOX model, which accuracy can reach 94.11%. In dataset, the negative sample ratio tends to be higher, and α factor is used to weigh the positive and negative samples. Thus, the positive and negative samples are balanced, and there is no longer a high AP value and a low value for the negative samples.

TABLE III: mAP(%) for different values of the loss function variable factor.

MODLE	LOSS	mAP(%)
YOLOX	α=0.1	94.01%
	α=0.2	93.80%
	α=0.3	93.40%
	α=0.4	94.11%

3) Ablation experiment

We conducted ablation experiments on two fabric defect datasets to further investigate the impact of each improvement on the network. The modules were added in the following order: CASA attention module, prediction head improvement, and loss function improvement. The results of the ablation experiments are shown in Table 4 and 5. The detection capability of the model is enhanced by the attention module, which enables the network to increase the weight of usable features from both spatial and channel dimensions. An average improvement of 2.58% was calculated on two datasets. Utilizing dynamic convolution kernels and assigning various weights to different feature layers, Dynamic convolution is used in the prediction head improvement. It can improve the expressiveness of the model and increase the detection of small targets by increasing an average of 1.42% on two datasets. Using the Varifocal loss solves the sample imbalance problem and the GIOU loss solves the problem of overlap between predicted and real frames, allowing for a more precise location and identification of the defect. With the improvement of the loss function, the accuracy was increased by an average of 0.40% on the datasets. The experimental results demonstrate the effectiveness of our proposed method and guarantee the accuracy of the model while reducing its size.

TABLE IV: Qualitative experiments for the proposed modules on Grid Fabric.

YOLOX	+CASA	+Prediction head	+Loss function	mAP(%)
√				88.78%
√	√			92.98%
√	√	√		93.90%
√	√	√	√	94.11%

TABLE V: Qualitative experiments for the proposed modules on Camouflage Fabric.

YOLOX	+CASA	+Prediction head	+ Loss function	mAP(%)
√				92.81%
√	√			93.78%
√	√	√		95.70%
√	√	√	√	96.30%

4) Comparative experiments

To validate the accuracy of the models in this paper, it was compared with five models, namely SSD, Efficient, YOLOv5, YOLOX, and YOLOv4-tiny. All models were trained using three fabric datasets in order to evaluate their performance. The results are shown in Table 6 and 7, and the visualization results are shown in Figure 7.

Comparing the data results reveals that our method obtains superior detection speed and detection accuracy. The YOLOv5 had an advantage in detection speed, which reached 38 FPS and met the production requirements of the industry. However, the average detection accuracy was 92.32%, which was lower than our network. The original YOLOX could reach 35 FPS in speed, but its accuracy was 4.27% lower in accuracy than our model. YOLOv4-tiny is a lightweight network and improvements in detection speed has an impact on the network accuracy. Despite achieving 50 FPS in speed, YOLOv4-tiny's detection accuracy did not meet industrial requirements. Compared to the SSD network, the accuracy of detection in Red Dotted Fabric reached 96.10%, but the size of the model and number of parameters were larger than our model. Compared to the Efficient network, the detection speed of Efficient was only 17 FPS, which was insufficient for real-time detection. In summary, our model outperforms other networks in terms of detection speed and accuracy.

According to the results of the visualization, we find that YOLOv5, YOLOX, and YOLOv4-tiny had missed detection. Although the SSD and Efficient detection results meet the industrial requirements, they were slower than our method. Through comprehensive consideration, our model was able to enhance detection speed while maintaining accuracy.

TABLE VI: mAP(%) comparison experiments with different networks on three datasets.

Networks	YOLOX	YOLOv5	YOLOv4-tiny	SSD	Efficient	Improved-YOLOX
Grid Fabric	88.78%	91.02%	77.38%	91.37%	89.49%	94.11%
Red Dotted Fabric	95.01%	94.38%	69.14%	96.10%	77.43%	99.00%
Camouflage Fabric	92.81%	91.57%	85.07%	91.21%	92.36%	96.30%

TABLE VII: Comparison of the four evaluation indicators for different networks.

Networks	YOLOX	YLOLv5	YOLOv4-tiny	SSD	Efficient	Improved-YOLOX
Parameter(M)	8.938	7.066	6.057	26.151	3.874	7.257
GFLOPs(G)	4.281	2.638	2.635	62.648	5.234	2.699
FPS	35	38	50	30	17	39

V. Conclusion

In this paper, a fabric defect detection model based on Improved-YOLOX is proposed in fabric production. The attention module is designed, which employs channel attention and spatial attention. We introduce Dynamic

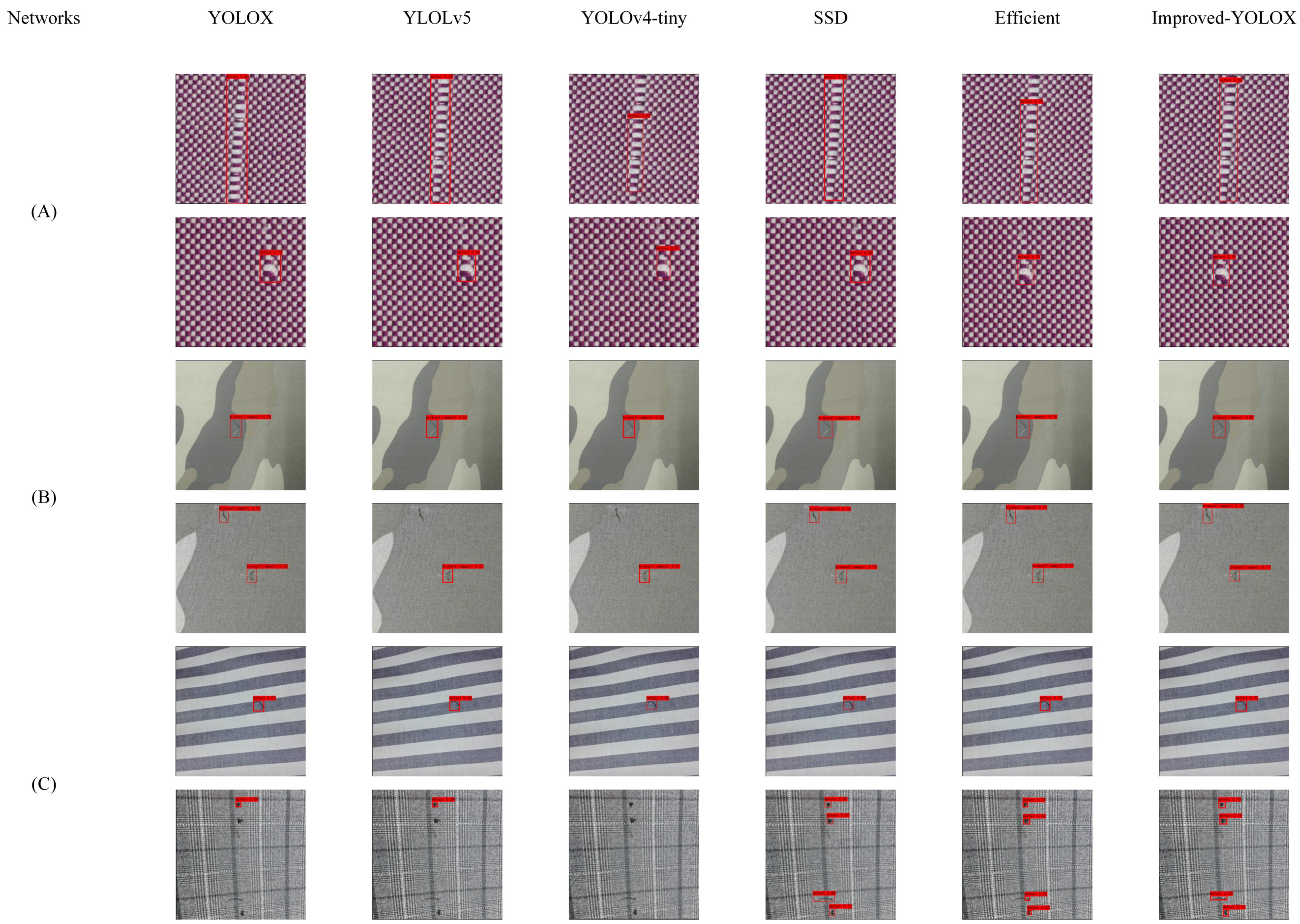

Figure 7 Visualization results. (A) represents the results of the test on the Red Dotted Fabric. (B) represents the results of the test on the Camouflage Fabric. (C) represents the results of the test on the Grid Fabric.

convolution in the detection head section to improve the detection of smaller targets. The Varifocal Loss function and GIOU Loss function are utilized to enhance the detection performance of the network. The experimental results indicate that the accuracy can reach an average of 96.47% in three fabric datasets, and the detection speed can achieve 39 FPS. In conclusion, our network achieves better performance in detection speed and detection accuracy and satisfies the requirements of industrial production. Compared to manual methods, our approach has clear advantages. Future research will focus on the following two directions: Firstly, the few samples or samples that are difficult to collect is a significant challenge for model. Secondly, the ability to generalize model is also an essential research topic.

References

[1] Yin, Z., Lu, H., Gan, L., & Zhang, Y. (2022). Electronic Fibers/Textiles for Health-Monitoring: Fabrication and Application. Advanced Materials Technologies, 2200654.

[2] Ma, L., Wu, R., Liu, S., Patil, A., Gong, H., Yi, J., ... & Wang, Z. L. (2020). A machine-fabricated 3D honeycomb-structured flame-retardant triboelectric fabric for fire escape and rescue. Advanced materials, 32(38), 2003897.

[3] Rocha, H., Semprimoschnig, C., & Nunes, J. P. (2021). Sensors for process and structural health monitoring of aerospace composites: A review. Engineering Structures, 237, 112231.

[4] Çelenk, E., & Tokan, N. T. (2022). All-textile on-body antenna for military applications. IEEE Antennas and Wireless Propagation Letters, 21(5), 1065-1069.

[5] Tulbure, A. A., Tulbure, A. A., & Dulf, E. H. (2022). A review on modern defect detection models using DCNNs–Deep convolutional neural networks. Journal of Advanced Research, 35, 33-48.

[6] Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object detection in 20 years: A survey. arXiv preprint arXiv:1905.05055.

[7] Liu, Q., Wang, C., Li, Y., Gao, M., & Li, J. (2022). A Fabric Defect Detection Method Based on Deep Learning. IEEE Access, 10, 4284-4296.

[8] Fan, J., Wong, W. K., Wen, J., Gao, C., Mo, D., & Lai, Z. (2021). Fabric Defect Detection Using Deep Convolution Neural Network. AATCC Journal of Research, 8(1_suppl), 143-150.

[9] Gopalakrishnan, K., & Vanathi, P. T. (2022). Fabric Defect Detection Using Deep Learning Techniques. In Ubiquitous Intelligent Systems: Proceedings of Second ICUIS 2022 (pp. 101-113). Singapore: Springer Nature Singapore.

[10] Li, M., Wan, S., Deng, Z., & Wang, Y. (2019). Fabric defect detection based on saliency histogram features. Computational Intelligence, 35(3), 517-534.

[11] Liu, G., & Li, F. (2022). Fabric defect detection based on low-rank decomposition with structural constraints. The Visual Computer, 38(2), 639-653.

[12] Qin, W., Wen, H., & Li, F. (2022). Fabric defect detection algorithm based on residual energy distribution and Gabor feature fusion. The Visual Computer, 1-15.

[13] Wan, G., Fang, H., Wang, D., Yan, J., & Xie, B. (2022). Ceramic tile surface defect detection based on deep learning. Ceramics International, 48(8), 11085-11093.

[14] Hang, J., Sun, H., Yu, X., Rodríguez-Andina, J. J., & Yang, X. (2022). Surface Defect Detection in Sanitary Ceramics Based on Lightweight Object Detection Network. IEEE Open Journal of the Industrial Electronics Society, 3, 473-483.

[15] Zhou, S., Zhao, J., Shi, Y. S., Wang, Y. F., & Mei, S. Q. (2022). Research on improving YOLOv5s algorithm for fabric defect detection. International Journal of Clothing Science and Technology, (ahead-of-print).

[16] Jing, J., Zhuo, D., Zhang, H., Liang, Y., & Zheng, M. (2020). Fabric defect detection using the improved YOLOv3 model. Journal of engineered fibers and fabrics, 15, 1558925020908268.

[17] Xie, H., Zhang, Y., & Wu, Z. (2021). An improved fabric defect detection method based on SSD. AATCC Journal of Research, 8(1_suppl), 181-190.

[18] Huang, F., Wang, B. W., Li, Q. P., & Zou, J. (2021). Texture surface defect detection of plastic relays with an enhanced feature pyramid network. Journal of Intelligent Manufacturing, 1-17.

[19] Li, F., & Li, F. (2021). Bag of tricks for fabric defect detection based on Cascade R-CNN. Textile Research Journal, 91(5-6), 599-612.

[20] You, S., Ye, S., Cui, C., & Wang, Y. (2022, July). Fabric Defect Detection Based on Improved Faster R-CNN and Two-Step Clustering Method. In 6th International Technical Conference on Advances in Computing, Control and Industrial Engineering (CCIE 2021) (pp. 282-292). Singapore: Springer Nature Singapore.

[21] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

[22] Chen Y, Dai X, Liu M, et al. Dynamic convolution: Attention over convolution kernels[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11030-11039.

[23] Zhang, H., Wang, Y., Dayoub, F., & Sunderhauf, N. (2021). Varifocalnet: An iou-aware dense object detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8514-8523).

[24] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 658-666).

[25] Mao, Y., Wang, S., Yu, D., & Zhao, J. (2021). Automatic image detection of multi-type surface defects on wind turbine blades based on cascade deep learning network. Intelligent Data Analysis, 25(2), 463-482.

[26] Lian, J., Jia, W., Zareapoor, M., Zheng, Y., Luo, R., Jain, D. K., & Kumar, N. (2019). Deep-learning-based small surface defect detection via an exaggerated local variation-based generative adversarial network. IEEE Transactions on Industrial Informatics, 16(2), 1343-1351.

[27] Ji, C., Wang, H., & Li, H. (2023). Defects detection in weld joints based on visual attention and deep learning. NDT & E International, 133, 102764.

[28] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

[29] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.

[30] Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13713-13722).