



Automatic fabric defect detection using a wide-and-light network

Jun Wu^{1,3} · Juan Le^{1,3} · Zhitao Xiao^{2,3} · Fang Zhang^{2,3} · Lei Geng^{2,3} · Yanbei Liu^{2,3} · Wen Wang^{2,3}

Accepted: 16 November 2020 / Published online: 6 January 2021
© Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Automatic fabric defect detection systems improve the quality of textile production across the industry. To make these automatic systems accessible to smaller businesses, one potential solution is to use limited memory capacity chips that can be used with hardware platforms with limited resources. That is to say, the fabric defect detection algorithm must ensure high detection accuracy while maintaining a low computational cost. Therefore, we propose a wide-and-light network structure based on Faster R-CNN for detecting common fabric defects. We enhance the feature extraction capability of the feature extraction network by designing a dilated convolution module. In a dilated convolution module, a multi-scale convolution kernel is used to adapt to defects of different sizes. Dilated convolutions can increase receptive fields without increasing the number of parameters used. Therefore, we replace a subset of ordinary convolutions with dilated convolutions to learn target features and use convolution kernel decomposition and bottleneck methods to simplify the feature extraction networks. Then, high-level semantic features are fused with bottom-level detail features (via skip-connection) to obtain multi-scale fusion features. Finally, a series of anchor frames (of different sizes) is designed to suit multi-scale fabric defect detection. Experiments show that compared with various mainstream target detection algorithms, our proposed algorithm can improve the accuracy of fabric defect detection and reduce the size of the model.

Keywords Fabric defect detection · Multi-scale · Dilated convolution · Feature extraction

1 Introduction

In the textile production process, quality control and inspection are essential, and fabric defect detection is the most essential part of the process. Experienced inspectors perform traditional fabric defect detection through visual observation; however, manual detection is not reliable. Situations such as work environment, labour intensity, the experience of inspectors and other factors can affect manual inspection [1]. We have determined, from textile industry information feedback, that the accuracy of manual detection is only 60–75% [2]. In recent years, machine vision-based

object detection methods have been widely used in the field of fabric defect detection [3, 4]. Many machine vision-based methods obtain higher accuracy than manual detection. As a result, using this technology for automatic fabric defect detection has become an inevitable trend in the textile industry.

In practical applications, fabric defect detection algorithms must not only ensure detection accuracy but also guarantee its applicability to hardware platforms with limited resources. Currently, the accuracy of existing detection models is low. This is due in large part to multi-scale defects in the fabric image; so, the fabric defect detection model must be able to meet multi-scale object detection. However, even the best model is still troubled by the large size of the problem. Therefore, we must consider ways to reduce the size of the model. Inspired by the successful use of deep convolutional neural networks (DCNN) for target detection [5, 34], we propose a wide-and-light network structure called WALNet. A WALNet network structure has the capability of standard feature extraction networks that is enhanced with a dilated convolution module. In the dilated convolution module, multi-scale convolution kernels are used to extract fabric defect features of different sizes.

✉ Zhitao Xiao
xiaoazhitao@tiangong.edu.cn

¹ School of Electronics and Information Engineering, Tiangong University, Tianjin, 300387, China

² School of Life Sciences, Tiangong University, Tianjin, 300387, China

³ Tianjin Key Laboratory of Optoelectronic Detection Technology and System, Tiangong University, Tianjin, 300387, China

We replace a subset of the ordinary convolutions used in multi-scale convolution kernels with dilated convolutions, which is beneficial to the detection of medium and large fabric defects. Convolution kernel decomposition and bottleneck methods are used to simplify the feature extraction components of the method, which reduces the number of parameters of the feature extraction network. However, the shallow features of the convolutional neural network contain an exhaustive amount of image detail, and the deep features include many abstractions. To make full use of the detailed information and abstract information of various feature levels, the shallow and deep features are fused; so that the model can extract features more comprehensively. To find defects of different sizes in a fabric image, a series of anchor frames of various sizes are designed in the network to adapt to multi-scale fabric defect detection and to improve detection accuracy.

The contributions of this article include the following three aspects:

- (1) Improve multi-scale defect detection in the fabric image using multi-scale convolution kernels, dilated convolutions and feature fusion;
- (2) Improve detection accuracy by obtaining comprehensive fabric features, and designing a series of candidate frames of different sizes;
- (3) Reduce the size of existing convolutional neural network models while maintaining or improving fabric defect detection by using convolution kernel decomposition and bottleneck methods to simplify feature extraction.

2 Related work

2.1 Fabric defect detection methods

Traditional automatic fabric defect detection methods [6–8] contain two parts: feature extraction and feature recognition. Feature extraction operations can extract the features that are most conducive to pattern classification based on the original pattern. This can significantly reduce the dimensionality of pattern sampling. Feature extraction is the preliminary work of feature recognition. The quality of fabric defect feature extraction directly affects the resulting of subsequent feature recognition. Effective feature extraction will improve the overall fidelity of automatic fabric defect detection. In the feature extraction phase, the most essential step is to design distinguishable defect features. At this stage, fabric defect detection methods using artificial design features can be roughly divided into four types: spectrum, statistical, structural, and modelling. Among these, the spectrum method is the

focus of research in this field. Commonly used spectrum methods include Fourier transforms, wavelet transforms, and Gabor transforms. Hanbay et al. [6] constructed a real-time fabric defect detection system based on Fourier transform, which can realize the detection of common circular knitting fabric defects. Zhu et al. [7] used wavelet decomposition methods to detect nonhomogeneous textured fabrics; wavelet decomposition meets the minimum accuracy requirements of automatic fabric detection. Liang et al. [8] proposed an automatic detection method for fabric defects based on lattice segmentation and Gabor filtering, which achieved a detection accuracy of 0.975 on the databases of star and box-pattern images. Statistics-based methods usually perform various feature statistics first, and then realize the detection of defects by distinguishing the difference between the defected area and the flawless area. Therefore, such methods cannot effectively detect small defects or defects similar to the background area. The structure-based method treats texture as a combination of texture primitives. Texture analysis is performed by acquiring texture features and inferring its replacement rules. However, the reliability of the structure method is very low, and the structured method can only remove fabric defects from unique textures. The model-based method performs defect detection by modelling the fabric texture and judging whether the test fabric image conforms to the model. However, model-based detection methods need to accurately estimate various parameters of the present model, which requires a large amount of calculation; and it is difficult to describe complex and changeable texture images with a unified model.

Although many traditional methods have obtained satisfactory results in detecting fabric defects with specific textures and patterns, in realistic environments, fabric characteristics are easily affected by light intensity, noise, shooting angle, and shooting distance. Therefore, designing and extracting fabric defect features that are more robust and easier to distinguish requires advanced methods. Deep learning is widely used in the field of object detection [9–12]. Compared with the method of manually designing features, deep learning methods can automatically extract image features; thus, the learned features have strong generalization and expression capabilities [35–37]. In recent years, CNN-based deep learning network structures, the mainstream of deep learning research, have proven that they are suitable for automatic fabric defect detection [13–16]. Li et al. [13] proposed an SDA framework based on the Fisher criterion, which realized the defect detection of periodic patterns and jacquard pattern fabrics. Mei et al. [14] proposed a fabric defect detection model, MSCADE, based on a multi-scale convolutional denoising autoencoder networks. MSCADE effectively detects various types of fabric defects. Jun et al. [15] used a two-step strategy for

fabric defect detection. The first step is to cut the original fabric image into squares of the same size, and then use the Inception-V1 model to classify the squares; the second step uses the LeNet-5 model to Identify the type of defect. Li et al. [16] proposed a compact convolutional neural network architecture to detect fabric defects and achieved high detection accuracy.

2.2 Target detection methods based on CNN

Currently, the two mainstream frameworks of CNN-based target detection technology are the two-stage methods, e.g., Region Convolutional Neural Network [25–27] (R-CNN), and the single-stage methods, e.g., Single-Shot Detector [28] (SSD) and You Only Look Once [29–31] (YOLO). Since the two-stage target detection framework has one more classification and regression than the single-stage target detection framework, the accuracy of detection is much higher.

This work is based on the two-stage target detection framework Faster R-CNN [27] network structure. In the target detection framework, the backbone network is located before the input layer, providing a path to the specific task layer. The backbone network is then used to extract different scale features of the target. Early CNN-based target detection models mostly used VGG-16 [17] and ResNet [18] as the backbone network. VGG-16 has 16 layers, and the extracted feature expression ability is limited. If the number of network layers is increased by superposition, the gradient will easily disappear or explode when it is transmitted in the network, which reduces the performance of the network. To solve the disappearance and explosion of deep network gradients, He et al. [18]

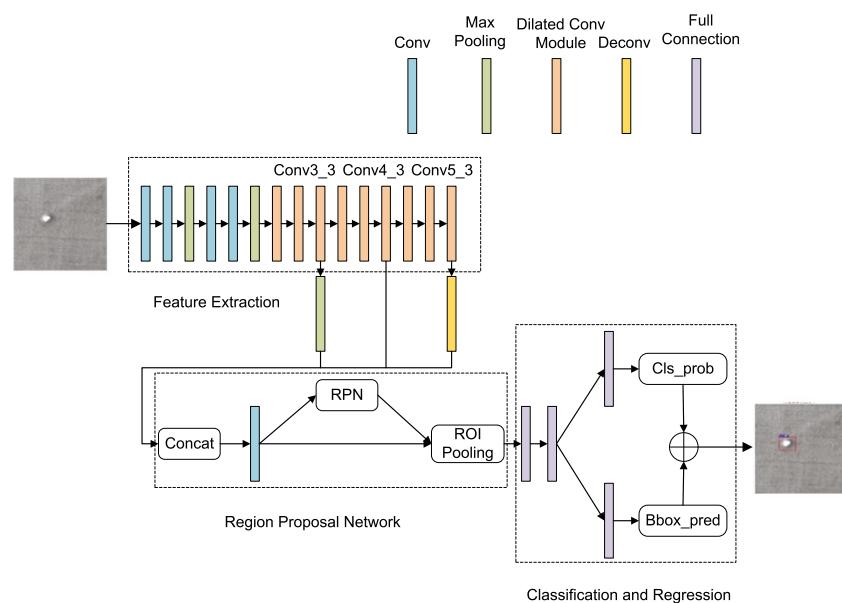
proposed a skip-connection residual network structure (ResNet). ResNet fuses the shallow feature information with subsequent layers to generate and transfer new features. This method effectively guarantees the transfer of characteristic information to the deep network and improves the overall performance of the network. Although using the residual structure can effectively improve the performance of the network, it doubles the number of parameters. However, fabric defect detection often needs to be performed in hardware platforms with limited resources, which limits the use of CNN models with larger sizes. Building a small and efficient CNN model, while maintaining satisfactory detection accuracy, is still an urgent problem in the field of fabric defect detection.

3 WALNet network architecture overview

3.1 WALNet network architecture

The WALNet network structure, shown in Fig. 1, includes three parts: feature extraction, region proposal network, and classification and regression. The feature extraction part can be divided into two stages: the shallow processing stage and the dilated convolution module processing stage. The shallow processing stage contains four convolutional layers and two pooling layers, which captures the texture and edge features of the fabric and reduces the feature map size. The processing stage of the dilated convolution module contains nine dilated convolution modules to obtain multi-scale features while reducing the number of parameters. To obtain more comprehensive and rich features, the three intermediate layers Conv3_3, Conv4_3, and Conv5_3 are

Fig. 1 WALNet network structure



used for feature fusion, and the fused features are further reduced by a 1×1 convolution, and the final feature is 512 dimensions. To adapt to fabric defects of different sizes, anchor frames with various scales are designed. Finally, the anchor frame generated by the region proposal network is sent to the classification and regression sections to correct the position of the anchor frame further. The algorithm flow chart of WALNet is shown as in Fig. 2.

In the entire WALNet network structure, the feature extraction network plays a vital role. Since, it maintains a small number of parameters, it not only increases the receptive field, but also obtains a more comprehensive fabric feature through the fusion of features at different scales and different levels. A series of candidate frames of different sizes are designed in the network to ensure the network can satisfy multi-scale fabric defect detection.

3.2 Dilated convolution module

Fabric defects have different sizes, which makes them difficult to detect. Latent features can be extracted for defect detection by increasing the number of convolution kernels, but this method will generate a large number of parameters. To ensure a small number of parameters and obtain multi-scale information, a dilated convolution module is designed in this paper. Four parallel paths are designed in the dilated convolution module, three of which are designed to extract features of different scales to detect fabric defects of different sizes, and the remaining one is to obtain non-linear features of defects of different positions. In addition,

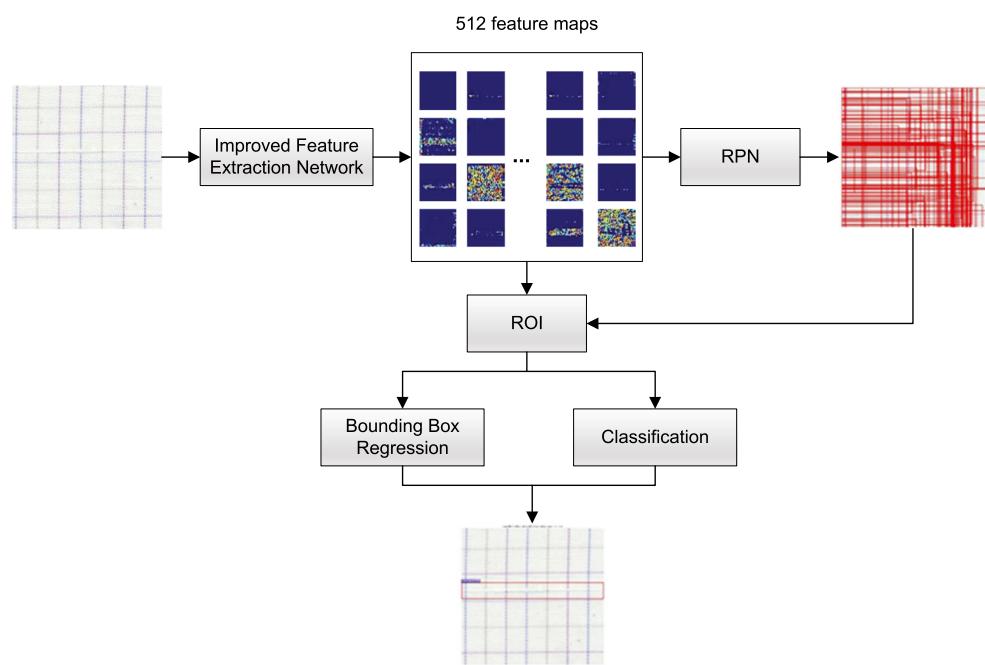
convolution kernel decomposition and bottleneck were used to keep the model size small.

3.2.1 Multi-scale convolution kernel

Multi-scale convolution kernel methods can effectively obtain multi-scale features [22, 23]. Considering that fabric defects are usually different sizes, three parallel convolution kernels are designed to extract fabric features. The convolution kernels can adapt to different size defects in the fabric image. The multi-scale convolution module is shown in Fig. 3. The outputs of the convolutional layers are connected to obtain the multi-scale features of the input feature map.

To further increase the receptive field, without increasing the number of parameters, a dilated convolution is used instead of ordinary convolution. The two images in Fig. 4 show the receptive fields of ordinary convolutions and dilated convolutions of the same size, respectively. (a) represents a 3×3 ordinary convolution and (b) represents a 3×3 dilated convolution. The dark blue squares represent the part of the image where parameters exist in the convolution kernel, and the light blue part squares represent the zeros in the convolution kernel. The figure shows that the number of parameters of both is 9, the receptive field of ordinary convolution is 3×3 , and the receptive field of dilated convolution is 5×5 . Dilated convolution makes the network increase the receptive field without increasing the number of parameters, so the ordinary convolutions in Path2 and Path3 are replaced with dilated convolutions.

Fig. 2 WALNet algorithm flow chart



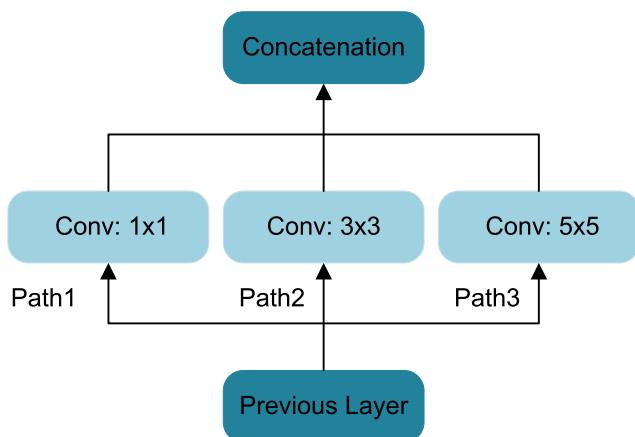


Fig. 3 Multi-scale convolution module

3.2.2 Convolution kernel decomposition

In the multi-scale convolution module, the number of parameters in each path can be calculated by (1)

$$P_i^l = W_i^l \cdot H_i^l \cdot X_i^{l-1} \cdot X_i^l + X_i^l, i = 1, 2, 3. \quad (1)$$

where W_i^l represents the width of the convolution kernel, H_i^l represents the height of the convolution kernel, X_i^{l-1} represents the number of output channels of the previous layer, and X_i^l represents the number of output channels. Note that the last X_i^l represents the number of offsets.

Since a 5×5 convolution kernel has 25 parameters, two consecutive 3×3 convolution kernels only need 18 parameters. Therefore, decomposing a 5×5 convolution kernel into two consecutive 3×3 convolution kernels can

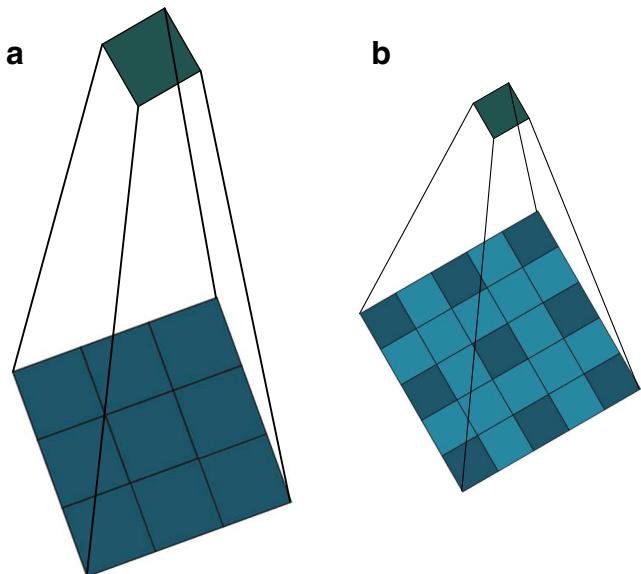


Fig. 4 Comparison of the receptive fields of (a) a 3×3 ordinary convolution and (b) a 3×3 dilated convolution with a dilation rate of 2

reduce the number of parameters by 28%. Also, since each convolution in the dilated convolution module is followed by ReLU [24], using two consecutive 3×3 convolutions can improve the ability of the dilated convolution module to represent nonlinearity. Therefore, the dilated convolution module is designed with two consecutive 3×3 convolutions instead of 5×5 convolutions. The module, after the convolution kernel decomposition, is shown in Fig. 5.

3.2.3 Bottleneck

Although the convolution kernel decomposition is used, the parameter number of the dilated convolution module is still large. To keep the number of parameter small, bottlenecking is used to reduce the number of feature maps on Path2 and Path3. The module behind bottleneck is shown in Fig. 6.

To extract the non-linear features of defects at different positions parallel to Path1, Path2, and Path3, an extra path (Path4) is added to the module. The dilated convolution module can extract features of different scales and positions and keep fewer parameters. Figure 7, shows the architecture of the dilated convolution module. Feature maps from the previous layer are input into four different parallel paths.

3.3 Feature fusion

In many recent deep learning tasks, feature fusion has a certain effect on performance improvement [21, 41]. Combining fine-grained details with highly abstracted information in the feature extraction layer can help subsequent region proposal network, and classification and regression networks detect objects of different sizes, which can further improve detection accuracy. As a result of the

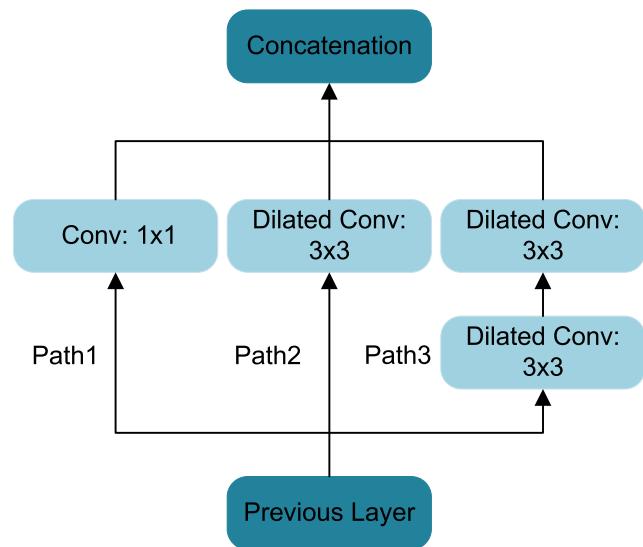


Fig. 5 Decomposed module of convolution kernel

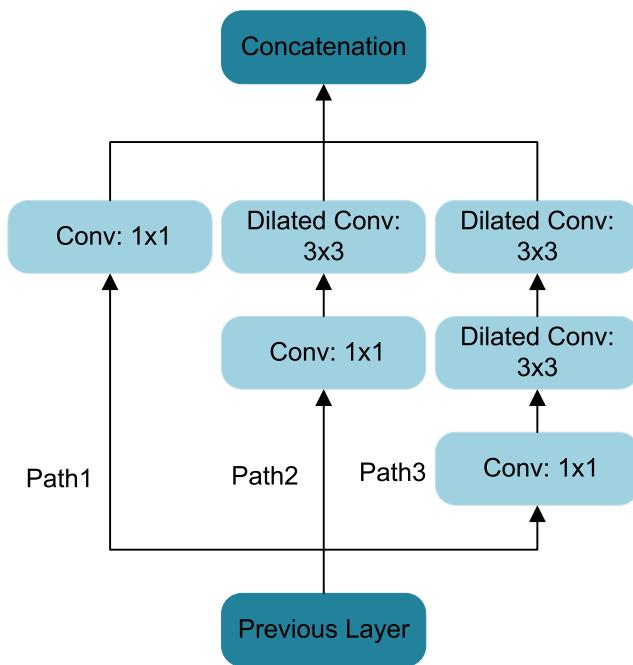


Fig. 6 Modules behind bottleneck

direct connection of all abstraction layers may produce redundant information and have more parameters, feature fusion is performed on Conv3_3, Conv4_3, and Conv5_3. Figure 8 shows the feature fusion module.

3.4 Anchor frames Design

The Faster R-CNN algorithm uses the anchor mechanism to generate anchor regions in the region proposal network.

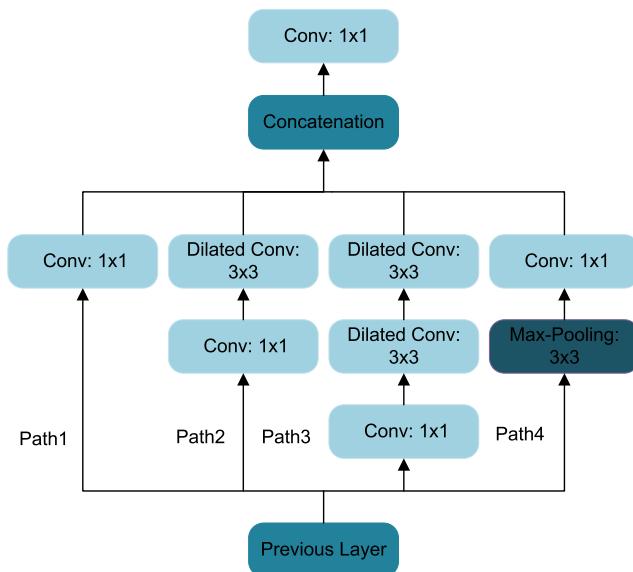


Fig. 7 Architecture of the dilated convolution module

The anchor mechanism includes 3 scales (8,16,32) and 3 aspect ratios (1:2,1:1,2:1). They can be combined to get nine anchor frames. In Faster R-CNN, the feature map obtained after the feature extraction network is 1/16 of the original image. Each pixel in the feature map corresponds to a 16×16 area of the original image. The area of the anchor frames can be calculated by (2), while the width and height of the anchor frames can be calculated by (3) and (4).

$$S_i = (\lambda_i \cdot B)^2, i = 1, 2, 3. \quad (2)$$

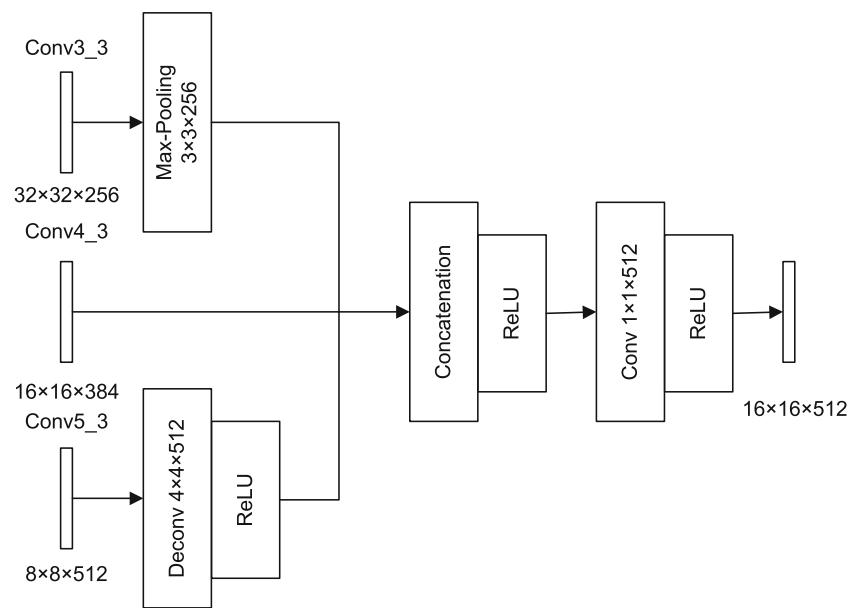
$$H_{i,j} = \sqrt{\frac{S_i}{R_j}}, i, j = 1, 2, 3. \quad (3)$$

$$W_{i,j} = \sqrt{S_i \cdot R_j}, i, j = 1, 2, 3. \quad (4)$$

where S_i represents the area of the anchor, λ_i represents the three scales of the anchor, B represents the size of each pixel on the feature map corresponding to the original image, R_j represents the three aspect ratios of the anchor, $H_{i,j}$ represents the height of the anchor and $W_{i,j}$ represents the width of the anchor.

In this work, we build a fabric database with different textures, including white grey fabrics from the TILDA fabric database, and dark red fabrics and grid fabrics created in a laboratory. Statistics on the size and aspect ratio of defects in the three fabric databases are shown in Fig. 9. Statistics show that the defects have statistical characteristics with significant differences in size and various morphological changes. To find fabric defects of different sizes, a series of anchor frames of different sizes are designed, so that the network can satisfy multi-scale fabric defect detection.

From the fabric defect size statistics shown in Fig. 9, the size of the fabric defect is usually between 14~255 pixels, and the aspect ratio is usually between 0.5~2. For this purpose, the scale of anchor is designed to be 4,8 and 16 and aspect ratio are designed to be 0.5,1 and 2. The area, width, and height of the candidate frames, before and after improvement, are calculated by (2), (3), and (4). Table 1 shows the area, width and height of the candidate frames before and after improvement. From the table, the minimum candidate frame before improvement is [91,181], the maximum candidate frame is [724,362]; the minimum candidate frame after improvement is [45,90], and the maximum candidate box is [362,181]. The size of the improved candidate frame is more suitable for fabric defect detection.

Fig. 8 Feature fusion module

4 Experimental results and analysis

4.1 Data set acquisition

We construct a fabric database with different textures to verify the accuracy of the algorithm. These include white grey fabrics from the TILDA fabric database and dark red fabrics and grid fabrics.

The TILDA fabric database is provided by the Computer Vision Group of the Computer Engineering Department of the University of Albert-Ludwig-Freiburg. TILDA is a

textile texture database with four types of fabrics, each type containing two representative textiles. Based on the analysis of the textile atlas, seven wrong categories and one correct category were defined. In the TILDA database, 360 defective images were selected. These images are divided into six categories (broken yarns, carryings, holes, fuzz balls, scratches and stains), each containing 60 samples. The size of the original image is 768×512 pixels. Figure 10 shows some typical fabric images.

Dark red fabrics and grid fabrics were collected using an industrial camera in a laboratory environment. Dark red

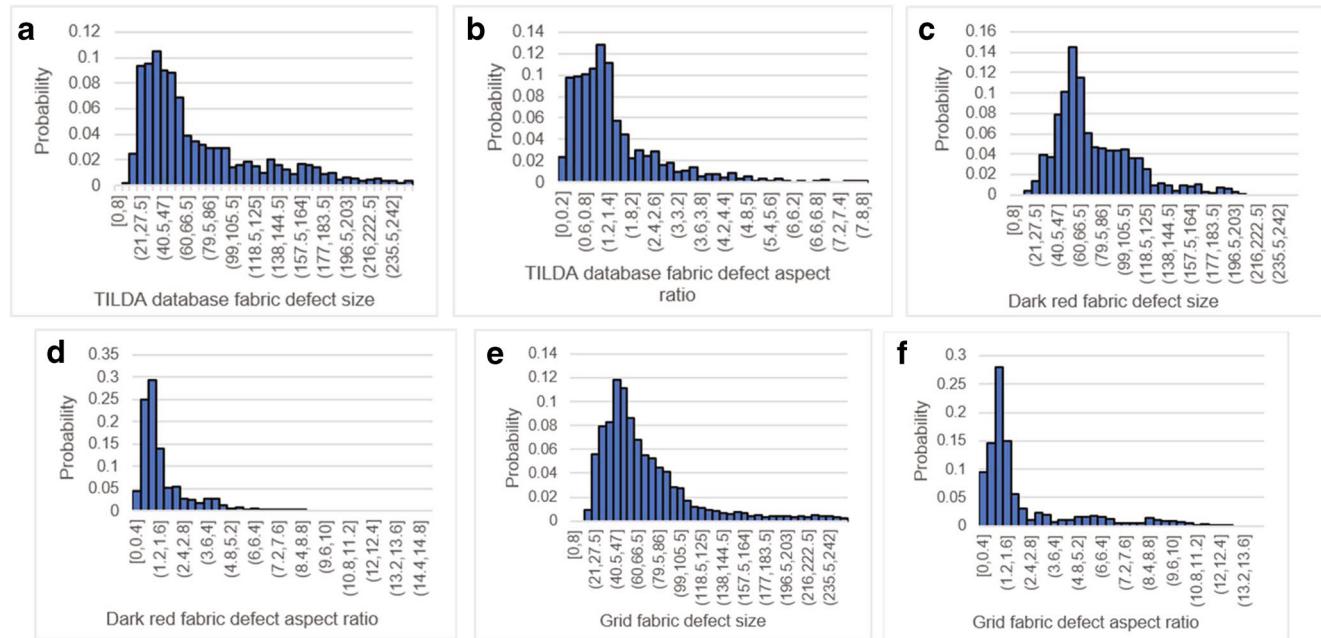
**Fig. 9** Statistics of fabric defect distribution

Table 1 Candidate box area, width and height

Anchor	Scale	Ratio	Area	W	H
No Improvement (base_size=16)	8	0.5		91	181
		1	128×128	128	128
		2		182	91
		0.5		181	362
	16	1	256×256	256	256
		2		362	181
		0.5		362	724
	32	1	512×512	512	512
		2		724	362
		0.5		45	90
	4	1	64×64	64	64
		2		90	45
		0.5		91	181
Improvement (base_size=16)	8	1	128×128	128	128
		2		181	91
		0.5		181	362
	16	1	256×256	256	256
		2		362	181

fabrics contain four defects such as carryings, broken yarns, knots, and fuzz balls. Grid fabrics contain four defects such as carryings, knots, holes and stains. Each type of defect in both fabrics contains 60 fabric images with an image size of 800×600. Figure 11 shows some of the fabric images used in the experiment.

Data augmentation is usually used to improve the quality of the data set, avoid overfitting of complex networks. In the experiment, the training data is augmented by random cropping and rotation, and a 256×256 image is randomly cropped from the experimental samples. Then, the cropped

image is rotated by 90° 180° and 270°. The final number of samples used for the experiment was 8400, of which the TILDA database fabric was 3600, and the dark red fabric and grid fabric were 2400. Each sample type is randomly selected to be 80% for the training set and 20% for the test set.

4.2 WALNet model size

Table 2 shows the parameter configuration of the dilated convolution module in the WALNet model. Of these,

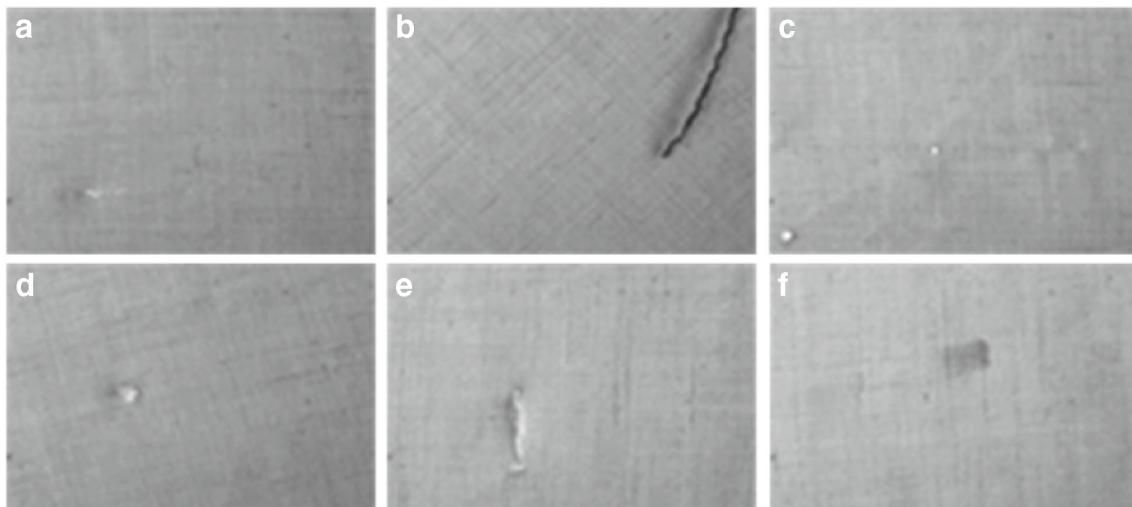


Fig. 10 Some typical fabric images in the TILDA database: **a** broken yarns, **b** carryings, **c** holes, **d** fuzz balls, **e** scratches, and **f** stains

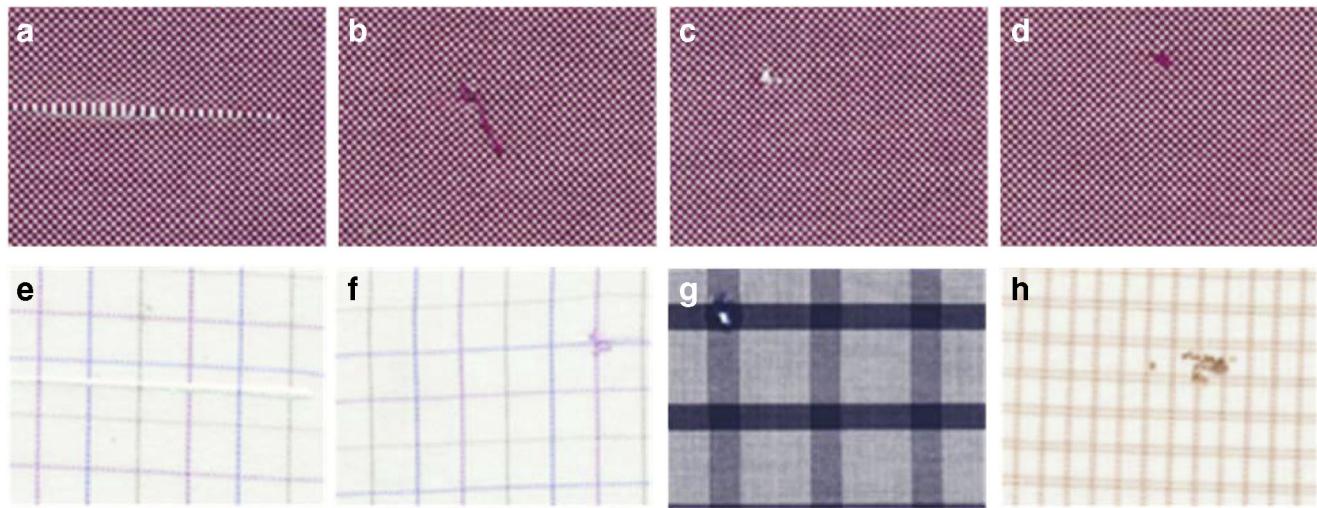


Fig. 11 Fabric image: The first row has a dark red fabric sample with **a** carryings, **b** broken yarns, **c** knots, and **d** fuzz balls; the second row has a grid fabric sample with **e** carryings, **f** knots, **g** holes, and **h** stains

DCM1~DCM9 represent nine dilated convolution modules used in the WALNet model, and Conv represents the number of 1×1 convolution kernel used in the last layer of the dilated convolution module.

Table 3 shows the specific architecture design of the feature extraction part in the WALNet model, where DCM is a dilated convolution module. Downscale indicates that Conv3_4 passes 3×3 max pooling with a step size of 2 and Upscale means that Conv5_3 passes 4×4 deconvolution with a step size of 2. Then, Downscale, Conv4_3, and Upscale are merged into Concat. The final output Convf is obtained after 1×1 convolution. A Dropout with a probability of 0.5 was used after Conv3_2, Conv4_2, and Conv5_2 to prevent overfitting.

4.3 Evaluation metrics

For each single class, the Precision(P) is the ability of a model to only identify the relevant objects. The recall(R)

is the ability of a model to find all of the relevant class (all ground truth bounding boxes). The P and R can be calculated by (5) and (6):

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

where True Positive(TP) predicts positive samples as positive samples; False Positive (FP) predicts negative samples as positive samples; False Negative (FN) predicts positive samples as negative samples.

After counting the number of instances in distinguished quality, a balanced metrics of Average Precision (AP) could be calculated and used for representing the average performance of detection. Because AP is different among every category, the mean AP of all categories could be used as the overall performance of detectors and it is named mAP.

Table 2 Parameter configuration of dilated convolution module in WALNet model

Module Name	Path1	Path2	Path3	Path4	Conv
	1×1	$1 \times 1 \ 3 \times 3$	$1 \times 1 \ 3 \times 3 \ 3 \times 3$	Pooling 1×1	1×1
DCM1	64	48 96	16 32 32	64 64	256
DCM2	64	48 96	16 32 32	64 64	256
DCM3	64	48 96	16 32 32	64 64	256
DCM4	96	72 144	24 48 48	96 96	384
DCM5	96	72 144	24 48 48	96 96	384
DCM6	96	72 144	24 48 48	96 96	384
DCM7	128	96 192	32 64 64	128 128	512
DCM8	128	96 192	32 64 64	128 128	512
DCM9	128	96 192	32 64 64	128 128	512

Table 3 Architecture design of feature extraction part in WALNet model

Name	Type	Stride	Output size	Parameters
Conv1_1	3×3 Conv	1	$256 \times 256 \times 64$	1.7k
Conv1_2	3×3 Conv	1	$256 \times 256 \times 64$	36k
Pooling1	3×3 Max-Pooling	2	$128 \times 128 \times 64$	
Conv2_1	3×3 Conv	1	$128 \times 128 \times 128$	73k
Conv2_2	3×3 Conv	1	$128 \times 128 \times 128$	147k
Pooling2	3×3 Max-Pooling	2	$64 \times 64 \times 128$	
Conv3_1	DCM	2	$32 \times 32 \times 256$	141k
Conv3_2	DCM	1	$32 \times 32 \times 256$	79k
Conv3_3	DCM	1	$32 \times 32 \times 256$	158k
Conv4_1	DCM	2	$16 \times 16 \times 384$	331k
Conv4_2	DCM	1	$16 \times 16 \times 384$	178k
Conv4_3	DCM	1	$16 \times 16 \times 384$	355k
Conv5_1	DCM	2	$8 \times 8 \times 512$	599k
Conv5_2	DCM	1	$8 \times 8 \times 512$	316k
Conv5_3	DCM	1	$8 \times 8 \times 512$	632k
Downscale	3×3 Max-Pooling	2	$16 \times 16 \times 256$	
Upscale	4×4 Deconv	2	$16 \times 16 \times 512$	8.2k
Concat	Concatenation		$16 \times 16 \times 1152$	
Convf	1×1 Conv	1	$16 \times 16 \times 512$	590k
Total				3644k

The AP and mAP can be calculated by (7), (8) and (9):

$$AP = \sum_n (R_{n+1} - R_n) P_{interp}(R_{n+1}), \quad (7)$$

$$P_{interp}(R_{n+1}) = \max_{\tilde{R}: \tilde{R} \geq R_{n+1}} P(\tilde{R}), \quad (8)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i, \quad (9)$$

where AP_i being the AP in the i th class and N is the total number of class being evaluated .

4.4 Experiment

Six experiments were performed on the resulting three fabric databases. The first experiment constructs a dilated convolution module using different expansion rates. The second experiment verifies the effectiveness of feature fusion. The third experiment verifies whether the improved anchor frame can improve model performance. The fourth experiment verifies the generalization ability of the model. The fifth experiment is to compare the proposed algorithm with traditional algorithms. The sixth experiment is to compare the proposed algorithm with various mainstream target detection algorithms.

Although the fabric data set has been expanded, it does not meet the requirements of large data volume DCNN. For

tasks with fewer samples, training a DCNN from scratch can easily lead to model overfitting, so this is not usually done. A common alternative is to pre-train on a large data set, and then use the weight of the trained DCNN as a fixed feature extractor for related tasks. This method is called transfer learning. We pre-train the model on VOC2007, and then set the trained parameters to the initialization parameters of the target task. The processor is Intel (R) Xeon (R) CPU E5-1620 v4 @ 3.50GHz 3.50 GHz, RAM 32.0GB, and the type of graphics card is NVIDIA GeForce GTX 1080Ti.

4.4.1 Experiment I

Dilated convolutions with different expansion rates correspond to receptive fields of different sizes. Therefore, the construction of a dilated convolution module (using dilated convolutions with different rates of expansion) will affect the detection results. For the white grey fabrics in the TILDA fabric database, the dilated convolution building modules with different expansion rates are used for training and the accuracy of the model is tested. Table 4 shows the experimental results.

Table 4 shows that as the expansion rate of the dilated convolution increases, mAP increases initially and then decreases, indicating that it is necessary to select a dilated convolution with a suitable expansion rate to build a dilated convolution module.

Table 4 Testing accuracy of dilated convolution module using dilated convolution with different expansion rates on TILDA database

Name	DCM1	DCM2	DCM3	DCM4	DCM5	DCM6	DCM7	DCM8	DCM9	mAP
midrule WALNet1	1	1	1	1	1	1	1	1	1	0.968
WALNet2	1	2	2	1	2	2	1	2	2	0.978
WALNet3	1	2	3	1	2	3	1	2	3	0.983
WALNet4	1	2	4	1	2	4	1	2	4	0.986
WALNet5	1	3	3	1	3	3	1	3	3	0.994
WALNet6	1	2	5	1	2	5	1	2	5	0.988
WALNet7	1	3	4	1	3	4	1	3	4	0.984
WALNet8	1	3	5	1	3	5	1	3	5	0.982
WALNet9	1	5	5	1	5	5	1	5	5	0.979

For the dark red fabric and the grid fabric created in the laboratory, the setting of the expansion rate of the dilated convolution in the dilated convolution module is the same as that in Table 4. The dark red fabric and the grid fabric are trained separately and then tested to find the accuracy of the model. Figure 12 shows the experimental results.

Figure 12 shows that for dark red fabrics and grid fabrics, as the expansion rate of dilated convolution increases, mAP also increases first and then decreases, which also shows that the choice of expansion rate of the dilated convolution is important.

If the expansion rate of the dilated convolution is too small, the corresponding receptive field will also be small, and the network can only learn the information of local smaller targets, but not the global information of larger targets. If the expansion rate of the dilated convolution is too large, it may contain too much redundant information, and some objects will be directly ignored and become part of the background, which will reduce the detection accuracy. Therefore, choosing an appropriate expansion rate is a key step of the algorithm. The experimental results show that

the mAP value is the highest when the expansion rate of the dilated convolution in the dilated convolution module is set to be consistent with WALNet5, so the expansion rate setting of the dilated convolution in the final model is consistent with WALNet5.

4.4.2 Experiment II

Feature fusion, at different levels, is an effective way to improve detection performance. To verify the effectiveness of the feature fusion method used in this paper, the white grey fabric in the TILDA fabric database and the dark red fabric and grid fabric created in the laboratory were tested by the algorithm. Table 5 shows the experimental results.

Low-level features have higher resolution and contain more position and detail information, but because they have fewer convolutional layers, they have lower semantics and more noise. High-level features have stronger semantic information, but the resolution is very low, and the ability to perceive detail is poor. Therefore, effectively integrating the two is the key to improve the model. Table 5 shows that for white grey fabrics in the TILDA fabric database, the mAP value increases by 1.2% after integrating features of different levels; for the dark red fabric created in the laboratory, the mAP value increases by 1% after integrating features of different levels; for the grid fabric created in the laboratory, the mAP value increases by 1.1% after integrating features of different levels. Experiments show that our feature fusion method used can improve the performance of the model.

4.4.3 Experiment III

To verify whether the improved candidate frame can improve model performance, the white grey fabric in the TILDA fabric database and the dark red fabric and grid

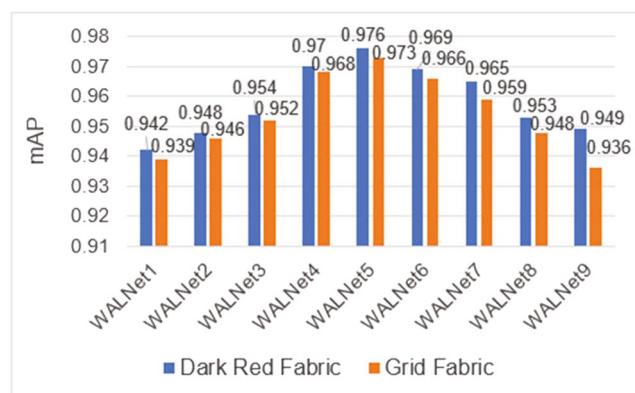


Fig. 12 Test Accuracy of dilated convolution module on dark red fabric and grid fabric using dilated convolution with different expansions

Table 5 Detection accuracy of model without and with feature fusion

Fabric Type	TILDA Fabric	Dark Red Fabric	Grid Fabric
No Feature Fusion			
mAP	0.982	0.966	0.962
Feature Fusion			
mAP	0.994	0.976	0.973

fabric created in the laboratory were trained and tested with the modified candidate frame before and after the improvement. Table 6 shows the experimental results.

Table 6 demonstrates that compared with the network without improved candidate frames, the accuracy of the model obtained by the network training with improved candidate frames is higher. For the white grey fabrics in the TILDA fabric database, the mAP value increased by 0.5%; for the dark red fabrics created in the laboratory, mAP values increased by 0.4%; for the grid fabrics created in the laboratory, mAP values increased by 0.5%. Experiments demonstrate that the candidate frame designed in this work is useful for improving the performance of the model.

4.4.4 Experiment IV

To verify the generalization ability of the WALNet model, a 10-fold cross-validation was performed in the experiment. In the experiment, each fabric database was divided into 10 copies. For each round of cross-validation, 1 copy was used as the test set, and the remaining 9 copies were used as Training set. Table 7 shows the test results.

The average mAP on the TILDA database is 0.995, and the average mAP of the dark red fabric and grid fabric are 0.976 and 0.974, respectively. Experimental results show that the WALNet network has good generalization ability.

4.4.5 Experiment V

To verify the effectiveness of the model proposed in this paper, we compare with the existing commonly used traditional algorithms. The detection result of the white grey

fabric in the TILDA fabric database, as shown in Fig. 13. (a) is the original fabric image to be inspected, (b) is the defect detection result obtained by the method in this paper, (c) is the defect detection result based on the method proposed by Zhu et al. [38], (d) is the defect detection result based on BVM and low-rank decomposition [39], (e) is the defect detection result based on HOG and low-rank decomposition [40]. It can be seen from the results that the method in this paper uses a red border to locate the defect area, which can accurately locate the defect. Other methods show the detection result by generating a binary image, and the white pixels are the defect areas. Although there is a small amount of noise in the detection image, the separation of the blemish and the background can be achieved. For dark red fabrics and grid fabrics, the comparison methods are still the above. The experimental results are shown in Figs. 14 and 15. The method in this paper can still accurately detect and locate defects. Other methods can still detect individual defects, and most of the defects cannot be separated from the background. The above experimental results prove that the method proposed in this paper can effectively locate the defect area, has a better detection effect and has strong adaptability to different fabric types.

4.4.6 Experiment VI

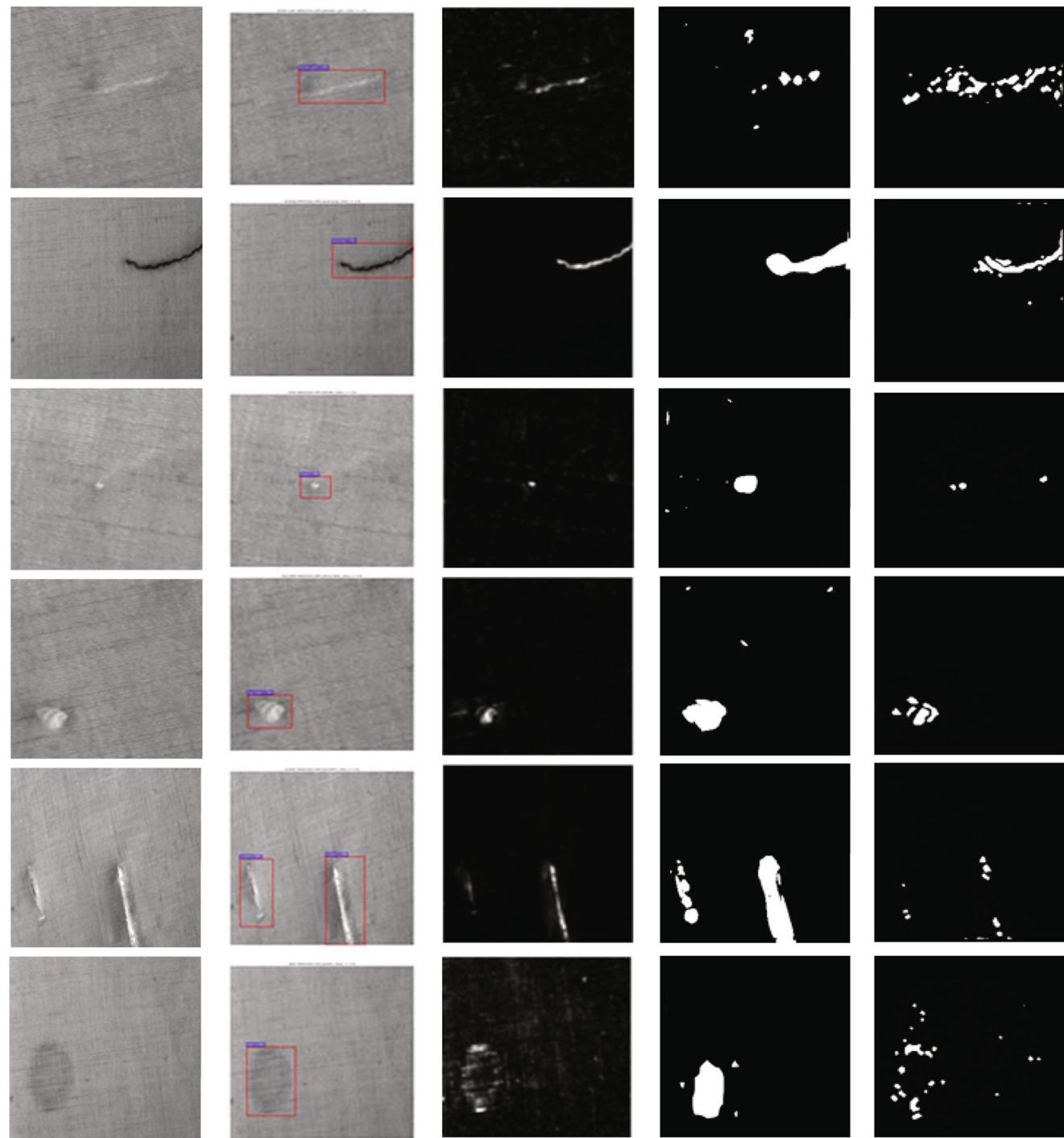
To evaluate the performance of our model, we compare it to various mainstream target detection networks. For white grey fabrics in the TILDA fabric database, the experimental results are shown in Table 8. For dark red fabrics, the experimental results are shown in Table 9. For grid fabrics, the experimental results are shown in Table 10.

Table 6 Detection accuracy of candidate box before and after improvement

Fabric Type	TILDA Fabric	Dark Red Fabric	Grid Fabric
No Improvement			
mAP	0.989	0.972	0.968
Improvement			
mAP	0.994	0.976	0.973

Table 7 Detection accuracy during 10-fold cross-validation

Fabric type	TILDA Fabric	Dark Red Fabric	Grid Fabric
Avg-mAP	0.995	0.976	0.974

**Fig. 13** Defect detection results of TILDA fabric database: **a** original images, **b** Ours, **c** Zhu et al. [38], **d** Li et al. [39], **w** Li et al. [40]

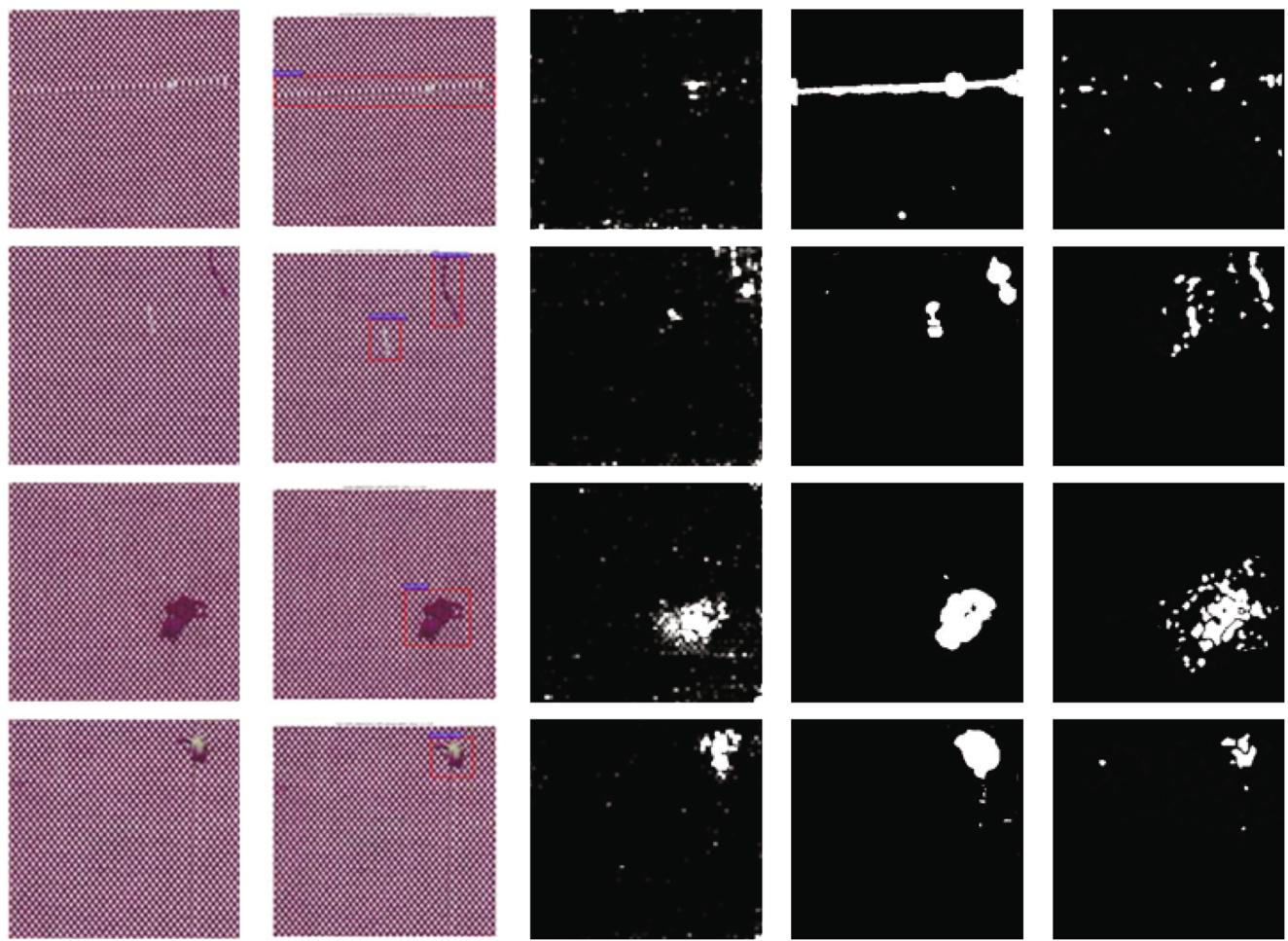


Fig. 14 Defect detection results of dark red fabrics: **a** original images, **b** Ours, **c** Zhu et al. [38], **d** Li et al. [39], **e** Li et al. [40]

Tables 8, 9 and 10 shows that the WALNet model obtains the highest mAP on all three data sets. The WALNet model also has the least number of parameters and has better performance than most current mainstream target detection models. We can draw the following conclusions from Table 8–10:

- (1) The WALNet model uses multi-scale convolution kernels and skip-connection to obtain multi-scale fusion features with rich structural levels. The rich and comprehensive feature effectively improves recognition performance. R-FCN, Faster R-CNN + R-101-FPN, Faster R-CNN + MobileNet, Faster R-CNN + ShuffleNet, Faster R-CNN + IGCV3, Mask R-CNN and RepPoints do not consider multi-scale feature extraction methods, which is why WALNet is superior to tested mainstream methods.
- (2) While improving the feature extraction capability of the feature network, WALNet uses convolution kernel

decomposition and bottlenecking to reduce the number of parameters, which can help the network spread feature information. Mask R-CNN, RepPoints and TridentNet all use various methods to improve the performance of the network, but they do not consider reducing the parameters of the model. Although Faster R-CNN + MobileNet, Faster R-CNN + ShuffleNet, Faster R-CNN + IGCV3 have fewer model parameters, the features extracted by the network are not sufficient.

- (3) WALNet sets the scale ratio and area of the anchor frame according to a specific fabric data set, which helps to improve the recognition performance. Although TridentNet considers multi-scale feature extraction, it does not change the scale ratio and area of the anchor frame. R-FCN, Faster R-CNN + MobileNet, Faster R-CNN + ShuffleNet, Faster R-CNN + IGCV3, Mask R-CNN and RepPoints are universal network, which means the previously set anchor frame may not be suitable for specific data sets.

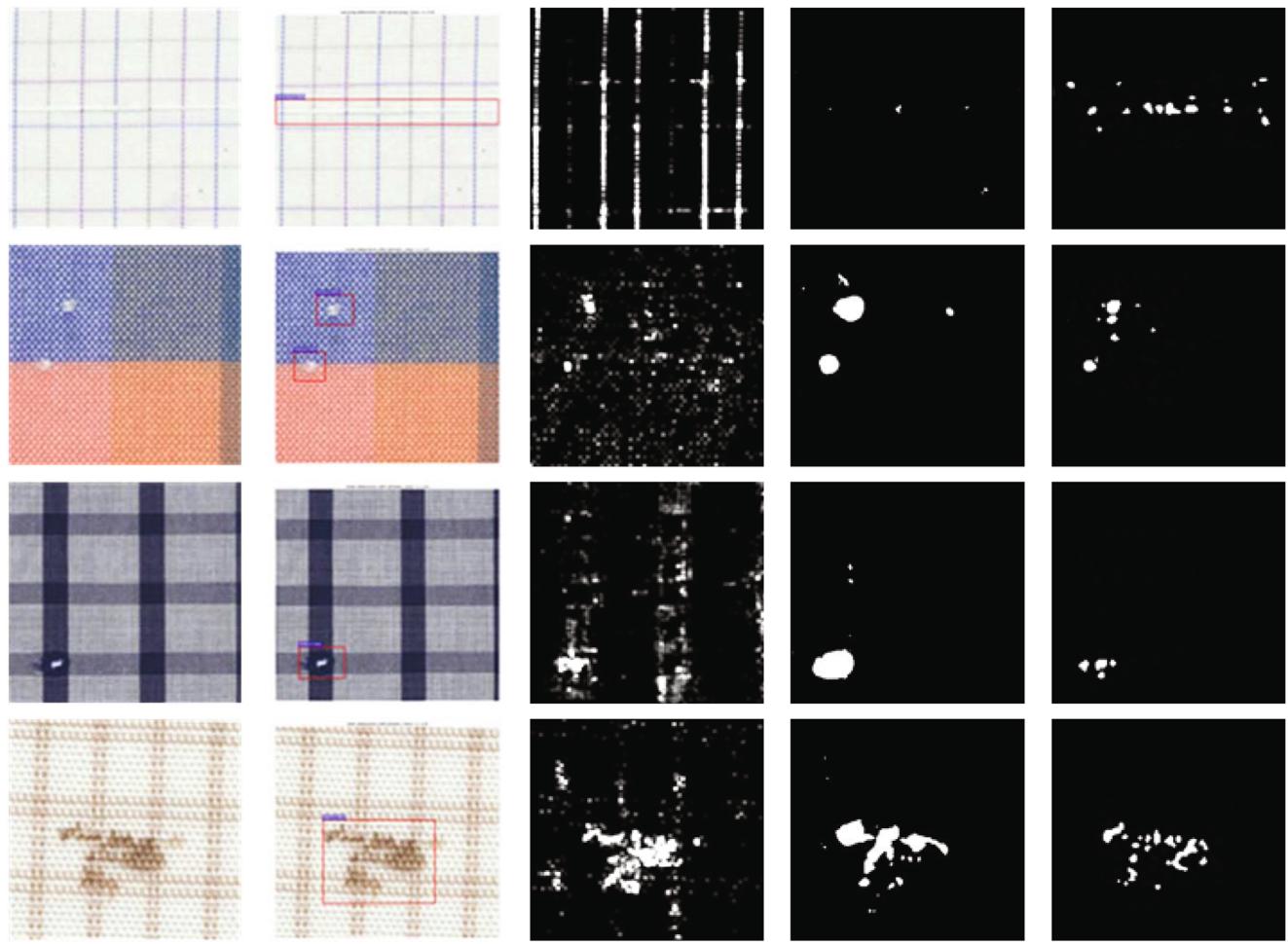


Fig. 15 Defect detection results of grid fabrics: **a** original images, **b** Ours, **c** Zhu et al. [38], **d** Li et al. [39], **e** Li et al. [40]

5 Conclusion

We propose a wide-and-light network for fabric defect detection that can deal with different types of fabrics. The model first uses a dilated convolution module to enhance the feature extraction capability of the feature extraction

network. The dilated convolution module uses a dilated convolution to learn fabric defect features and uses multi-scale features to adapt to defects of different sizes in the fabric image. The main components of the module use multi-scale convolution kernels, convolution kernel decomposition and a bottleneck cut-off. The results show

Table 8 Comparison of mAP by different algorithms on TILDA fabric database

Models	Backbone network	Parameters(MB)	mAP
R-FCN [32]	R-101	-	0.969
Faster R-CNN	R-101-FPN	60.52	0.979
Faster R-CNN	MobileNet	54.45	0.974
Faster R-CNN	ShuffleNet	53.72	0.972
Faster R-CNN	IGCV3	53.72	0.975
Mask R-CNN [33]	R-101-FPN	62.81	0.980
RepPoints [20]	R-101	55.62	0.989
TridentNet [19]	R-101	85.82	0.992
WALNet	-	53.89	0.994

Table 9 Comparison of mAP by different algorithms on dark red fabrics

Models	Backbone network	Parameters(MB)	mAP
R-FCN	R-101	-	0.960
Faster R-CNN	R-101-FPN	60.52	0.972
Faster R-CNN	MobileNet	54.45	0.966
Faster R-CNN	ShuffleNet	53.72	0.963
Faster R-CNN	IGCV3	53.72	0.967
Mask R-CNN	R-101-FPN	62.81	0.969
RepPoints	R-101	55.62	0.972
TridentNet	R-101	85.82	0.976
WALNet	-	53.89	0.976

Table 10 Comparison of mAP by different algorithms on grid fabrics

Models	Backbone network	Parameters(MB)	mAP
R-FCN	R-101	-	0.955
Faster R-CNN	R-101-FPN	60.52	0.965
Faster R-CNN	MobileNet	54.45	0.962
Faster R-CNN	ShuffleNet	53.72	0.960
Faster R-CNN	IGCV3	53.72	0.963
Mask R-CNN	R-101-FPN	62.81	0.966
RepPoints	R-101	55.62	0.970
TridentNet	R-101	85.82	0.972
WALNet	-	53.89	0.973

that fewer parameters extract the effective features of the input image. Then multi-scale fusion features, with rich structural levels, are obtained through skip connections. Finally, a series of candidate frames, with different sizes, are adopted to adapt to multi-scale fabric defect detection. Experimental results show that on white gray fabrics in the TILDA fabric database, dark red fabrics and grid fabric created in the laboratory, the model proposed in this work has achieved an accuracy rate of more than 97%. Our method accurately detecting common defects in the fabric, such as broken yarns, carryings, holes, fuzz balls, knots, scratches and stains. Additionally, the size of the model is reduced while obtaining accurate detection results.

References

- Korman S, Reichman D, Tsur G, Avidan S (2013) Fast-match: Fast affine template matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 2331–2338
- Jia L, Chen C, Liang J, Hou Z (2017) Fabric defect inspection based on lattice segmentation and gabor filtering. Neurocomputing 238:84–102
- Yapi D, Allili MS, Baaziz N (2017) Automatic fabric defect detection using learning-based local textural distributions in the contourlet domain. IEEE Trans Autom Sci Eng 15(3):1014–1026
- Huangpeng Q, Zhang H, Zeng X, Huang W (2018) Automatic visual defect detection using texture prior and low-rank representation. IEEE Access 6:37965–37976
- Nasrabadi NM (2019) Deeptarget: An automatic target recognition using deep convolutional neural networks. IEEE Trans Aerosp Electron Syst 55(6):2687–2697
- Hanbay K, Talu MF, Ozguven OF (2017) Real time fabric defect detection by using fourier transform. Journal of the Faculty of Engineering and Architecture of Gazi University 32(1):151–158
- Zhu B, Liu J, Pan R, Gao W, Liu J (2015) Seam detection of inhomogeneously textured fabrics based on wavelet transform. Textile Research Journal 85(13):1381–1393
- Jia L, Chen C, Liang J, Hou Z (2017) Fabric defect inspection based on lattice segmentation and gabor filtering. Neurocomputing 238:84–102
- Lin CH, Wang SH, Lin CJ (2019) Using convolutional neural networks for character verification on integrated circuit components of printed circuit boards. Applied Intelligence 49(11):4022–4032
- Lu M, Hu Y, Lu X (2020) Driver action recognition using deformable and dilated faster r-cnn with optimized region proposals. Applied Intelligence 50(4):1100–1111
- Zhang YY, Feng Y, Liu Dj, Shang Jx, Qiang Bh (2020) Frwcae: joint faster-rcnn and wasserstein convolutional auto-encoder for instance retrieval. Applied Intelligence pp 1–14
- Li Y, Zhang Y, Huang X, Yuille AL (2018) Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing 146:182–196
- Li Y, Zhao W, Pan J (2017) Deformable patterned fabric defect detection with fisher criterion-based deep learning. IEEE Transactions on Automation Science and Engineering 14(2):1256–1264
- Mei S, Wang Y, Wen G (2018) Automatic fabric defect detection with a multi-scale convolutional denoising autoencoder network model. Sensors 18(4):1064
- Gao, XJWZMP (2020) Fabric defect detection based on a deep convolutional neural network using a two-stage strategy. Textile Research Journal
- Li Y, Zhang D, Lee D (2019) Automatic fabric defect detection with a wide-and-compact network. Neurocomputing 329:329–338
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409-1556
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
- Li Y, Chen Y, Wang N, Zhang Z (2019) Scale-aware trident networks for object detection. arXiv: Computer Vision and Pattern Recognition
- Yang Z, Liu S, Hu H, Wang L, Lin S (2019) Repoints: Point set representation for object detection pp 9657–9666
- Zhang Z, Tang Z, Zhang Z, Wang Y, Qin J, Wang M (2019) Fully-convolutional intensive feature flow neural network for text recognition. arXiv: Computer Vision and Pattern Recognition
- Geng H, Liang Y, Liu Y, Alsaadi FE (2018) Bias estimation for asynchronous multi-rate multi-sensor fusion with unknown inputs. Information Fusion 39:139–153
- Geng H, Wang Z, Liang Y, Cheng Y, Alsaadi FE (2017) Tobit kalman filter with time-correlated multiplicative sensor noises under redundant channel transmission. IEEE Sensors Journal 17(24):8367–8377
- Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 807–814
- Girshick R, Donahue J, Darrell T, Malik J (2015) Region-based convolutional networks for accurate object detection and segmentation. IEEE transactions on pattern analysis and machine intelligence 38(1):142–158
- Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 1440–1448
- Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp 91–99
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: Single shot multibox detector. In: European conference on computer vision, Springer, pp 21–37
- 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
- Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7263–7271
- Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. arXiv:180402767

32. Dai J, Li Y, He K, Sun J (2016) R-fcn: Object detection via region-based fully convolutional networks. arXiv: Computer Vision and Pattern Recognition
33. He K, Gkioxari G, Dollar P, Girshick R (2020) Mask r-cnn. IEEE Transactions on Pattern Analysis and Machine Intelligence 42(2):386–397
34. Weimer D, Scholzreiter B, Shpitalni M (2016) Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. Cirp Annals-manufacturing Technology 65(1):417–420
35. Wang T, Chen Y, Qiao M, Snoussi H (2018) A fast and robust convolutional neural network-based defect detection model in product quality control. The International Journal of Advanced Manufacturing Technology 94(9):3465–3471
36. Gao XJ, Zhang Z, Mu TJ, Zhang XD, Cui CR, Wang M (2020) Self-attention driven adversarial similarity learning network. Pattern Recognition
37. Liu L, Zhang H, Xu X, Zhang Z, Yan S (2019) Collocating clothes with generative adversarial networks cosupervised by categories and attributes: A multidiscriminator framework. IEEE Transactions on Neural Networks pp 1–15
38. Zhu W, Liang S, Wei Y, Sun J (2014) Saliency optimization from robust background detection. computer vision and pattern recognition pp 2814–2821
39. Li C, Gao G, Liu Z, Yu M, Huang D (2018) Fabric defect detection based on biological vision modeling. IEEE Access 6:27659–27670
40. Li CL, Gao GS, Liu ZF (2017) Fabric defect detection algorithm based on histogram of oriented gradient and low-rank decomposition. Journal of Textile Research 38(3):153–158
41. Kong T, Yao A, Chen Y, Sun F (2016) Hypernet: Towards accurate region proposal generation and joint object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 845–853

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.