Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Defect Detection of Fabric Printing with Differential Attention

Xinhai Li*, Qifeng Luo, Dehe Fan
Zhongshan Power Supply Bureau of Guangdong Provincial Power Grid Co., Ltd., Zhongshan, Guangdong, China
**Corresponding author: Xinhai Li (e-mail: zslixinhai@163.com)*

ARTICLE INFO

Received: 15 Nov 2024

Revised: 24 Dec 2024

Accepted: 12 Jan 2025

ABSTRACT

Fabric defect detection is a major quality control process in the textile industry. However, compared with common detection tasks, printed fabric defect detection has several difficulties as follows: printed fabric pattern texture is complex; defects are of numerous types and vary greatly in size and shape; most of the defects are extremely small.Aiming at these problems and difficulties in real scenes, this paper firstly adopts the idea of template matching in change detection to eliminate the background pattern texture features of target images by using template image features.Aiming at the problem of large scale difference of defects, this paper proposed a better modeling ability of Ratio DCN on the basis of deformable convolutional network for defects with different scales, especially the abnormal Ratio of long and short edges. To solve the problem of a large number of small target defects, this paper uses the template image features as an auxiliary structure, and fuses the high-level and low-level features of the detected and template images to improve the detection rate of small target defects.In addition, the Diff Attention structure was also proposed in this paper.Based on the differential features of the target image and template image, the model's Attention was more focused on the region where the defect was located, which could strengthen the feature extraction ability of the model for the defect.DPDAN's good generalization is verified on both its own data set and public data set.

Keywords: Computer Vision, Deep Learning, Texture Representation

1. INTRODUCTION

Fabric defect detection is an important quality control process in textile industry. It is beneficial for workers to monitor fabric quality if the defect location can be quickly located in the production process. The existing fabric defect detection is completed by manual detection, not only the efficiency is low, but also the accuracy and recall rate is not high, easy to miss and error detection.

Traditional fabric defect detection methods are mainly divided into three categories: Based on the structure of the method through the analysis of the placement of the cloth grain to infer the overall composition of fabric texture[1] . Based on gray co-occurrence matrix, histogram statistics and mathematical morphology and other statistical methods of woven fabric defect detection method[2].

Based on Fourier transform method, wavelet transform method[3], Gabor filter[4] and other spectrum flaw detection method, As well as the flaw detection method based on traditional machine learning[5].

The rapid development of deep learning computer vision also provides a more powerful processing model for printing defect detection. Target detection models can be classified into two-stage model, one-stage model, target detection model without atnchor and target detection model based on Transformer: Two-stage model[6] is to borrow from some specific methods to generate suspected contain the target candidate box, and then through the detection head to analyze these candidate box positioning and identified the goal; One-stage model[7] is one step in place, after the input picture through model conversion directly output target location and category information; Anchor detection model[8] no longer presets anchor points, but instead generates category heat maps and corner

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

offsets at each pixel of the feature map; Based on the transformer[9] target detection model[10] by the supervision methods use transformer to replace convolution neural network learning to more stable characteristics .

Change Detection is generally applied in multi-temporal remote sensing data and has important applications in urban planning, forest protection, environmental protection and so on. Change detection is to analyze images in two different time dimensions and find out the changes. Deep learning-based change modeling models are mainly divided into three categories: Supervised learning change detection[11] according to the change area looking for change region and difference image characteristics of contact; However, remote sensing change detection is marked as pixel level professional tag cost is too high, unsupervised learning change detection[12][13] by inhibiting irrelevant variables inherent means looking for difference figure characteristics, such as data distribution to change the regional discriminant ability training; Migration study change detection[14] is to use a small amount of markup and training after migration to the unsupervised learning model, Then the classification accuracy and convergence speed of the model are improved.

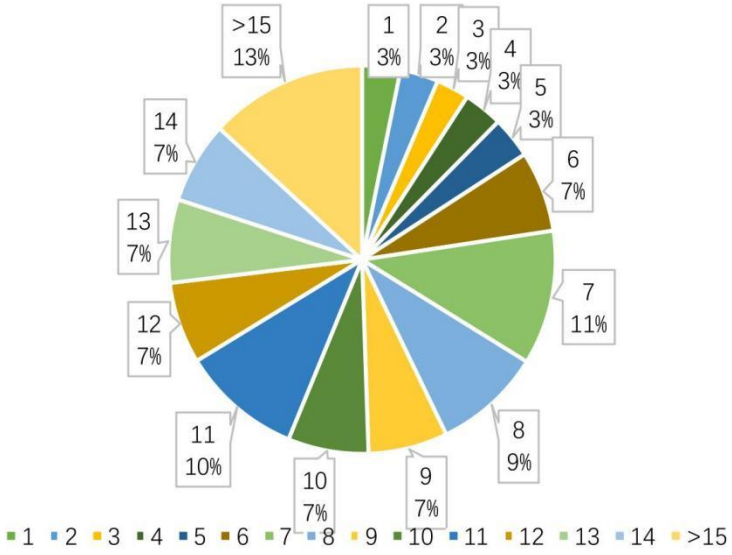
In general target detection models[15], the target to be detected often has richer location and semantic information than the background information, and the aspect ratio of the target to be detected is always close to 1. However, printed fabrics have the following characteristics:First of all, the patterns of printed fabrics are diverse, and some defects can hardly be distinguished even by human eyes under the cover of patterns. In general target detection model, printing pattern is easily regarded as target recognition, resulting in high false positive rate.Secondly, the geometric deformation of printed fabric[16] was very different, not only the different kinds of defects were very different, but also the same kind of defects had a huge difference in shape characteristics.Moreover, a large number of printed fabric defects are too small in area, occupying less than 3\% of the whole fabric image area, which is easy to produce missed detection under the interference of pattern.

This paper mainly focuses on the defects detection of printed fabrics, analyzes the existing problems and difficulties, puts forward some solutions based on the characteristics of data, and carries out experiments to verify the effect of the model[17]. In general, the main contributions of this paper are as follows:Firstly, by introducing the change detection template matching idea into the printing defect detection, the twin network can learn the difference between the target image and the template image, and reduce the noise effect caused by the pattern texture in the image to be detected.Secondly, for the problem of large differences in the geometric scale of the defect, the Ratio DCN was proposed, which enabled the model to have better modeling ability for the changes in the geometric scale of the defect, especially for the defect with obvious differences in the length and length.Thirdly, the defect problem, for a small target in the image to be detected and the template image characteristics between the figure characteristics of composite structures, the features of the network to extract figure has both the images and template images to be detected high-level features and shallow, and shallow features abundant space position information of small targets in defect detection plays an important role.Thirdly, diff Attention was proposed, which made the model focus on the defect area with the Diff feature graph as the query guide and improved the feature extraction ability of the model to the defect.Fourthly, the DPDAN (introduced at Section 2) defect detection model is proposed, which makes the detection accuracy of printed fabric defects significantly improved compared with the general target detection scheme and change detection scheme, and achieves the best experimental results on both the own data set and the public data set.

Figure 1 presents the statistical results of the length-width ratio of fabric defects in the FDUPFD dataset. Clearly visible from the figure, the abscissa represents different intervals of the defect length-width ratio, covering a wide range from extremely small ratios to ratios greater than 15. The ordinate indicates the number or proportion of defects in each interval (determined according to the actual markings on the chart). Through this figure, it can be intuitively observed that the length-width ratios of different types of defects vary significantly. This distribution characteristic is quite different from common target detection datasets, highlighting the uniqueness of printed fabric defects in terms of geometric scale and providing important data basis for the subsequent research on targeted detection methods.

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



[Statistics on the ratio of defect length to width]

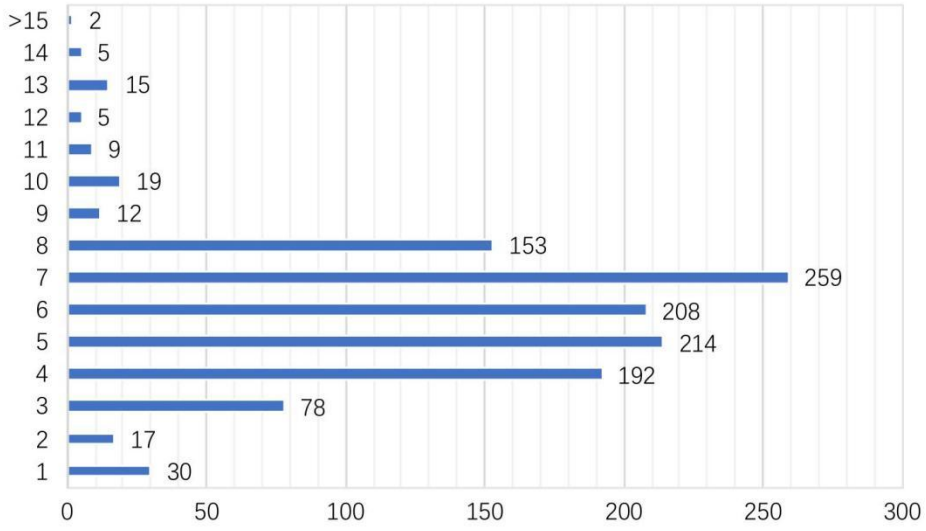


Figure 1 Length-width ratio statistics

Table 1. Category and number of defects

f	s	t	q	w	d	t
1218	1562	1105	834	571	662	176
b	w	t	z	h	p	s
178	164	223	181	234	80	109

2. DPDAN MODEL

FL loss function[18] is also used in DPDAN model to solve the problem of the imbalance of the number of different defects in FDUPFD dataset, and the output subnetwork of RetinaNet is used in DPDAN model, and the mask prediction subnetwork is added after the box regression and classification subnetwork to predict the result of

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

instance segmentation. DPDAN model mainly includes dual input module, Diff Attention module and output subnetwork module. The structure of DPDAN model is shown as figure 2

The dual-channel input module is mainly used to construct the backbone network structure of DPDAN, which consists of an input structure with dual-channel weight sharing, three sub-modules of Ratio-DCN and multi-scale feature fusion.The structural idea of dual-path weight shared input comes from change detection, and the difference between the image to be detected and the template image is learned by using the twin network, that is, the defect area.

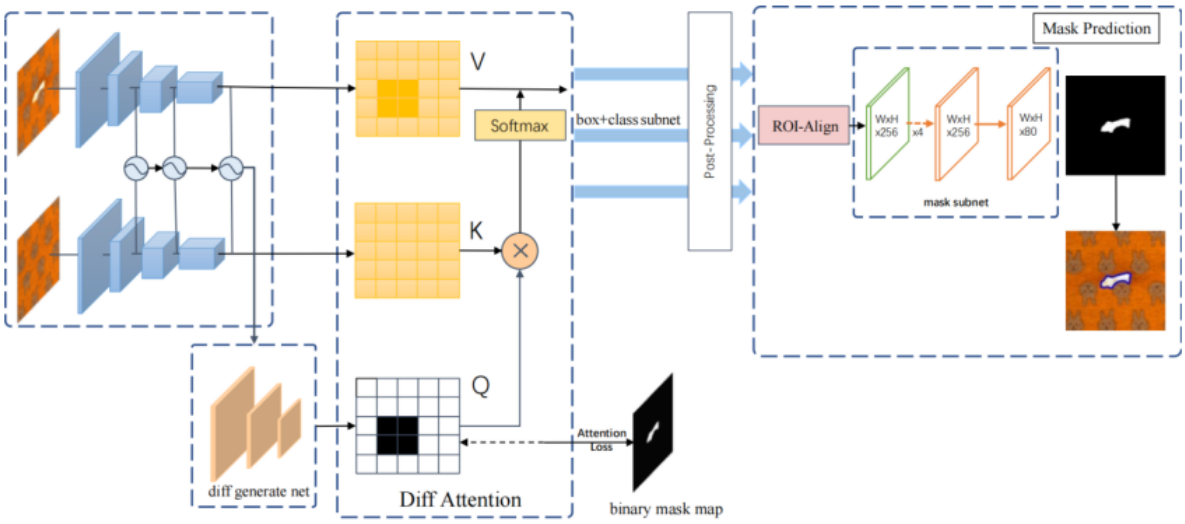


Figure 2 Schematic diagram of DPDAN model structure

Ratio-DCN By learning an extra offset, the variable convolutional network (DCN) enables CNN to adaptively learn the information of object position change[19][20], thus greatly improving the modeling ability of the model for geometric deformation. In order to better adapt to the geometric scale change of printed textile defects, we propose a Ratio DCN, in which the offset learned in DCN is multiplied by a Ratio function $g(\delta) = 2 * Sigmoid(\delta)$, so that the offset is larger in the direction of large scale and smaller in the direction of small scale. $\delta$is given in the formula 1:

$$\delta = \begin{cases} \frac{\bar{x}}{\bar{y}} - 1, & \text{In the x direction} \\ \frac{\bar{y}}{\bar{x}} - 1, & \text{In the y direction} \end{cases} \tag{1}$$

where $\bar{x}$ represents the size of the defect in the direction of x, and h represents the size of the defect in the direction of y. When the defect scale ratio is approximately equal to 1, the result of δ is approximately equal to 0, and the value of the ratio function $g(\delta)$is approximately equal to 1, which does not affect the learned offset of the variability convolution.The implementation form of Ratio-DCN is similar to ordinary two-dimensional convolution, and it can be applied to most backbone networks and has the advantages of high flexibility in plug - out and play We replace the second convolution layer of ResNet50 with Ratio DCN, and the input and output size, receptive field, step size and zero complement remain unchanged with the original ResNet50.

Feature fusion structure Multi-scale feature fusion[21] has been studied and used in many task scenarios, such as FPN[22], PANet[23], NAS-FPN[24] and BiFPN[25], these feature fusion methods are based on the same input in different stages of fusion, so that the final output contains feature information of different stages.

Different from the above common single-channel network structure Feature fusion[26], this part of the structure is in the form of Composite Feature (CF), that is, the template image features are used as the auxiliary structure to compound the template image features and the image features to be detected. The final output not only integrates the high-level and low-level features of the backbone network to be detected and the template, but also introduces the difference and similarity information between the image to be detected and the template image.

As shown in figure 3, three different characteristic composite structures are designed.

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

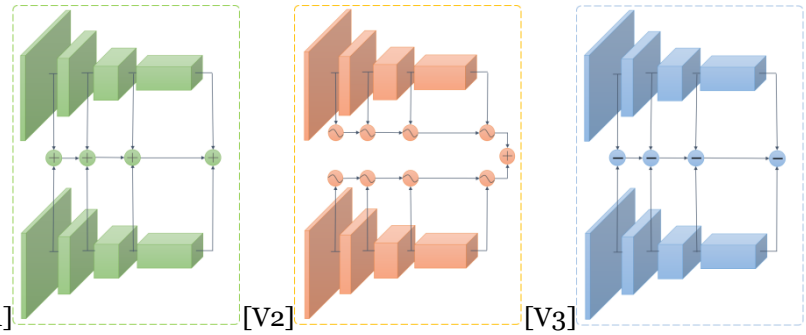


Figure 3 Three different multi-scale cross fusion structures

Suppose the basic backbone network of the image to be detected and the template image are $B_1$and B_2respectively. Each backbone network contains Lstages ($L = 4$), and each stage contains several convolution layers and batch normalization layers. The backbone network of the first L-th stage is realized by a nonlinear transformation function $F^L(\cdot)$.

In a typical convolutional neural network, each structure contains only one path of the backbone network, and the output of the previous L-th stage $L-1$-th stage is the input, which can be expressed by the following formula: $x^l = F^l(x^{l-1}), l \geq 2$.

The CF method combines features in the two-path input structure[27], iteratively intersects and fuses the corresponding features of each backbone network stage, and propagates them forward. ${B_1}_L^x$and ${B_2}_L^x$represent the output of two backbones network in L-th stage, c^Lrepresent the output of feature cross fusion in L-th stage, $g(\cdot)$represent smooth operation, ${B_1}_L^x$and ${B_2}_L^x$represent the output of two backbones network in L-th stage. Consists of a 1×1convolution layer. So for V1-V3 structure, its combination can be expressed by the following formula.

in V1:

$$x_{cat}^l = concat(x_{B_1}^l, x_{B_2}^l) \quad (1)$$

$$c^l = c^{l-1} + g(x_{cat}^l) \quad (2)$$

in V2:

$$x_{B_1}^l = F^l(x_{B_1}^{l-1} + g(x_{B_1}^{l-1})) \quad (1)$$

$$x_{B_2}^l = F^l(x_{B_2}^{l-1} + g(x_{B_2}^{l-1})) \quad (2)$$

$$c^l = concat(x_{B_1}^l, x_{B_2}^l) \quad (3)$$

in V3:

$$x_{diff}^l = abs(x_{B_1}^l - x_{B_2}^l) \quad (1)$$

$$c^l = c^{l-1} + g(x_{diff}^l) \quad (2)$$

Composite Feature In deep neural network, the spatial resolution of input features becomes smaller after being down-sampled by multiple layers of the network[28], while the high-level semantic information and contextual information become richer and richer. In terms of implementation, the final output of V1, V2, and V3 structures will be slightly different. V1 and V2 feature fusion output $c$fusion has the feature of each stage, so the composite feature can be taken as the final output, as defined $x_{out} = c$; However, V3 composite features only contain the differential features of the image to be detected and the template image, lacking the features of the global context of the original image, so it is necessary to add the features of the image to be detected before the composite operation, as defined $x_{out} = x + c$,where x_{out}represents the final output feature graph, x represents the feature graph to be detected before the composite operation, and $c$represents the composite feature.

Loss Function The loss function of DPDAN model is divided into several parts: classification loss, regression loss, mask loss and attention loss, as shown in the formula:

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$Loss = L_{cls} + L_{reg} + L_{mask} + L_{att},$$

where L_{CLS} is the IOU loss L_{CLS} , $L_{att}$for attention Loss, Smooth L1 Loss is used. We add multiple loss functions directly because we pay the same attention to these results and want them to be balanced. Direct addition also makes training easier and more intuitive.

Diff Attention

The Diff Attention module consists of two parts, one is the Diff feature graph generation network[29], the other is Diff Attention. The Diff feature graph network is used to generate queries in the attention mechanism, and the key values are acted by the template image feature graph and the image to be detected feature graph respectively.

Diff feature graph generation network The structure of Diff generate net is similar to the pyramid network of FPN, which is a bottom-up multi-layer network structure. In order to better learn the characteristics of Diff, we added a Attention Loss to modify the learned parameters. Attention Loss adopts Smooth L1 Loss, which can be expressed as:

$Att\_Loss(x,t,y) = \frac{1}{n}\sum_{i=1}^{n} 0.5 * (g(x_i,t_i) - y_i)^2$ if $\backslash absg(x_i,t_i) - y_i) < 1$ and

$Att\_Loss(x,t,y) = \frac{1}{n}\sum_{i=1}^{n} \backslash abs\, g(x_i,t_i) - y_i) - 0.5$ otherwise,

where $g(x,t)$represents the Diff feature graph generation network, x,t and y represent the image input to be detected, template image input and label respectively, and n represents the number of images.

Diff Attention Mechanism Combined with the characteristics of template images in the data, images to be detected and template images are used as input to generate Diff feature graph[30]. On this basis, the Diff Attention mechanism is proposed, with Diff feature graph as query, template image as key and images to be detected as value, to calculate Attention distribution.The weight calculated by the query will be largely concentrated in the area where the defect is located, while ignoring the information of other areas such as background, which can greatly improve the effectiveness and differentiation of attention calculation and thus improve the effect of the model.Diff Attention consists of dual Attention, which calculates the weight of Attention in spatial resolution and channel respectively. Spatial attention focuses on strengthening the global location information, while channel attention mainly focuses on strengthening the semantic information between channels. After obtaining these two outputs, they are superimposed and fused to further obtain the characteristics of global dependencies and enhance the understanding ability of the model to semantic information.

3. EXPERIMENT

3.1 FDUPFD dataset

All data sources are obtained from the factory's production lines and are repeatedly marked by textile technicians to ensure reliability.

In the textile factory, the fabric can be simply divided into two kinds, one is plain fabric, the color is relatively single, the background is relatively simple, only the texture, the other is printed fabric, the background often has a lot of patterns, pattern pattern form, very prominent. As shown in figure 4, it is the comparison between plain fabric and printed fabric.

FDUPFD data set is composed of printed fabrics, which are collected by production workers in a unified standard in several workshop production lines. There are 123 patterns in total, and the texture of different patterns is very different, as shown in Figure 4(b).

The data collected in the project are recognized and cut out images of fixed size by the machine, and then classified and marked by professional and technical personnel and summarized and uploaded to the system data management source. After marking, refer to the annotation format of Microsoft COCO data set. Programs were written to retrieve the corresponding image and category information from the database to generate annotation

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

files, and the image data and annotation files were downloaded and consolidated. Finally, FDUPFD (FDU Printing Fabric Defect) data set was made.

There are 7297 images in this dataset, and each image is 400 pixels wide and high, divided into 14 categories:(1) seam head (2) plug net (3) escape flower (4) before and after color difference (5) false color (6) interference at the bottom (7) desoiling (8) white (9) stains (10) drag color (11) dirty (12) flower paste (13) hole (14) water stains.

Figure 4 Plain and Printed fabrics Examples(a) lines are plain fabrics and (b) lines are printed fabrics

Table 1 shows the data distribution. After statistical analysis of defect area, defect maximum aspect ratio, defect minimum aspect ratio, inter-class scale size and intra-class scale size of the data set, it can be found that there are several prominent features in the data set. The feature information of each feature is analyzed in detail below.

Printed fabric defects: Floral fabric defects differed greatly in geometric scale, which was manifested in two aspects: one was the extreme length to width ratio of different types of defects, with the maximum length to width ratio reaching 20:1; the other was the large difference between different types of defects, and the same type of defects were also different. The defects with the aspect ratio greater than 7:1 accounted for about 75% of the total, while the defects with equal or little difference in length and width accounted for only a small part, which was quite different from the common target detection data set.

3.2 Experimental Results

Table 2. Comparison of exprimental results of target detection model at FDUPFD dataset

model	AP@50-95(%)	AP@50(%)	AP@75(%)	FPS
Faster-RCNN[6]	23.60	41.62	25.28	5.2
Cascade-RCNN[20]	25.39	43.48	26.57	6.5
YOLOv3[21]	19.29	38.35	21.13	19.5
CornerNet[8]	18.44	36.84	20.47	16.8
FCOS[22]	23.82	42.73	22.35	17.3
Reppoint[23]	23.05	44.62	23.26	18.6
RetinaNet[24]	19.43	39.79	20.5	18.2
DPDAN(ours)	**43.82**	**66.28**	**49.53**	13.4

In order to illustrate the effectiveness of DPDAN model, the design experiment is compared with the current general model, including two-stage, single-stage and no anchor point target detection model, the experimental results are shown in Table 2.It can be seen from the table that in the printed fabric data set, the DPDAN model proposed in this paper has the best detection performance. This is due to the template matching twin network structure based on change detection, which greatly reduces the influence of pattern texture of printed fabric data background. Combined with the Ratio-DCN, composite features and Diff Attention module, the model detection effect has been greatly improved. In terms of reasoning speed, it can be seen that the two-stage model algorithm is obviously slower than the single-stage algorithm, but in terms of accuracy, the two-stage detector is higher. The

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

DPDAN model was about 5 to 6 FPS slower than the single-stage model, but 7 to 8 FPS faster than the two-stage detector. In general, DPDAN model can greatly improve the detection accuracy of defects while maintaining good inference speed.Because DPDAN model adopts dual-path input structure, it uses the idea of template matching to detect the defects of images to be detected, while the general target detection algorithm does not input template information, which will appear unfair.In order to verify the effectiveness of the Ratio-DCN, combined feature and Diff Attention module in DPDAN, this paper compares the current commonly used change detection algorithms. Since the output of change detection algorithm is mostly panoramic semantic segmentation, while flaw detection only needs to pay attention to the change of local flaw area, without paying attention to the global change information, this paper tries its best to modify the output of the last layer while keeping the main body of the algorithm unchanged, so that it becomes the instance segmentation output.The experimental results are shown in table 3.

Table 3.Comparison of experimental results of change detection model

Model	AP@50 - 95(%)	AP@50(%)	AP@75(%)
CDNet [25]	33.49	60.59	39.29
DASNet[26]	32.80	61.40	41.75
FC-Siam-diff[27]	37.77	63.03	42.87
FCN-PP[28]	38.62	63.34	44.31
FFAN	38.45	64.06	44.92
DPDAN(ours)	**43.82**	**66.28**	**49.53**

As shown in 3, it can be seen that DPDAN model has better effect than the current change detection algorithm with better performance. This is because the general change detection method usually only focuses on change and unchanged change, which is a binary learning task. It pays more attention to global change and does not pay too much attention to the geometric shape and size of defects. Compared with general change detection methods, DPDAN model proposed improved methods such as Ratio DCN, compound feature and Diff Attention, which could make DPDAN pay more Attention to the changes of local defects rather than detect the global changes.Deep learning-based defect detection method for woven fabrics, currently the effective method models are yFDD-yOLO , NSRA . The YFDD-YOLO method is based on the improved YOLOV2 target detection model to optimize the selection of hyperparameters. NSRA is a non-local sparse representation method, which includes image preprocessing, restoration and threshold operation. In this paper, the basic model is reproduced by referring to relevant codes and combining these papers, and experiments are carried out on the FDUPFD data set. The comparison results are shown in Table 4.

Table 4. Comparison of experimental results of woven fabric defect detection methods

modle	Acc(%)
YFDD - YOLO[29]	73.32
NSRA[30]	80.14
DPDAN(ours)	**85.54**

To further test the generalization performance of the model, a comparative experiment is carried out on the public data set of tianchi fabric defects. Tianchi fabric defect detection competition is divided into several stages. The fabric in the preliminary stage is pure color fabric, and the fabric in the semi-final stage is printed fabric. Each stage has its corresponding highest result. The data set of Tianchi Fabric defect Detection Competition is similar to our data set, and also provides the images to be tested and the corresponding template images. In order to fit our application scenario, we use the printed fabric data set of the second round to conduct experiments and compare

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

with the results of the first place in the second round. The mAP of DPDAN is 58.08, while the mAP of the rank first model is 56.74.As can be seen from the table, the result of our method is 1.34% higher than that of the first place in the second round, indicating that our model has good generalization.

3.3 Ablation Study

Table 5. Results of main module ablation experiment

2*model	innovation point				2*mAP(%)
Baseline	Dual - Path	Composite	Ratio - DCN	DA	37.39
5*DPDAN	Yes				60.48
5*DPDAN	Yes		Yes		62.87
5*DPDAN	Yes			Yes	64.32
5*DPDAN	Yes	Yes			63.30
5*DPDAN	Yes	Yes	Yes		64.48
5*DPDAN	Yes	Yes	Yes	Yes	66.28

In this experiment, the performance of DPDAN model's main modules was compared by ablation experiment, and the experimental effect of each algorithm module was quantitatively analyzed. The experimental results are shown in table 5.Among them, dual-path is dual-input network structure, Composite is feature Composite, DA is Diff Attention, Baseline is a Baseline model excluding the above improvement points, which is composed of a ResNet50 and a detection head. It can be seen from the table that the model detection accuracy can be greatly improved by adding dual-path input network structure to the Baseline, which indicates that the introduction of template matching can reduce the influence of background pattern texture by referring to the idea of change detection algorithm. On the basis of dual-channel input network structure, the Ratio DCN, composite feature and Diff Attention have been improved by 2.39% and 2.82% respectively, among which Diff Attention has the biggest improvement, reaching 3.84%. Adding Composite and Ratio-DCN to Dual-Path can improve the detection accuracy by 4 %. Finally, the final detection model of these three factors is improved to 5.8%.The model performance of each improved point was further compared, and a comparative experiment was conducted on each improved point. In the input structure, the difference between single input and double input was compared. In the composite characteristics, the differences of V1, V2 and V3 are compared. In the

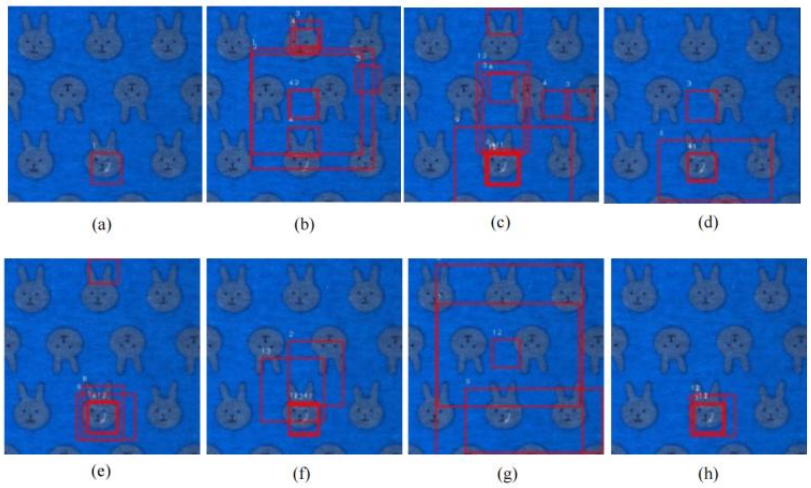


Figure 5 Prediction results of each target detection model

Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

anamorphic convolution, the difference between DCN and Ratio-DCN is compared. The experimental results are shown in table 6.Figure 5 shows the prediction results of each target detection model for the same image. (a) is DPDAN proposed by us, (b) is ConerNet, (c) is YOLOv3, (d) is ftF-RCNN, (e) is Cascade-RCNN, (f) for FCOS, (g) for RetinaNet, (h) for Reppoint. It can be seen that in the absence of template matching, each target detection model will yield many wrong prediction results. However, DPDAN weakens numerous pattern texture interference in the target image due to template matching, so that the model can focus on the area where the defects are located.

Table 6. Improved point control experiment

2*model	input structure		composite character			Deformable Convolutional Networks		2*mAP(%)
7*DPDAN	x Yes	x+t Yes	V1	V2	V3	DCN	Ratio-DCN	 39.79 **60.48**
		Yes Yes Yes	Yes	Yes	Yes			61.24 61.77 **63.3**
		Yes Yes				Yes	Yes	61.32 **62.87**

Figure 6 shows the DPDAN model mask prediction results. Each picture is composed of real label results and prediction results. The mask formed by real labels is in the top line, and the mask formed by prediction results is in the bottom line. Compared with the real label result, the predicted mask result could identify the pattern or misidentify the defect around the defect, making the predicted mask range larger, but it was still able to distinguish the defect instance from the background.

Figure 6 The results with mask

Declaration of Conflicting Interests

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

Data Sharing Agreement

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

Funding

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

REFERENCES

[1] Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine computation of complex Fourier series. Math. Comp., 19, 297-301.
[2] Haykin, S. (2002). Adaptive Filter Theory (4th ed.). Prentice Hall. (Information and System Science series)
[3] Morgan, D. R. (2005, Aug.). Dos and Don'ts of Technical Writing. IEEE Potentials, 24(3), 22-25.
[4] Bertalmio, M., Sapiro, G., Caselles, V., & Ballester, C. (2000). Image inpainting. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (pp. 417--424).
[5] Barnes, C., Shechtman, E., Finkelstein, A., & Goldman, D. B. (2009). PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Trans. Graph., 28(3), 24.
[6] Levin, A., Zomet, A., & Weiss, Y. (2003). Learning How to Inpaint from Global Image Statistics. In ICCV (Vol. 1, pp. 305--312).
[7] Li, J., He, F., Zhang, L., Du, B., & Tao, D. (2019). Progressive reconstruction of visual structure for image inpainting. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 5962--5971).
[8] Li, J., Wang, N., Zhang, L., Du, B., & Tao, D. (2020). Recurrent feature reasoning for image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7760--7768).
[9] Nazeri, K., Ng, E., Joseph, T., Qureshi, F. Z., & Ebrahimi, M. (2019). Edgeconnect: Generative image inpainting with adversarial edge learning. arXiv preprint arXiv:1901.00212.
[10] Guo, X., Yang, H., & Huang, D. (2021). Image Inpainting via Conditional Texture and Structure Dual Generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 14134--14143).
[11] Wang, T., Ouyang, H., & Chen, Q. (2021). Image Inpainting with External-internal Learning and Monochromic Bottleneck. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5120--5129).
[12] Ballester, C., Bertalmio, M., Caselles, V., Sapiro, G., & Verdera, J. (2001). Filling-in by joint interpolation of vector fields and gray levels. IEEE transactions on image processing, 10(8), 1200--1211.
[13] Esedoglu, S., & Shen, J. (2002). Digital inpainting based on the Mumford--Shah--Euler image model. European Journal of Applied Mathematics, 13(4), 353--370.
[14] Darabi, S., Shechtman, E., Barnes, C., Goldman, D. B., & Sen, P. (2012). Image melding: Combining inconsistent images using patch-based synthesis. ACM Transactions on graphics (TOG), 31(4), 1--10.
[15] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
[16] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2536--2544).
[17] Ren, Y., Yu, X., Zhang, R., Li, T. H., Liu, S., & Li, G. (2019). Structureflow: Image inpainting via structure-aware appearance flow. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 181--190).
[18] Yan, Z., Li, X., Li, M., Zuo, W., & Shan, S. (2018). Shift-net: Image inpainting via deep feature rearrangement. In Proceedings of the European conference on computer vision (ECCV) (pp. 1--17).
[19] Wang, Y., Tao, X., Qi, X., Shen, X., & Jia, J. (2018). Image Inpainting via Generative Multi-column Convolutional Neural Networks. In NeurIPS.
[20] Liu, H., Jiang, B., Song, Y., Huang, W., & Yang, C. (2020). Rethinking image inpainting via a mutual encoder-decoder with feature equalizations. In Computer Vision--ECCV 2020: 16th European Conference, Glasgow, UK, August 23--28, 2020, Proceedings, Part II 16 (pp. 725--741). Springer.

Research Article

[21] Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., & Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 85--100).
[22] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234--241). Springer.
[23] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6), 1452--1464.
[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998--6008).
[25] Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., & Vedaldi, A. (2014). Describing Textures in the Wild. In Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).
[26] Ngan, H. Y. T., Pang, G. K. H., & Yung, N. H. C. (2011). Automated fabric defect detection—a review. Image and vision computing, 29(7), 442--458.
[27] Chan, C.-h., & Pang, G. K. H. (2000). Fabric defect detection by Fourier analysis. IEEE transactions on Industry Applications, 36(5), 1267--1276.
[28] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 580--587). (Colombia, USA: IEEE)
[29] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 779--788). (Las Vegas, USA: IEEE)
[30] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International Journal of Computer Vision, 88(2), 303--338.