Research on Fabric Defect Detection Technology Based on RDN-LTE and Improved DINO

Li Yao, Zhongqin Chen, and Yan Wan(✉)

Donghua University, Shanghai 201620, China
winniewan@dhu.edu.cn

Abstract. In order to solve the problem of detecting various types of complex fabric defects such as different scale sizes, high fusion with the background and extreme aspect ratios generated in actual production environment, this paper proposes a defect detection method that combines super-resolution reconstruction technology with object detection technology. Firstly, the dataset is reconstructed using the super-resolution reconstruction technology RDN-LTE, which effectively solve the problem of high fusion between defects and background. Furthermore, the copy-paste technology is employed for data augmentation to enhance model robustness. Then the dataset is fed into the detection network DINO for training. To improve the receptive field of the model, Swim Transformer is used as the backbone network of the model instead of ResNet-50, and the scale features extracted by the model are increased from 4 to 5. The deformable attention mechanism is also introduced in the third and fourth stages of Swim Transformer to enhance the global relationship modeling. Finally, multi-scale training method is introduced to capture the defect features at different scales to further improve the model detection effect and training speed. The results of the three kinds of comparative experiments show that the method based on RDN-LTE and improved DINO has a better overall recognition rate for multiple kinds of fabric defects than other current methods.

Keywords: Fabric defect detection · RDN-LTE · DINO · Swim Transformer · Copy-paste

1 Introduction

The textile industry plays an important role in China's economic development, and the quality of fabrics has a great impact on the economic benefits of the textile industry. Statistics show that defective fabrics can cost enterprises 45% to 65% of their profits [1], so the accurate detection of defects is a crucial step in enterprise production. The types of defects are complex and diverse, for example, some types of defects are extremely small, and some types of defects are highly integrated with the background. Even experienced workers may have problems

B. Sheng et al. (Eds.): CGI 2023, LNCS 14497, pp. 137–148, 2024.
https://doi.org/10.1007/978-3-031-50075-6_11

such as missed detection, false detection, slow detection speed, long time consumption. Object detection algorithms have been applied in many fields, such as [2,3], effectively saving labor and improving enterprise profits, so it is necessary to research automated defect detection.

In the model-based method, [4] constructed a new low rank decomposition model to handle defects. [5] first preprocessed the dataset with denoising, and then used the Gaussian Markov random field (GMRF) model to model fabric texture features. However, model-based methods have the problem of difficulty in identifying small defects and high computational complexity.

In the statistical analysis method, [6] obtained the set of fabric texture feature vectors based on Haralick parameters. However, statistical analysis requires a large number of defect datasets, and it is difficult to accurately describe defects with complex background.

In the spectrum analysis method, [7] proposed a new method for local tuning of amplitude spectrum to preserve defect areas while suppressing background patterns. [8] used Gabor filters to enhance the contrast between defects and surrounding background textures, while obtaining the Histogram of Oriented Gradient (HOG) features of the background image, and identified defects based on the differences between defect features and HOG features.

In the one-stage algorithm based on deep learning, [9] added deformable convolution network (DCN) to the backbone of YOLOv4 in order to improve the ability of the model to describe geometric changes. Although the one-stage algorithm is fast in detection, the detection accuracy is low due to the absence of the link to generate candidate regions.

In the two-stage algorithm, [10] used the cascade approach to first identify defects as bars or blocks in the Inception-ResNet-v2 network, and then adjusted the anchor frame ratio to detect them in the Faster R-CNN network. The detected defect categories were not comprehensive and did not have generalization ability.

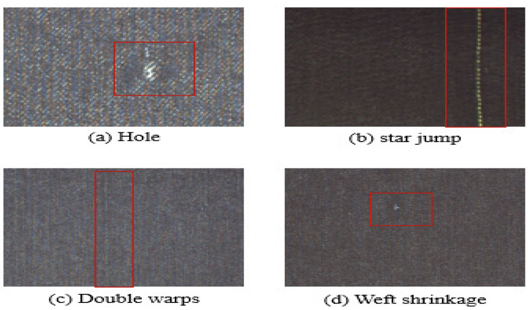

Fig. 1. Comparison of hard-to-detect defects and easy-to-detect defects.

Although the above methods can achieve the detection of some obvious defect types, as shown in (a) and (b) in Fig. 1, the detection effect is not good for complex defect types with different scale sizes, high fusion with background, and extreme aspect ratios, as shown in (c) and (d) in Fig. 1, and these types of defects are also common, this paper proposes a universal method for detecting multiple types of defects in actual production environments. Firstly, to effectively address the issue of high fusion between defects and background, RDN-LTE fabric defect image reconstruction model is proposed, thus obtaining a clearer texture dataset for fabric defects. Further utilizing copy-paste [11] technology for data augmentation to enhance the robustness of the model. Then the enhanced dataset is input into the detection network DINO [12] for training. In order to increase the receptive field of the model, the original backbone network ResNet-50 [13] is replaced by Swim Transformer [14]. At the same time, in order to better capture the shape and size characteristics of different types of defects, and enhance global relationship modeling, deformable attention [15] is introduced in the third and fourth stages of Swim Transformer. And increase the scale features extracted by the model from 4 to 5, in order to obtain more rich and diverse representation of defect features. Finally, multi-scale training is introduced to capture the defect features at different scales to improve the model detection effect and training speed. This paper detects 20 common types of fabric defects provided by the Alibaba Cloud Tianchi platform. The experiment proves that our method has a good improvement in the detection accuracy of different types of defects, especially for difficult to identify defect types. The three sets of comparative experiments demonstrate that our method is superior to other methods, the average detection accuracy reaches 86.7%.

2 Model Structure Based on RDN-LTE and Improved DINO

In response to the above issues, this paper proposes a fabric defect detection method based on RDN-LTE and improved DINO, with the model structure shown in Fig. 2.

The main idea of the model is to first solve the problem of high fusion between defects and background, which is a major difficulty in defect detection. Therefore, it is necessary to study the super-resolution reconstruction technology. The study found that compared to the Deeply-Recursive Convolutional Network (DRCN) [16], the local residual learning and dense feature fusion mechanism of the Residual Dense Network (RDN) [17] can better adapt to different types of defects and fully learn the texture features of the defects. Moreover, the introduction of Local Texture Estimator (LTE) [18] algorithm into RDN can further enhance the local texture feature representation of defects, help the model quickly capture the main frequency information of defects, thus accelerating the model convergence. Therefore, this paper proposes the RDN-LTE fabric defect image reconstruction model. Furthermore, copy-paste technology is used for data augmentation, allowing the model to better adapt to new datasets and enhance its robustness. The enhanced dataset is then fed into the detection network DINO. The receptive field of the original backbone network ResNet-50 is

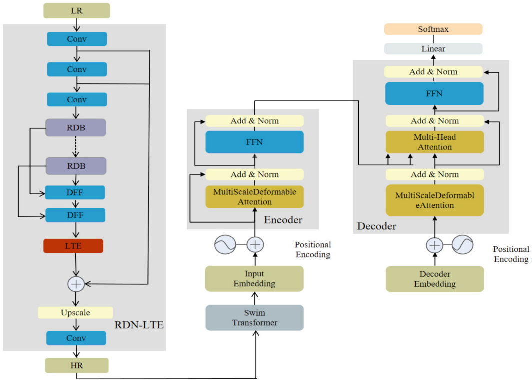

Fig. 2. Network structure based on RDN-LTE and improved DINO.

limited, which can only focus on local features of the defective images. It cannot fully learn the relationships between global features, making it prone to information loss. The study found that the adaptive pooling layer in Swim Transformer can dynamically adjust the pooling size based on the sizes of different defective images, thereby increasing the receptive field. In addition, the grouping convolution and cross layer connection technology in Swim Transformer can also be used. The grouping convolution can divide the defect feature map into multiple subgroups, and each subgroup can be convolved, thus further increasing the receptive field while keeping the model parameters unchanged. The cross layer connection technology can connect the shallow and deep defect feature maps, so that broader context information can be learned. Therefore this paper uses Swim Transformer as the backbone network instead of ResNet-50, and increases the scale features extracted from the model from 4 to 5 to capture more abundant defect features. However, during the model training process, it was found that Window-based Multi-head Self-Attention (W-MSA) in Swim Transformer perceives local defect information well, but may ignore the correlation between global features, and the Shifted Window-based Multi-head Self-Attention (SW-MSA) cannot self-adapt to defects of different scales due to the fixed size of the shifted window. Through experiments, it was found that deformable attention can enable the model to adaptively adjust the size and shape of the receptive field, flexibly adapt to different types of defects with different scale sizes, and can also enhance global relationship modeling, to a certain extent, reducing the computational complexity of the model. Therefore, deformable attention is introduced in the third and fourth stages of Swim Transformer. Finally, the introduction of multi-scale training methods can not only enable the model to learn defect features of different scales, improve the detection performance and robustness of the model, but also improve the computational efficiency of the model and accelerate its convergence.

2.1 RDN-LTE Network

This paper studies the RDN model and improves it to adapt to the fabric defect dataset. In order to adapt to the complex and diverse types of defects, this paper deepens shallow feature extraction network by adding a layer of convolution on top of the original structure. Then the extracted shallow defect features are input into residual dense block (RDB) for further processing, by utilizing dense connections in RDB modules, defect features of different scales can be obtained, and local feature fusion and residual learning mechanisms are utilized to enhance the representation of defect features. Subsequently, defect dense feature fusion is performed in the dense feature fusion (DFF) layer. Research has found that introducing the LTE algorithm into RDN model can effectively detect and repair local defect texture features, enhance the local expression ability of defect features, and learn the dominant frequency information of defects faster, thereby accelerating model convergence. Finally, the upsampling layer utilizes grouped sub-pixel convolution to generate high-resolution defect images. The RDN-LTE model is shown in Fig. 3.

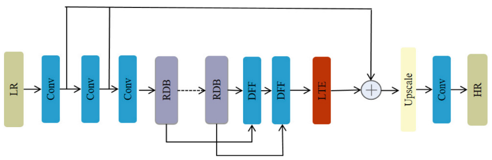

Fig. 3. Network structure of RDN-LTE.

Introducing LTE Algorithm. In order to further enhance the local texture feature representation of defects, this paper introduces the LTE algorithm, which estimates the Fourier coefficients corresponding to high-frequency defect texture information to assist RDN in obtaining more precise texture features during fabric defect image reconstruction. The LTE algorithm is shown in formula (1), Where θ represents the trainable weight of MLP, χ represents two-dimensional coordinates in continuous image domain, S denotes the predictive value space function of the decoder, ψ is used to estimate the principal frequency and the corresponding Fourier coefficients, $\mathbf{I}_{\uparrow}^{\mathbf{LR}}$ indicates a jump connection between the low-resolution defective image and the decoder.

$$\hat{\mathbf{s}}\left(\mathbf{x}\right) = \mathbf{s}\left(\mathbf{x}, \mathbf{I}^{\mathbf{LR}}; \Theta, \psi\right) + \mathbf{I}_{\uparrow}^{\mathbf{LR}}\left(\mathbf{x}\right) \tag{1}$$

The original dataset is used as the high-resolution image, and the 2× downsampling is used as the low-resolution image, and the dataset is made in DIV2K format and trained in the RDN-LTE model, the PSNR is continuously adjusted to obtain the desired image, and the model with LTE has a shorter training time. Figure 4 shows a comparison of the effects before and after reconstruction.

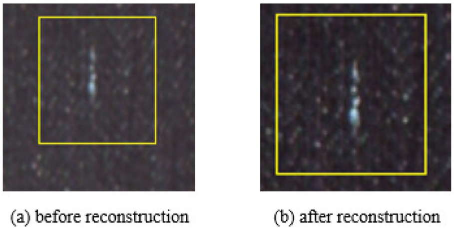

Fig. 4. Comparison of three silk type defect before and after reconstruction.

2.2 Copy-Paste Data Augmentation

In order to improve the robustness of the model, this paper uses copy-paste data augmentation technology to expand the dataset by randomly copying each type of defect and performing random horizontal or vertical flipping, random zooming in and out, and randomly adjusting the contrast within a certain range. Figure 5 shows a comparison of the effects before and after copy-paste.

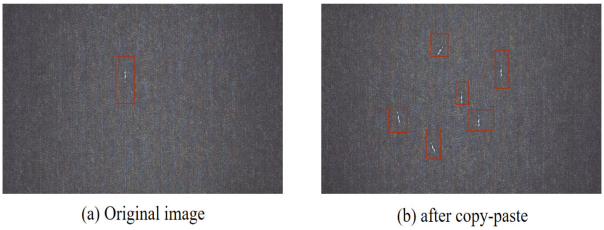

Fig. 5. Comparison of before and after copy-paste.

2.3 Improved DINO

The original DINO model consists of a backbone network, multi-layer Transformer encoder, multi-layer Transformer decoder, and multiple prediction head modules. First, the fabric defect image is input to the backbone network for feature extraction, and the backbone network ResNet-50 is commonly used to extract the features of 4 scales, and then enter the multi-layer transformer encoder for feature enhancement and global relationship modeling. At this stage, the contrast noise reduction training method of the DINO model is fully utilized, specifically, adding noise to the edge of the anchor frame of the defect, research has found that using this contrastive denoising training method can enable the model to better handle noise interference and increase its adaptability to defects of different shapes and sizes. Then, the hybrid query and look forward twice algorithm are simultaneously used in the decoder to obtain defect context information and optimize the size and position of the defect anchor box. Finally, perform the final defect classification and regression in the prediction header module.

However, the types of defects are complex and diverse, and the resolution of the reconstructed image becomes larger. The receptive field of ResNet-50 is limited, and it can only focus on the local features of the defective image, unable to fully learn the relationship between the global features, and prone to information loss, resulting in poor detection effect of the model. Therefore, the backbone network needs to be improved.

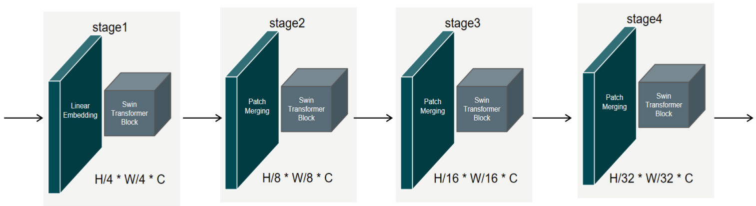

Fig. 6. Network structure of Swim Transformer.

Using Swim Transformer as Backbone Network. Research has found that Swim Transformer, through its flexible hierarchical structure and shifted window design, can obtain defect features at different scales while maintaining the same model parameter quantity, and weighted fuse shallow and deep features. In addition, the adaptive pooling layer of Swim Transformer can dynamically adjust the pool size according to different defect sizes, thus increasing the receptive field, which is very suitable for training multi-species fabric defect dataset. The model structure of Swim Transformer is shown in Fig. 6, which consists of four stages, in each stage the images are downsampled by 4×, 8×, 16× and 32×, and the sampled images are subjected to Patch_Embedding operation, while each individual patch is subjected to W-MSA computation, and then global information connection is established by shifted window, this structure can better adapt to defects of different scale sizes.

Introducing Deformable Attention Mechanism. The main attention mechanisms used in Swim Transformer are W-MSA and SW-MSA. Although W-MSA is able to reduce the computational complexity relative to the global attention mechanism, its slow receptive field growth rate limits the model's ability to model the reconstructed defect images. In addition, SW-MSA focuses on the spatial location of the defect, while ignoring the differences in the specific shape and size of defects. In contrast, the deformable attention mechanism can learn the shape and size differences of defects through the deformation module, while enhancing the connection between local and global receptive fields. This paper introduced deformable attention in the third and fourth stages of Swim Transformer, and the results showed a certain improvement in the detection effect for each type of defect. The deformable attention mechanism is shown in formula (2), where $x \in R^{C*H*W}$, p_q is the two-dimensional reference point for the query element, K represents the sampling point, m represents the attention head, $Amqk$ and $\triangle pmqk$ represent the attention weight and offset of the k-th sampling point in the m-th attention head, respectively.

$$DeformAttn\left(z_q, p_q, x\right) = \sum_{m=1}^{M} W_m \left[\sum_{k=1}^{K} A_{mqk} * W_m^{'} x\left(p_q + \Delta p_{mqk}\right)\right] \tag{2}$$

Introducing Multi-scale Training. The size of the input image is one of the important factors that affect the detection effect of the model. If the image size is too small, it is difficult for the model to capture the key features, while the image size is too large, it will increase the computation and memory consumption. To solve this problem, we adopt a multi-scale training approach. In multi-scale training, the model performs random scale transformations on the input image, such as scaling, cropping or rotating operations, and then trains on the transformed data. In this way, the model can learn features at different scales, thus improving its robustness and generalization ability.

3 Experiments

3.1 Experimental Environment and Dataset

All the experiments are based on the computing cloud platform, and the server uses two V100-SXM2-32 GB graphics cards. During the experiment, the learning rate is continuously adjusted, we try to set the learning rate to 1e−4, 1e−5 and 5e−5 respectively, and the results show that the best training effect is achieved when the learning rate is set to 5e−5. All experiments were conducted using multi-scale training methods.

The experimental dataset is derived from the publicly available fabric defect dataset on the Tianchi platform, containing 9576 data samples, of which there are 5913 images with defects, each containing one or more different defects. In this paper, we re-clean the dataset, and the defects are classified into 20 categories according to their characteristics, the category names are shown in Table 2, and the numbers 1 to 20 are used to indicate the corresponding categories in order.

3.2 Experiments Process and Results

The first experiment inputs the original dataset without any processing into the original DINO model with ResNet-50 as the backbone network to extract 4 scale features. The second experiment with Swim-L as the backbone network to extract 5 scale features. The third experiment reconstructed the dataset based on the second experiment. The fourth experiment uses copy-paste data augmentation technology. The fifth experiment introduces deformable attention mechanism. The detection results of each experiment are shown in Table 1. Table 2 shows the detection results for each type of defect in each round of experiments.

Table 1. Experimental results of each step of our algorithm.

Object detection algorithm	mAP@[IoU = 0.5]
(a) DINO (ResNet-50) Baseline	**65.9%**
(b) DINO+Swim-L	70.4%
(c) DINO+Swim-L+RDN-LTE	77.9%
(d) DINO+Swim-L+RDN-LTE+copy-paste	83.8%
(e) DINO+Swim-L+RDN-LTE+copy-paste+Deformable Attention	**86.7%**

Figure 7 shows the detection results of a representative defect type after each improvement experiment, with labels (a), (b), (c), (d), and (e) corresponding to each experiment in Table 1.

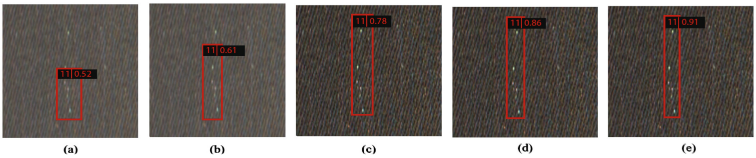

Fig. 7. The detection effect of hanging warp type defect in each step of our method.

3.3 Comparison Experiments

In order to verify the superiority of the method proposed in this paper, comparative experiments are conducted using Transformer based object detection algorithm Deformable DETR, one-stage algorithm YOLOv4, and two-stage algorithm Cascade R-CNN. The experimental results are shown in Table 3.

According to the comparative experimental results, our method performs better than other methods in fabric defect detection tasks. Figure 8 shows the comparison of our method with other methods in detecting knot defect. This type of defect has a small area, irregular shape, and high fusion with the background. In order to display the differences more clearly, we enlarged the image. It can be observed from the comparison graph that the detection effects of other methods are not satisfactory, and Yolov4 even produces false detection. This is because other methods require the design of appropriate anchor frame ratios, which makes it difficult to cover various defect shapes, especially for irregularly shaped defects. In addition, it is difficult for the model to learn effective defect features because the defect occupies a small area and is highly integrated with the background. In contrast, our method automatically determines the anchor box during the training process and first performs super-resolution reconstruction on the defect image to solve the problem of high fusion between the defect and the background. In addition, we introduce Swim Transformer, which can adaptively adjust the size of the pool based on the characteristics of defects, and better perceive the specific shape and size of defects by introducing deformable attention

Table 2. Detection results of each type of defect in each round of experiments.

Defect Type	experiments 1	experiments 2	experiments 3	experiments 4	experiments 5
Hole	95.5%	95.7%	96.4%	97.0%	97.1%
Water stains	51.9%	58.4%	69.5%	72.7%	75.5%
Three threads	76.6%	80.5%	86.3%	90.1%	93.1%
Knot	71.5%	73.6%	79.9%	85.9%	90.0%
Flower board jump	67.9%	70.4%	81.4%	86.9%	89.8%
Hundred feet	62.2%	71.3%	77.0%	81.8%	82.1%
coarse grain	56.8%	62.2%	70.7%	79.3%	83.9%
Thick warp	85.1%	85.8%	87.2%	93.8%	94.6%
loose warp	63.9%	66.1%	73.7%	79.0%	81.2%
Broken warp	44.2%	50.3%	66.2%	72.5%	77.2%
Hanging warp	51.6%	62.1%	77.8%	88.8%	90.5%
Thick weft	63.9%	70.5%	77.1%	75.9%	80.8%
Weft shrinkage	45.2%	48.1%	57.8%	77.0%	81.3%
Pulp spot	88.1%	88.5%	87.1%	91.6%	94.0%
Warping knot	72.5%	74.8%	78.1%	84.2%	92.9%
star jump	84.4%	85.6%	89.1%	99.2%	98.6%
Broken spandex	77.1%	80.5%	82.3%	84.5%	84.2%
Thin and dense paths	43.7%	50.1%	60.1%	68.4%	76.0%
Wear mark	60.6%	63.5%	81.1%	83.3%	83.1%
Double warps	55.1%	69.8%	79.1%	84.8%	88.2%

Table 3. Comparison with mainstream methods.

Object detection algorithm	mAP@[IoU = 0.5]
Deformable DETR (ResNet-50)Baseline	**51.5**%
Deformable DETR + Swim-L	66.7%
Deformable DETR + Swim-L+RDN-LTE	69.1%
Deformable DETR + Swim-L+RDN-LTE+copy-paste	72.7%
Deformable DETR + Swim-L+RDN-LTE+copy-paste+Deformable Attention	75.8%
Cascade R-CNN(ResNet-50)Baseline	**52.7**%
Cascade R-CNN + Swim-L	55.4%
Cascade R-CNN + Swim-L+RDN-LTE	59.7%
Cascade R-CNN + Swim-L+RDN-LTE+copy-paste	64%
Cascade R-CNN + Swim-L+RDN-LTE+copy-paste+Deformable Attention	66.7%
YOLOv4 (CSPDarknet53)	**54.3**%
YOLOv4+ Swim-L	50.7%
YOLOv4+ Swim-L+RDN-LTE	55.6%
YOLOv4+ Swim-L+RDN-LTE+copy-paste	61%
YOLOv4+ Swim-L+RDN-LTE+copy-paste+Deformable Attention	62.8%
Our method	**86.7**%

mechanisms. Finally, we also use multi-scale training to capture defect features at different scales, enabling the model to fully learn the features of defects. Therefore, our method has made significant improvements in defect detection.

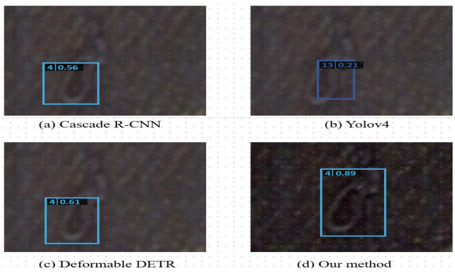

Fig. 8. Comparison of detection effects of knot type defect in different methods.

4 Conclusion

This paper proposes a detection method for various common defects in the actual environment of fabric production. Firstly, the defect dataset is reconstructed using super-resolution reconstruction technology RDN-LTE to solve the problem of high fusion between defects and background. Then, copy-paste technology is used to enhance the dataset and improve the robustness of the model. Finally, the processed dataset is input into the DINO detection model. In the DINO model, the Swim Transformer is introduced as the backbone network, and the deformable attention mechanism is introduced in the third and fourth stages of the network. In addition, multi-scale training is used to capture different scale defect features to improve the detection effect. Three sets of comparison experiments show that the method in this paper is better than other methods. However, due to the high performance requirements for GPU, high memory consumption, and long training time of the super-resolution reconstruction technology, further improvement is needed in subsequent work.

References

1. Selvi, S.S.T., Nasira, G.: An effective automatic fabric defect detection system using digital image processing. J. Environ. Nanotechnol. **6**(1), 79–85 (2017)
2. Chen, Z., Qiu, J., Sheng, B., Li, P., Wu, E.: GPSD: generative parking spot detection using multi-clue recovery model. Vis. Comput. **37**(9–11), 2657–2669 (2021). https://doi.org/10.1007/s00371-021-02199-y
3. Li, J., et al.: Automatic detection and classification system of domestic waste via multimodel cascaded convolutional neural network. IEEE Trans. Ind. Inf. **18**(1), 163–173 (2021)
4. Liu, G., Li, F.: Fabric defect detection based on low-rank decomposition with structural constraints. Vis. Comput. **38**(2), 639–653 (2022). https://doi.org/10.1007/s00371-020-02040-y
5. Xu, Y., Meng, F., Wang, L., Zhang, M., Wu, C.: Fabric surface defect detection based on GMRF model. In: 2021 2nd International Conference on Artificial Intelligence and Information Systems, pp. 1–4 (2021)

6. Tola, S., Sarkar, S., Chandra, J.K., Sarkar, G.: Sparse auto-encoder improvised texture-based statistical feature estimation for the detection of defects in woven fabric. In: Chakraborty, M., Jha, R.K., Balas, V.E., Sur, S.N., Kandar, D. (eds.) Trends in Wireless Communication and Information Security. LNEE, vol. 740, pp. 143–151. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-6393-9_16
7. Liu, G., Zheng, X.: Fabric defect detection based on information entropy and frequency domain saliency. Vis. Comput. **37**(3), 515–528 (2021). https://doi.org/10.1007/s00371-020-01820-w
8. Tang, X., Huang, K., Qin, Y., Zhou, C.: Fabric defect detection based on Gabor Filter and HOG feature. Comput. Measur. Control **26**(9), 39–47 (2018)
9. Liu, T., Chen, S.: YOLOv4-DCN-based fabric defect detection algorithm. In: 2022 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 710–715. IEEE (2022)
10. Zhao, Z., Gui, K., Wang, P.: Fabric defect detection based on cascade faster R-CNN. In: Proceedings of the 4th International Conference on Computer Science and Application Engineering, pp. 1–6 (2020)
11. Ghiasi, G., et al.: Simple copy-paste is a strong data augmentation method for instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2918–2928 (2021)
12. Zhang, H., et al.: DINO: DETR with improved denoising anchor boxes for end-to-end object detection. In: The Eleventh International Conference on Learning Representations (2022)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
14. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
15. Xia, Z., Pan, X., Song, S., Li, L.E., Huang, G.: Vision transformer with deformable attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4794–4803 (2022)
16. Kim, J., Lee, J.K., Lee, K.M.: Deeply-recursive convolutional network for image super-resolution. IEEE (2016)
17. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2472–2481 (2018)
18. Lee, J., Jin, K.H.: Local texture estimator for implicit representation function. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1929–1938 (2022)