Fabric defect detection via a spatial cloze strategy

Article *in* Textile Research Journal · November 2022

DOI: 10.1177/00405175221135205

4 authors, including:

Zhengyang Lu
Jiangnan University
16 PUBLICATIONS **123** CITATIONS

SEE PROFILE

Textile Research Journal

Original article

Textile Research Journal
0(0) 1–16

DOI: 10.1177/00405175221135205
journals.sagepub.com/home/trj
SAGE

Fabric defect detection via a spatial cloze strategy

Zhengyang Lu[1], Yudian Zhang[2], Han Xu[3] and Han Chen[3]

Abstract

Deep-learning models have achieved state-of-the-art performances in a wide range of defect detection tasks. However, an inescapable criticism of one-stage fully supervised models is the lack of interpretability, which not only reduces the reliability of fabric defect detection systems but also limits the scope of their applications in production environments. To tackle the data imbalance and low interpretability of defect samples, we proposed a spatial cloze strategy for fabric defect detection, which reconstructs a local normal image and then feeds it into the detection model with the original image simultaneously. Specifically, we formulate the defect detection task as a novel image completion problem. Firstly, an end-to-end deep neural network is trained to finely restore the defect image by completing each image slice removed in sequence. Next, the progressive attention mechanism fuses the repaired normal image with the raw image, replacing the input layer of the cascade region-based convolutional neural network. Eventually, accurate instance-level defect segmentation can be obtained by comparing the repaired defect-free and the raw images.

On the Tianchi dataset, the proposed method displays superior accuracy in 92% of defect classes, with a breakthrough in various categories that have hardly ever been detected. Extensive experiments on various complex fabric defect samples demonstrate that our strategy outperforms existing advanced methods.

Keywords

Fabric defect detection, object detection, convolutional neural network, image completion

With the rapid development of deep neural network theory, significant progress in computer vision has been made in the field of defect detection. Nowadays, aided by the rapid gains in hardware, convolutional neural networks (CNNs) can be applied to production lines in various fields. Fabric defect detection is a direct application that can be implemented directly in industrial production lines, which intuitively improves textile quality and alleviates laborious work.

Fabric defect detection is a target-specific object detection with hard samples that are difficult to distinguish on the fabric texture. Thus, the widely used detection methods are not applicable to fabric tasks due to the defect samples' scarcity and the fabric texture samples' ambiguity. In fabric manufacturing lines, various factors can have an impact on the final product, such as the material type, machine failure, dye quality, yarn diameter and human interaction.[1] Typically, fabric defects represent surface faults in textiles, comprising various common and rare categories, the majority of which result from process issues and machine faults. Despite many efforts on models, defect detection remains challenging: (1) scarcity— as defects are usually rare, collecting all kinds of real defects for training is often difficult or impossible; (2) ambiguity:—the defects do not possess static semantic information and possibly refer to different events depending on various contexts, so they can be highly changeable and unpredictable. As a result, the imbalanced sample distribution leads to over-fitting of most deep-learning

[1]The Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, China
[2]School of Art and Design, Zhejiang Sci-Tech University, China
[3]School of Design, Jiangnan University, China

Corresponding author:
Han Chen, Jiangnan University, No.1800, Lihu Avenue, Binhu District Wuxi, Jiangsu 214122, China.
Email: chenhanisaac@163.com

models and difficulty in learning features from rare samples.

Inspired by the cloze test that completes a text with the original words erased,[2] we incorporate a standard two-stage deep-learning object detection method with the spatial cloze strategy. For the spatial cloze module, an end-to-end CNN is employed to predict normal slices from the remaining regions in the same image. We can estimate the instance-level defect semantic segmentation by comparing the raw image with the normal image produced from the stitched slices. The cascade region-based convolutional neural network (C-RCNN), a standard state-of-the-art object detection, is the backbone network for fabric defect detection. However, the complement module provides the expected defect-free image and the real image as input, rather than the single image in widely used detection methods. Hence, a progressive attention mechanism creates a multi-channel spatial mapping that establishes a plausible input for the detection backbone.

Extensive experiments are compared with existing advanced methods on the challenging Tianchi fabric defect dataset.[3] We provide quantitative and qualitative results to demonstrate the validity and interpretability of the proposed method. Further, reliable instance segmentation results for this defect dataset are available online.

1. We incorporate a deep-learning fabric defect detection with the spatial cloze strategy. This framework predicts each normal slice from the defect image that masks each slice and then stitches them together into a complete defect-free image.
2. A progressive attention mechanism conducts a multi-channel spatial convolution operation as the input layer of C-RCNN, which improves the feature utilization for simultaneous input of raw and predicted defect-free images.
3. The experimental results show that the proposed fabric defect detection approach achieves state-of-the-art performance on the Tianchi fabric defect dataset. Moreover, instance-level segmentations demonstrate unsurpassed interpretability and are published online.

Related works

Fabric defect detection is a critical quality control technique that aims to recognize and locate defects appearing in fabrics, replacing unstable human observation. The detection procedure is regularly deployed at the output of air-jet looms directly. In Figure 1, the detection region is highlighted with a dashed red line.

Fabric defect detection approaches can be organized into three categories: classic algorithms and one- and two-stage detection methods. The first paradigm restricts itself by only utilizing specific statistical models or formulas, such as feature engineering with prior information, preliminary statistical, structure, spectrum and model-based methods. The deep-learning approaches can be further classified into one- and two-stage detection methods. Among the existing methods, most one-stage CNNs are simply transplants from widely used object detection models

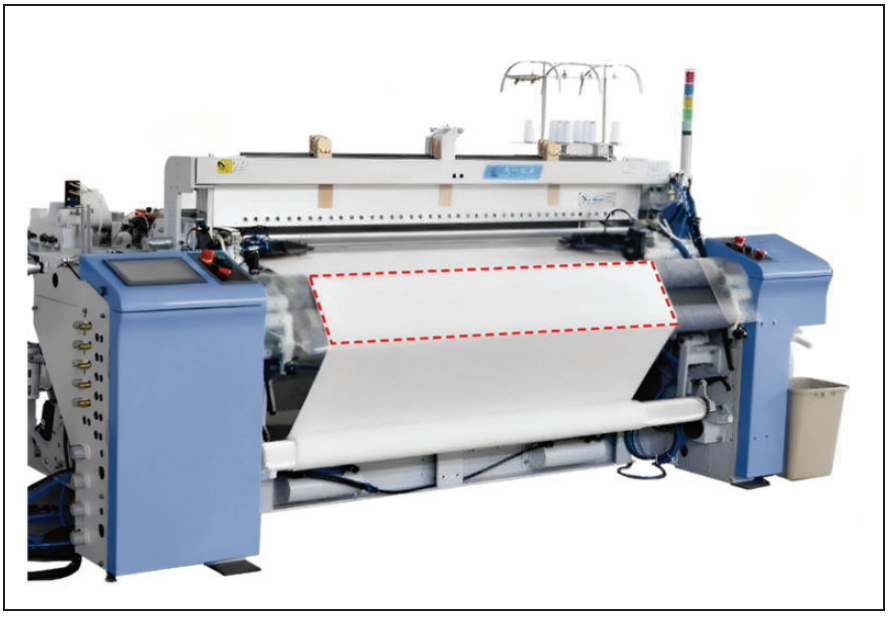

Figure 1. Defect detection deployment on air-jet looms. The red dashed line marks the detection region. (Color online only.)

and suffer from numerous applicability problems. Two-stage detection methods are primarily deployed to generate additional defects to alleviate the shortage of negative samples and the imbalance in sample distribution.

Classic methods

For the preliminary statistical processing, classic methods employ the value distribution of pixels in single frames, for example, gray-level co-occurrence matrices, auto-correlation decomposition and high-dimensional manifold features. Kumar and Hafedh[4] developed a statistical model for detecting fabric defects through eigenvalues. Fabric defective portions were recognized with the variation coefficients. Song et al.[5] estimated the similarity of each sliced region in order to detect fabric defects rapidly. The saliency of defective regions is determined by combining the image's extreme point density map with the features of the similarity region. Ben Gharsallah and Ben Braiek[6] provided an modified anisotropic diffusion filter and saliency extractor for fabric defect recognition. Because standard anisotropic diffusion techniques cannot distinguish between the defect edge and the background texture, the modified anisotropic diffusion model combined the local gradient magnitude and a saliency map.

In structural approaches, the most efficacious solution to segment defective regions on patterned textile images is the golden image subtraction approach.

Ngan et al.[7] developed an automatic visual inspection method for identifying defects in textured fabrics using wavelet transform golden image subtraction. Abouelela et al.[8] introduced a fabric defect detection system incorporating basic statistical metrics to reach real-time performance.

Thus, this proposed method traded the low complexity of the basic statistical features for speedy detection in production deployments. Meanwhile, this method provided better performance in detecting defects with drastic physical changes.

According to spectral approaches, the frequency domain theory primarily incorporates Fourier and wavelet transform. Li et al.[9] provided a solution for adaptive textile defect detection and improved production quality on the basis of multi-scale wavelet transforms and Gaussian mixture models. To build an efficient and robust system, Rebhi et al.[10] proposed a textile defect detection model based on local homogeneity features and discrete cosine transform (DCT). The DCT was employed on the computed homogeneous image and then extracted the various energy features from all DCT blocks, which were treated as input to the back-propagation neural network.

In model-based methods, Ngan et al.[11] proposed an ellipsoidal decision region for motif-based patterned textile defect detection. Extensive experiments proved the validity of the previous detection through the decision region of energy-variance values. The textile defect detection approach in homogeneous and structured fabrics introduced by Bissi et al.[12] has been validated on the TILDA dataset.[13]

One-stage detections

With the rapid development of neural network theory, numerous deep-learning techniques have been applied to fabric defect detection tasks, achieving unsurpassed results for improving the quality of textile products and replacing laborious human checks.

Whilst deep-learning methods have proven to be simple and efficacious when processing recognition and detection problems, various problems exist for target-specific tasks. The first problem is the real-time performance of the models, which is a fundamental requirement for efficient production line applications. The second problem is that defect samples are difficult to capture. Thus, the majority of the fabric dataset is defect-free samples, which leads to sampling imbalance and long-tailed data distribution, challenging the optimization process of the model.

For deep-learning-based object detection techniques, existing solutions can be categorized into one-stage and two-stage detectors. Most one-stage detectors are fast but have unreliable detection accuracy, making it hard to reach production accuracy demands. Nevertheless, two-stage approaches provide more accurate detection results, but reaching the real-time requirements for production scenarios is problematic.

In contrast to two-stage detectors, single-stage methods omit an independent proposal generation stage. Such methods regularly perceive arbitrary areas on an image as potential objects and seek to classify each region of interest (RoI) as a target object.

The single-shot multi-box detector (SSD)[14] has demonstrated a superior detection performance, but the original SSD model suffers in detecting small objects. Therefore, Liu et al.[15] presented a one-stage SSD model for fabric defect detection. By combining motif determination, candidate generation and a neural network, Ouyang et al.[16] introduced a deep-learning method for on-loom defect detection. In addition, a paired-potential activation layer was inserted into the neural network, resulting in highly accurate defect segmentation on fabrics with complex features. Li et al.[17] proposed a compact deep-learning architecture that combines multi-scale features, filter discretization, multi-position ensembles and parameter simplification to detect a few regular textile defects. Further, the

compact network was optimized through micro-architecture with multi-layer perceptions. Considering the fabric surface properties, resolution and defect appearance, Zhou et al.[18] proposed an efficient deep CNN, called an efficient defect detector. This method also presented a feature fusion strategy, namely the L-shaped feature pyramid network, to extract complete low-level texture features. Because factories necessitate higher real-time performance, Jing et al.[19] introduced the Mobile-Unet, a highly efficient backbone, to perform the defect segmentation task.

Two-stage detections

Compared to single-stage networks, two-stage detectors have higher accuracy and complexity. In the two-stage detector, a set of candidate RoIs is estimated in the first stage. Next, the features of the candidate RoIs are determined with deep CNNs in the second stage.

As a milestone in two-stage detectors, the fast region-based convolutional neural network (RCNN)[20] improved the raw RCNN[21] by deploying a region proposal network (RPN).

Hence, some research developed from the RCNN[21] has been adapted to fabric defect detection.[22] Li and Li[23] employed numerous tricks on the C-RCNN to improve the accuracy of textile defect detection, but there are no developments in the model structure or algorithmic strategy.

Recently, generative adversarial networks (GANs)[24] provide the most surprising idea that can be adopted in most computer vision tasks, including defect detection and segmentation. The GAN-based fabric defect detection method is able to fit various fabric patterns through learning from the labeled textile defect samples. Liu et al.[25] introduced a complex semantic segmentation network to detect texture defects. The multi-stage GAN model synthesizes plausible defects from defect-free samples. Le et al.[26] employed Wasserstein generative adversarial nets (WGANs) to construct a promising learning-based model for fabric defect detection. Further, this method was enabled to tackle severe imbalances in defective samples, which is necessary for real-world scenarios. To further improve the detection accuracy, Zhou et al.[27] introduced a mixed semi-supervised approach for textile defect detection using a variational auto-encoder, which is an efficient generator for rare samples, and a Gaussian mixture model. Aiming at high precision, Dlamini et al.[28] developed a defect detection system with extensive data augmentations and a YOLOv4 backbone.[29]

In summary, one-stage detectors are simple in principle and lightweight for deployment. On the other hand, the second-order method provides a preliminary solution to the sample imbalance problem, by providing reasonable intermediate region proposals. Previous methods remain major problems that need to be addressed in fabric defect detection, for example, inaccurate hard sample classification and missing instance-level segmentation.

Methodology

This section introduces the proposed fabric defect detection pipeline before the implementation details. The pre-complemented detection framework is shown in Figure 2, which utilizes a spatial cloze network for image completion and a detection network for defect recognition. As shown in Figure 3, the spatial cloze strategy for defect detection is simple in concept, aiming to reconstruct plausible fabric texture slices from the pre-masked spatial contexts and stitching to predict the complete defect-free image. The C-RCNN, a standard advanced object detector, is the backbone for fabric defect detection. Next, a progressive attention mechanism is inserted to the input layer of the C-RCNN, which improves the feature utilization for simultaneous input of normal and predicted defect-free images.

Spatial cloze strategy

The cloze test, which completes a text with the original words erased,[1] provides a solution to exploit spatial context information.

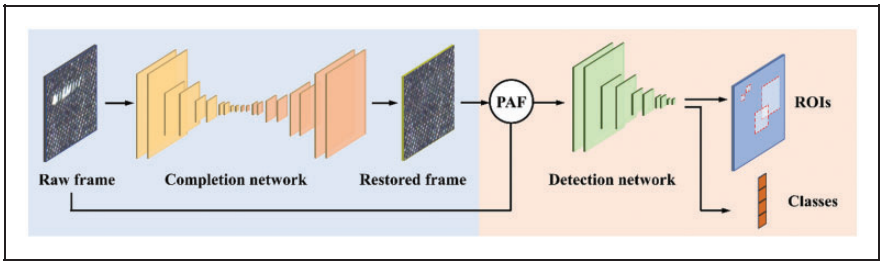

Figure 2. The framework of fabric defect detection via the spatial cloze strategy. The left-hand part aims to reconstruct accurate defect-free frames. In the right-hand part, the original and restored images are fed into the detection network via progressive attention fusion (PAF). RoIs: regions of interest.

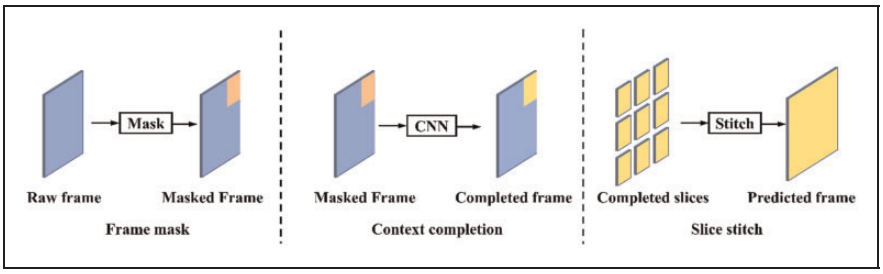

Figure 3. The framework of fabric defect detection via the spatial cloze strategy. The left-hand part aims to reconstruct accurate defect-free frames. In the right-hand part, the original and restored images are fed into detection network via progressive attention fusion (PAF). CNN: convolutional neural network.

Previous deep-learning works cannot fully extract spatial context information due to the direct object detection approaches. Most methods for fabric defect detection are simply transplanted from widely used object detection frameworks, such as YOLOv3,[30] RCNN,[21] fast RCNN[20] and C-RCNN.[31] However, the extremely imbalanced data distribution is a significant challenge, especially for fabric defect samples, where standard methods are not applicable.

As a solution, we propose a spatial cloze strategy to avoid fabric defect detection of end-to-end network over-fitting due to long-tailed distribution. Figure 3 illustrates the core framework for the cloze strategy, which aims to reconstruct defect-free fabric images. Firstly, we mask each region where the raw frames are divided equally in sequence, aiming to collect masked and normal frame pairs. It is worth noting that only defect-free slices and corresponding normal slices are applied for the next stage, since the purpose is to predict normal slices from spatial contexts. The second step is to train a standard end-to-end deep-learning model to complement each masked image region. Finally, as each complemented frame is full size, we crop the predictions from each masked region and stitch them together to compose complete predicted frames. The prediction produces a defect-free image that can be fed into the detection method by fusing it with the raw image. Furthermore, the cloze strategy has unsurpassed interpretability, as instance-level segmentation results can be calculated directly.

In the experiment implementation, it is worth noting that some fundamental details are required, including the cloze strategy's slice number and the complementary network's loss function. For the split-complement network, the complementary loss L_c comprising the structure similarity index measure (SSIM)[32] and mean square error (MSE) between the predicted defect-free and raw image, is

$$L_c = \mathrm{MSE} + \alpha \mathrm{SSIM}$$

where

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where μ_x is the average of x, μ_y is the average of y, σ_x^2 is the variance of x, σ_y^2 is the variance of y and σ_{xy} is the covariance of x and y.

After determining the loss function, the optimal number of slices, another critical parameter, will be estimated in further ablation experiments.

Progressive attention fusion

In most existing methods, the input to the detection network is one single image.

Nevertheless, we simultaneously obtain the repaired normal image and the raw defect image in the spatial cloze module. Therefore, it is crucial to establish an efficient multi-frame fusion approach as an alternative to simple concatenation.

As shown in Figure 4, the progressive attention mechanism of the detection framework provides a solution incorporating multiple relevant images as input, which is efficient and easy to implement. Regarding the symbols in the diagram, the kernel of the convolution layer is 1×1, and the plus sign represents a concatenation operation. Different from classic attention methods, this attention not only performs the convolution operation for multiple channels but also establishes a multi-channel progressive convolution operation, which concatenates the predicted frames with the original frames that have been operated by three convolutions and then concatenates them with the predicted frames with three convolution operations as input to the C-RCNN.

Progressive interaction in fusion processes provides rich feature interaction and thus enhances information utilization. The validity of the proposed progressive attention fusion will be demonstrated in the following experiments.

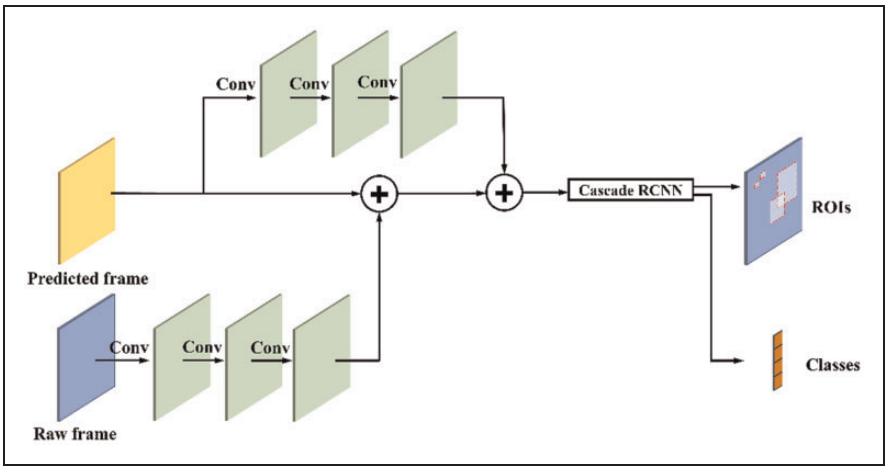

Figure 4. Progressive attention fusion for the detection framework. Inspired by the attention mechanism, we seek to establish an efficient way to fuse the repaired and raw frames, replacing simple concatenation. RCNN: region-based convolutional neural network; RoIs: regions of interest.

Cascade RCNN

Currently, object detection is one of the best developed fields in computer vision, and thus the advanced C-RCNN[31] is adopted as the detection network. In contrast to standard detection with a single image input, the spatial cloze mechanism provides real and completed normal image pairs.

The cascade RCNN is introduced as a cascade regression problem by decomposing complex tasks into simple steps. In Figure 5, "I" donates the input image, "conv" donates the feature encoder, "pool" donates the region-wise feature decoder, "H" represents the network head, "B" is the annotation box and "C" is the class. The main function of the RPN is to provide region proposals, which can be regarded as numerous potential anchors. In the classic RCNN,[20] most of the proposals selected by the RPN are low quality, making it hard to adopt a high-threshold detector directly. The C-RCNN utilizes cascade regression as a re-sampling mechanism to increase the intersection of union (IoU) value stage by stage, thus enabling the re-sampled proposals from the previous stage to fit the next one with a higher threshold. The improved connection is to concatenate a series of regressors

$$f(x,b) = f_T(x,b_T) \times f_{T-1}(x,b_{T-1}) \times \cdots \times f_1(x,b_1)$$

where T donates the candidate quantity of the cascade steps. In the cascade structure, each regressor f_T is optimized against the sample distribution b_t reaching the relevant step, rather than the initial distribution of b_1. This cascade structure improves box location accuracy in a step-by-step manner.

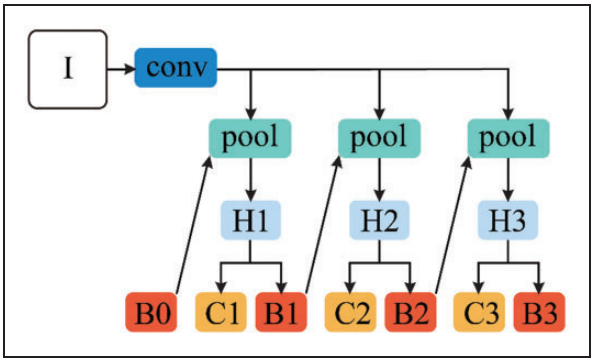

Figure 5. The cascade detection framework extends beyond the standard parallel connection.

Experimental details

Spatial-cloze-assisted C-RCNN is extensively validated on fabric defect detection and localization. Firstly, various components and hyper-parameters of the proposed method are determined from the ablation experiments. Next, the quantitative and qualitative metrics are placed in a wider perspective by comparing our model with existing methods for fabric defect detection. Furthermore, failure cases are discussed at the end of the section to inspire further research directions.

Fabric defect dataset

To prove the proposed method's validity, we utilize the Tianchi fabric defect dataset[3] captured from the real textile production scenes. The Tianchi fabric defect dataset is the only one used to validate the proposed method because previous datasets, such as the TILDA dataset,[13] do not have sufficient defect categories and samples. Furthermore, this dataset is the most

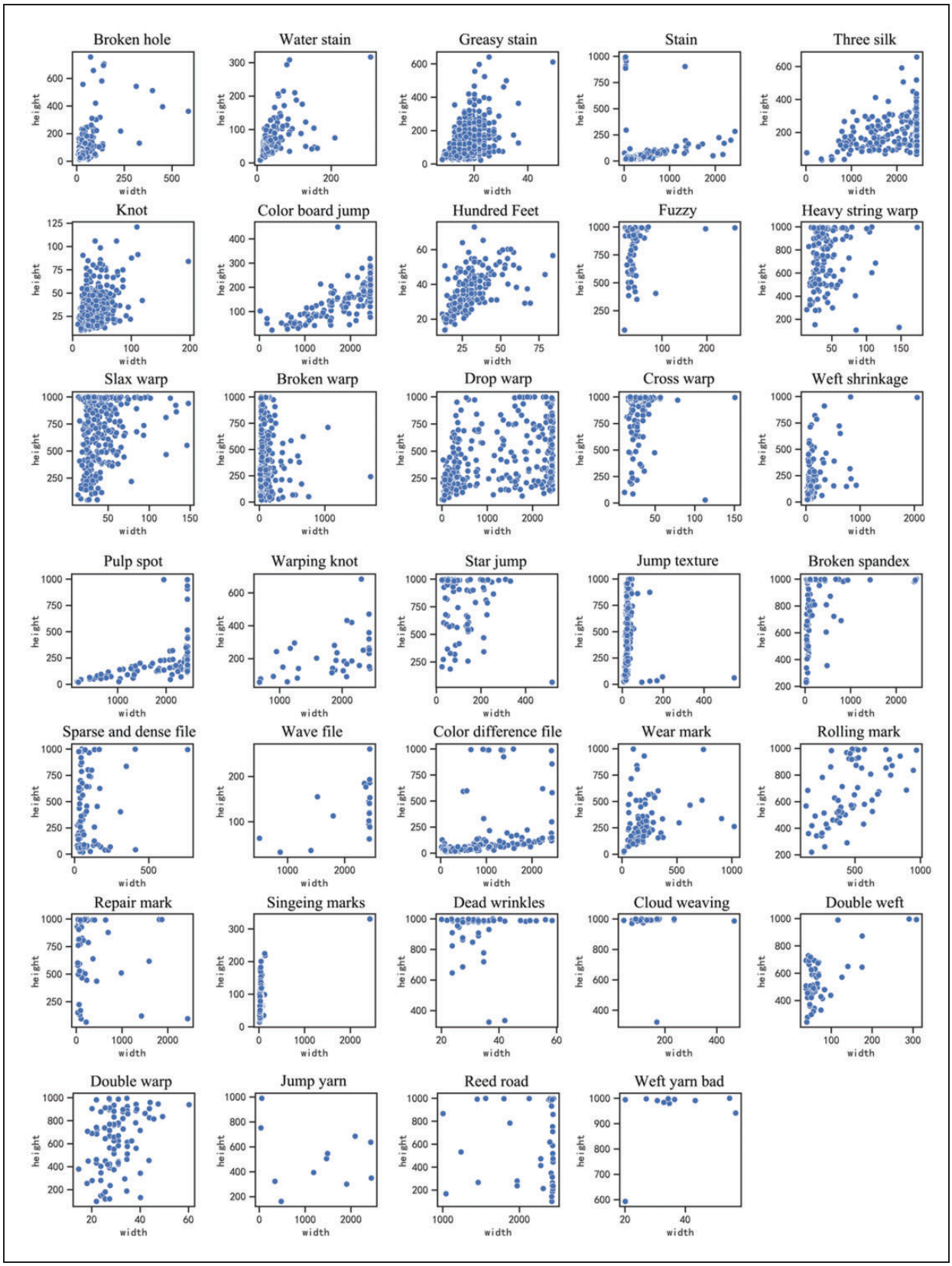

Figure 6. The Tianchi fabric dataset contains various defect categories and uncertain aspect ratios.

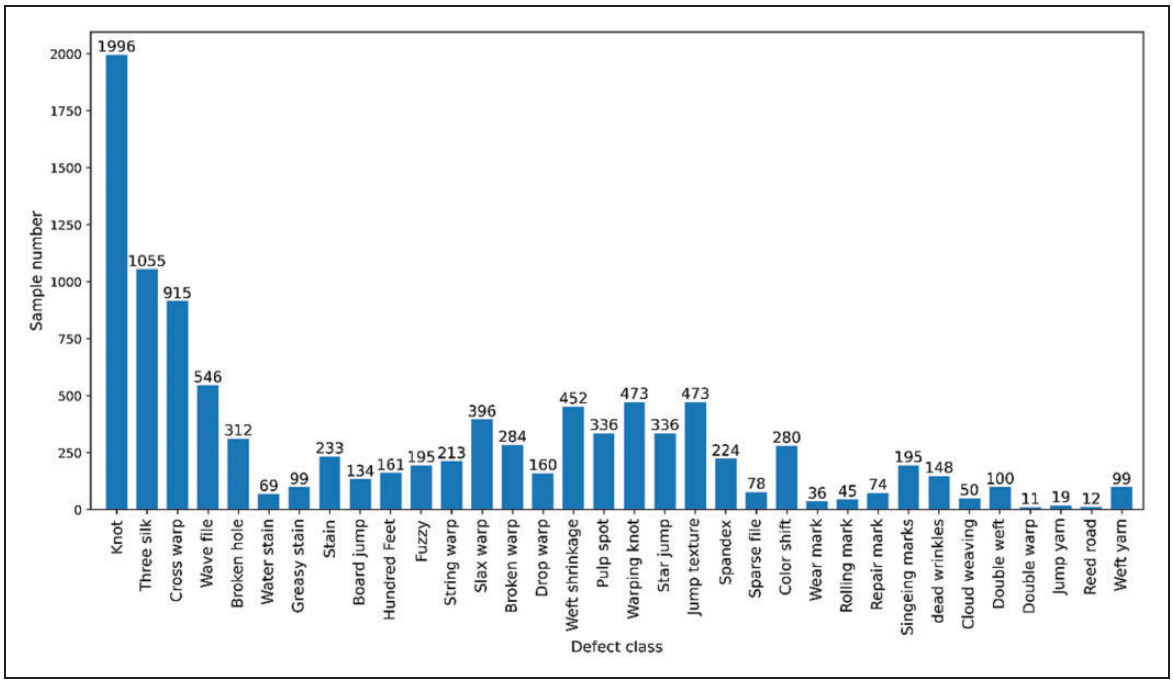

Figure 7. The distribution of defect categories in the Tianchi fabric dataset is extremely imbalanced.

comprehensive and complex defect dataset. This dataset contains 9576 images, whose size is 2446×1000, comprising 3663 normal images and 5913 defect images. The total number of defects is 13,186, containing 34 defect classes with variable aspect ratios, as shown in Figure 6. We split the dataset into training and test sets, with 80% and 20%, respectively. The image number in the training set is 7660, including 2902 normal and 4758 defect images, with the defect number of 10,468. Regarding the test set, the number of fabric images in real-world scenes is 1916, comprising 761 normal images and 1155 defect images.

Figure 7 displays the distribution of defect categories and quantities in the Tianchi fabric defect dataset. It can be noticed that the number distribution of defects is extremely imbalanced, with some images having up to several thousand defects and others having almost none. The extreme data imbalance demonstrates the advantages of adopting the spatial cloze method. As shown in Figure 8, the multiple co-existing defects are difficult to detect.

Evaluation metrics

Following MS-COCO 2017,[33] the widely used average precision (AP) averages precision across IoU thresholds from 0.5 to 0.95 at intervals of 0.05. These evaluation metrics allow measuring the detection performance with various requirements. Due to sampling complexity and data imbalance, a fault-tolerant AP metric whose IoU is greater than the threshold is regarded as a correct result is adopted for the fabric defect task. Furthermore, the following experiments provide mAP, AP_{50}, AP_{70} and AP_{90} metrics for detection results.

Implementation details

We implement the proposed method using PyTorch on a single Nvidia 2080s GPU. We build the cloze model based on Unet,[34] ResNet-18, ResNet-34 and ResNet-50.[35] In completion processing, the mask and the image that removed the mask part are fed into the completion network simultaneously. It is worth noting that only defect-free slices are employed for completion training. Furthermore, the removed part is filled with the pixel average of the complete image. For the split-completion network, the loss weight α is set to 0.1. The learning rate of the completion network is set to 10^{-4}, and the regular term is 10^{-5}.

The C-RCNN uses the MS-COCO[33] pre-trained ResNet-34 as the model encoder. In the fabric detection network, learning rates are 10^{-4} for the pre-trained convolution layers and 10^{-2} for the other layers, reduced after each epoch by a factor of 0.1. In this framework, both networks have a selected stochastic gradient descent (SGD)[36] as the optimizer.

Ablation experiments

To determine the hyper-parameter for the spatial cloze module, Table 1 displays ablation experiments of slice

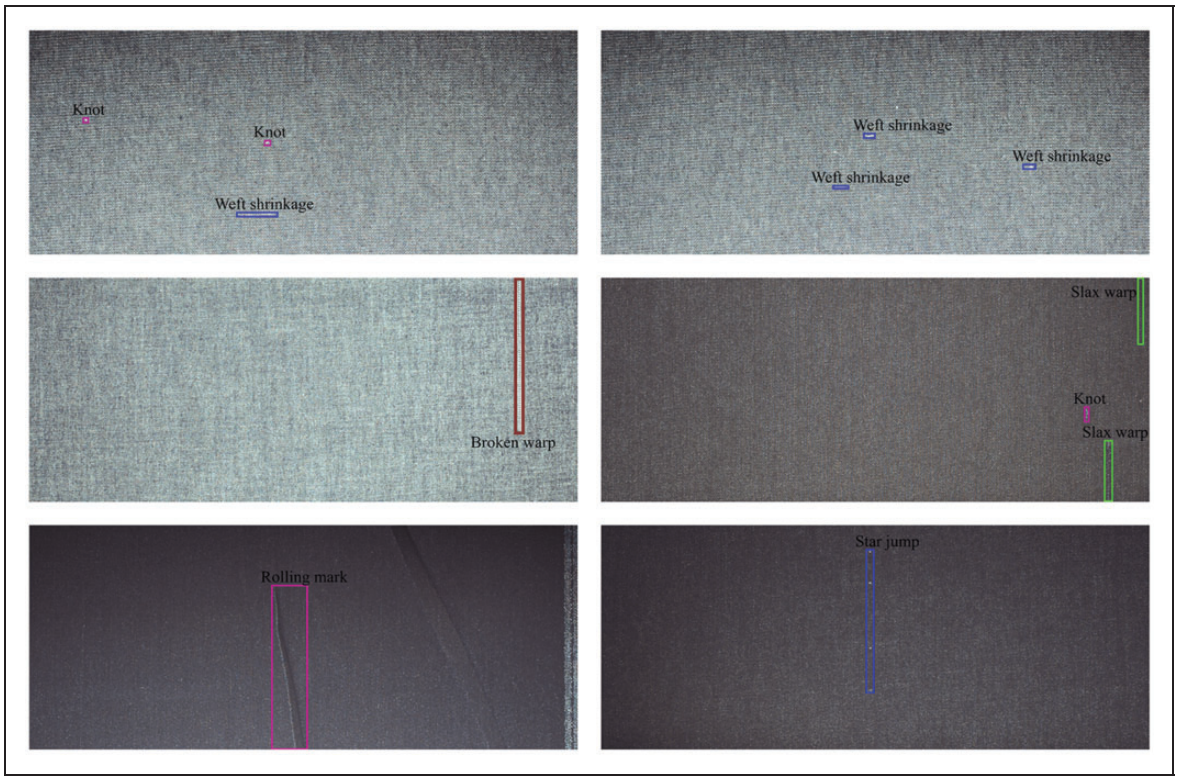

Figure 8. Hard samples from the Tianchi fabric defect dataset.

Table 1. Ablation experiments to determine the optimal slice number on the Unet backbone

Slices					Accuracies			
1	2	3	4	5	AP_{50}	AP_{70}	AP_{90}	mAP
√					28.91	22.65	4.54	14.82
	√				35.43	28.97	10.32	18.14
		√			38.42	32.48	12.85	19.20
			√		38.51	32.58	12.93	19.29
				√	38.52	33.61	12.99	19.34

numbers on the fabric defect dataset, which illustrates the effectiveness of the proposed pre-complemented framework. Intuitively, the slice number is an independent variable, and thus the optimal parameter is determined individually on the Unet backbone.

In contrast to theoretical expectations, the accuracy does not consistently increase as the size of the predicted image slice gets smaller. The reconstruction performance reaches a bottleneck when the slice number is three, with mAP stabilizing at 19.20. From the qualitative results shown in Figure 9, the estimated defect-free images confirm the existence of a bottleneck. In other words, continuing to increase the slice numbers has a negligible improvement in accuracy. Therefore, a compromise has been made between complexity and accuracy to determine the slice number for the following experiments as three. Furthermore, Figure 9 displays instance-level defect segmentations, which are determined through the error normalization between the predicted defect-free image and the real image. These results demonstrate that it is feasible to localize defects accurately without labeled information, despite it not being possible to determine defect classes.

In Table 2, we focus on selecting the optimal backbone of the completion module and simultaneously demonstrating the accuracy benefit from the progressive attention fusion. For comparison, the baseline method is set to the original C-RCNN without the spatial cloze strategy. From apparent results in Table 2, the ResNet series[35] achieves higher accuracies than other backbones involved in the comparison. The predicted images shown in Figure 10 demonstrate that the selected backbones are sufficient for the complementation task. On the other hand, the high complexity brings high accuracy. Considering the computational complexity, ResNet-18 can meet the speed of the production line, while the accuracy reaches a bottleneck approximately. It is evident that progressive attention fusion yields far more accuracy improvement than a simple concatenation operation. The progressive attention, a bidirectional interactive attention mechanism,

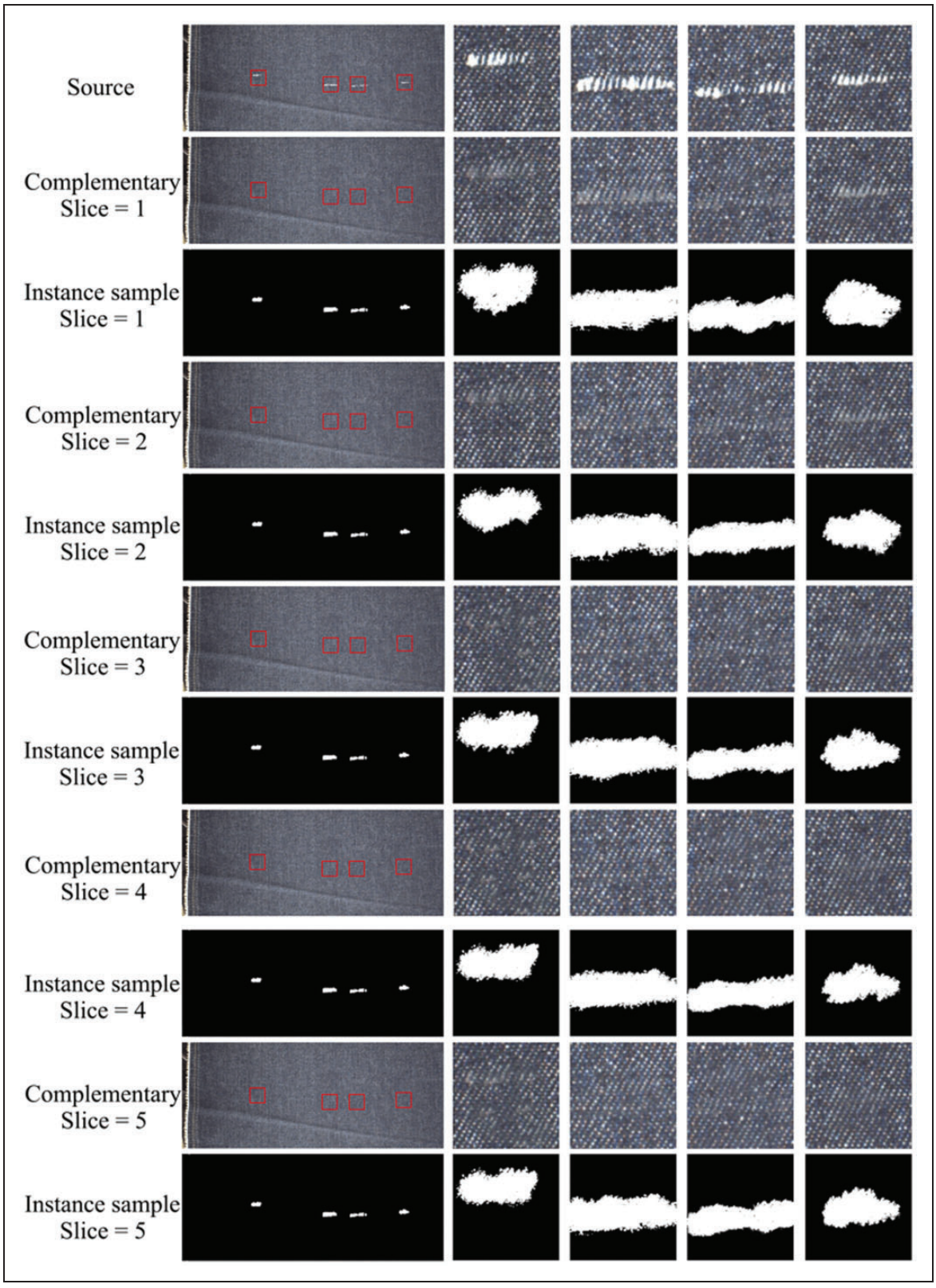

Figure 9. Left-hand images show the full defect samples. Right-hand images focus on the corresponding reconstruction results and zoomed details.

Table 2. Ablation experiments to determine the cloze backbone and fusion method

	Fusion method		Accuracies				Speed
Backbone	Concatenation	P-attention	AP_{50}	AP_{70}	AP_{90}	mAP	(frame/s)
Baseline	–	–	28.95	22.61	4.57	14.89	13.8270
Unet-5	√		34.08	29.26	12.03	17.72	3.3110
ResNet-18	√		34.80	29.37	12.10	18.07	6.2670
ResNet-34	√		34.87	29.51	12.12	18.11	4.8740
ResNet-50	√		34.91	29.65	12.21	18.21	3.4780
Unet-5		√	38.42	32.48	12.85	19.20	3.2320
ResNet-18		√	38.70	32.62	12.93	19.42	6.1100
ResNet-34		√	38.78	32.69	12.98	19.49	4.7811
ResNet-50		√	38.83	32.78	12.99	19.53	3.4289

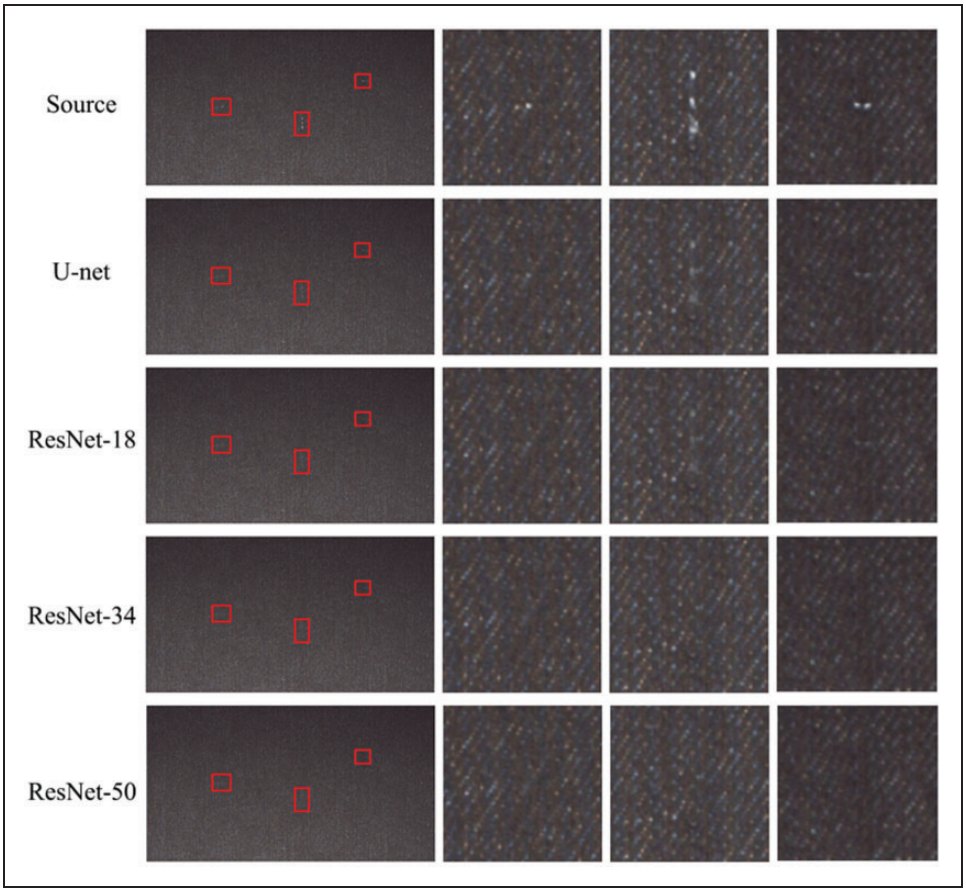

Figure 10. The qualitative results of the completion network with various backbones demonstrate their validity, but the visual performance improves as the network complexity increases.

Table 3. Accuracy comparison with existing advanced methods

Method	AP_{50}	AP_{70}	AP_{90}	mAP
YOLOv4	28.1299	21.8381	4.3484	14.6848
SSD	28.1546	20.7802	4.1265	13.8863
RCNN	23.5810	16.4770	3.5090	10.3530
C-RCNN	28.9510	22.6120	4.5770	14.8870
Ours	**38.4880**	**32.4830**	**12.8900**	**19.2090**

SSD: single-shot multi-box detector; RCNN: region-based convolutional neural network; C-RCNN: cascade region-based convolutional neural network.

provides extra accuracy improvement from more sophisticated intermediate pathways. Furthermore, the weights of both the original and defect-free images can be automatically adjusted for the model input. In the following experiments, ResNet-18 is selected as the backbone, and progressive attention fusion replaces direct concatenation.

Comparisons with state-of-the-arts

In the following experiments, we evaluate the proposed deep-learning fabric defect detection with the spatial cloze strategy on the Tianchi dataset. For fair comparisons, an official split of the training and testing data is applied to the following experiments.

The quantitative results are displayed in Table 3, and these values are collected from the experiments we conducted. It is worth noting that the basic model architecture and pre-training parameter setting of the RCNN and the C-RCNN are the same as in our method. As we have observed, our model, which incorporates the ResNet-18-based completion model and the C-RCNN detection framework, yields the best performance in all metrics, with a shocking 50% accuracy improvement.

In particular, in the AP_{90} metric, the proposed method improves by approximately 182% relative to the basic C-RCNN, which implies a breakthrough in high-precision defect detection.

For detailed comparisons, we provide the values of mAP metrics for each category, as shown in Table 4. Compared with the basic C-RCNN, our method has superior accuracy in 88.24% of defect classes, with a breakthrough in various categories that have hardly

Table 4. Category-level accuracy comparison with existing methods

Defect category	YOLOv4	SSD	RCNN	C-RCNN	Ours
Broken hole	38.1820	38.5200	34.6060	39.3900	**49.7100**
Water stain	9.0070	7.7050	3.1600	10.2360	**18.2630**
Greasy stain	31.6550	31.2900	33.7800	32.4310	**40.7610**
Stain	19.1850	18.6840	19.1010	20.0380	**23.2240**
Three silk	34.6820	34.9400	31.1700	**35.8900**	35.0810
Knot	19.3040	19.1130	8.2980	21.8010	**23.7710**
Board jump	44.7170	44.1190	38.2010	45.9370	**59.7220**
Hundred feet	20.6670	20.1960	15.7140	22.3430	**47.7700**
Fuzzy	6.0820	5.8410	2.7740	6.7160	**8.8240**
String warp	3.5590	2.4770	0.6100	4.1630	**47.4860**
Slax warp	14.6180	14.8860	11.1602	15.8420	**39.7960**
Broken warp	12.1900	11.3340	13.4720	12.0770	**22.0700**
Drop warp	3.6900	3.7280	2.8580	4.7990	**21.0530**
Cross-warp	19.1970	18.7620	16.4670	**20.6400**	19.2280
Weft shrinkage	4.9270	4.5430	4.1860	5.9080	**21.4430**
Pulp spot	40.2530	38.9420	35.9230	**41.1940**	39.2460
Warping knot	15.3730	14.4080	3.7780	18.1950	**26.3840**
Star jump	2.8190	2.7760	0.7710	3.7910	**21.2660**
Jump texture	26.3110	26.1710	24.5160	27.7730	**45.7100**
Broken spandex	14.8220	15.0770	11.3512	16.1980	**29.1630**
Sparse file	1.6050	2.3620	2.5310	2.3830	**15.3810**
Wave file	0.0000	0.0000	0.0410	0.0000	**2.1100**
Color shift	4.9540	4.5100	0.0000	6.3370	**8.1550**
Wear mark	4.2930	4.4090	5.5410	5.0440	**9.8890**
Rolling mark	14.3590	12.9950	10.6001	15.1200	**25.2770**
Repair mark	4.1880	3.3790	3.1720	4.5240	**33.6120**
Singeing marks	55.6700	54.7590	49.1280	57.0280	**62.8560**
Dead wrinkles	0.5860	0.2410	0.2610	0.3110	**3.1440**
Cloud weaving	0.7210	0.0000	0.0000	0.0000	0.0000
Double weft	0.0000	0.2320	0.0000	0.0000	**47.2510**
Double warp	0.0000	0.0000	0.0000	0.0000	**15.8660**
Jump yarn	3.2200	1.8450	1.9640	3.9110	**45.9250**
Reed road	0.0000	0.0000	0.0000	0.0000	**31.5480**
Weft yarn	36.6320	36.7740	27.1970	38.7870	**51.1880**

SSD: single-shot multi-box detector; RCNN: region-based convolutional neural network; C-RCNN: cascade region-based convolutional neural network.

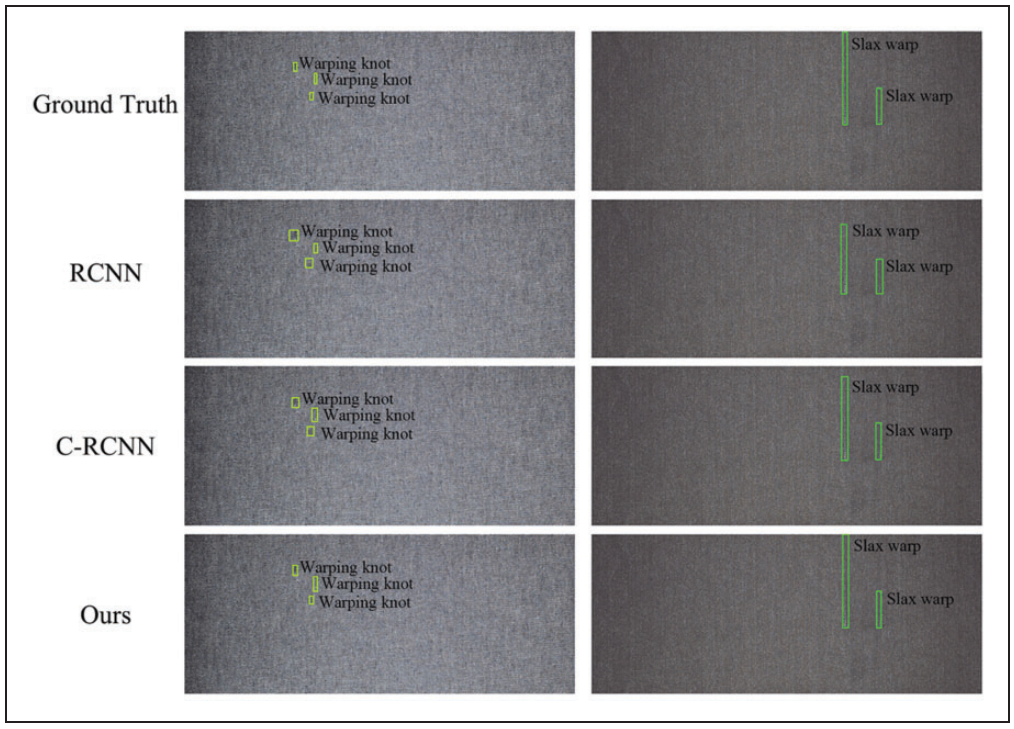

Figure 11. Comparison of qualitative experimental results with dense and high aspect ratios. RCNN: region-based convolutional neural network; C-RCNN: cascade region-based convolutional neural network.

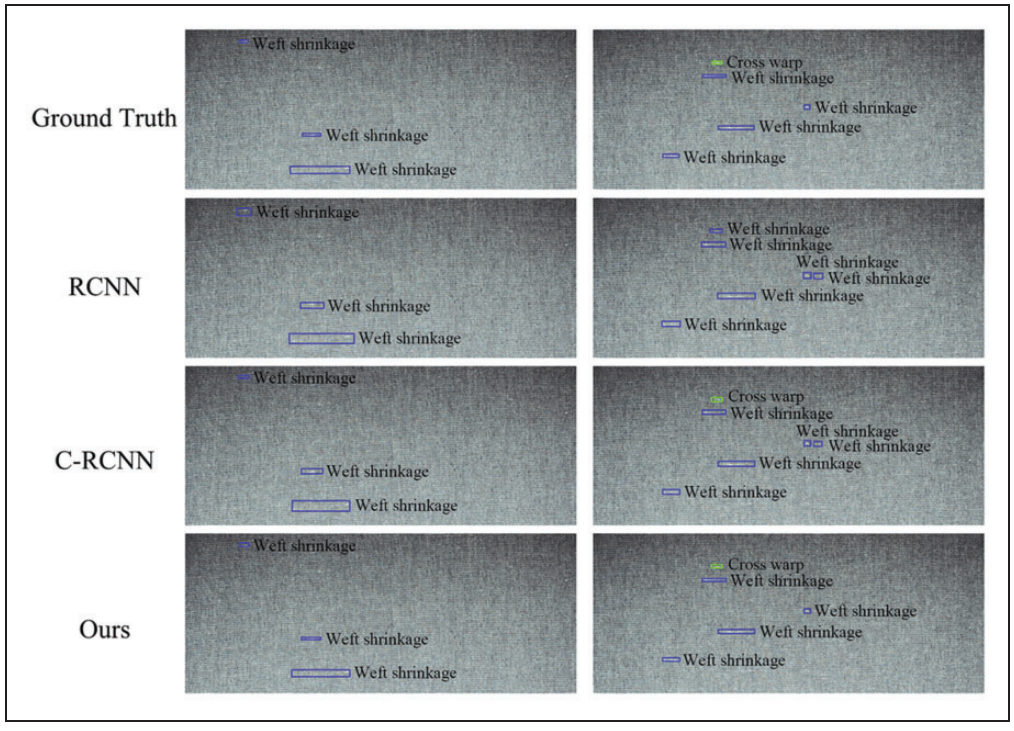

Figure 12. Comparison of qualitative experimental results with multiple complex defects. RCNN: region-based convolutional neural network; C-RCNN: cascade region-based convolutional neural network.

ever been detected. It is worth considering that the results are not favorable in four categories; in particular, all models fail to detect cloud weaving. For quantitative comparisons, this method is unrivalled, not only for specific categories but also for comprehensive comparisons.

In addition to the mAP, a more important metric is the intuitive evaluation of the image reconstruction because the deep-learning method aims to go beyond human visual perception. Qualitative comparisons are displayed in Figures 11 and 12. Note that Figure 11 demonstrates the defect detection results on dense

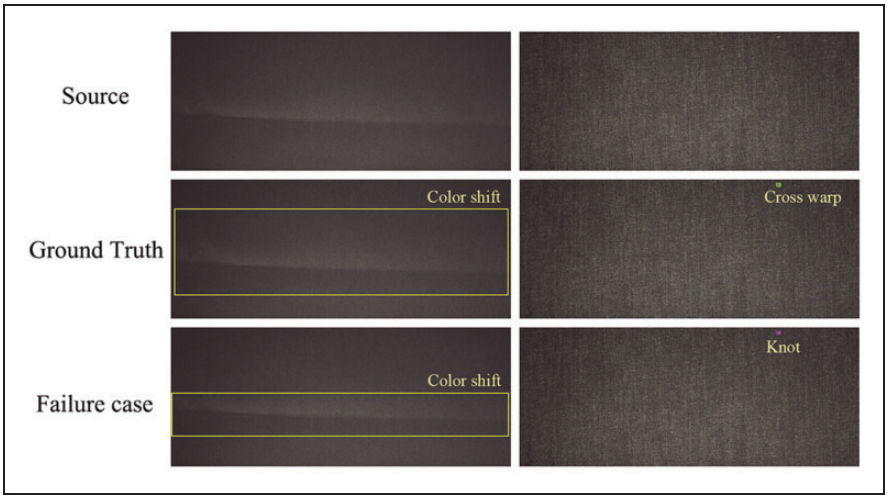

Figure 13. Specific samples demonstrate inaccurate location boxes and misclassification categories.

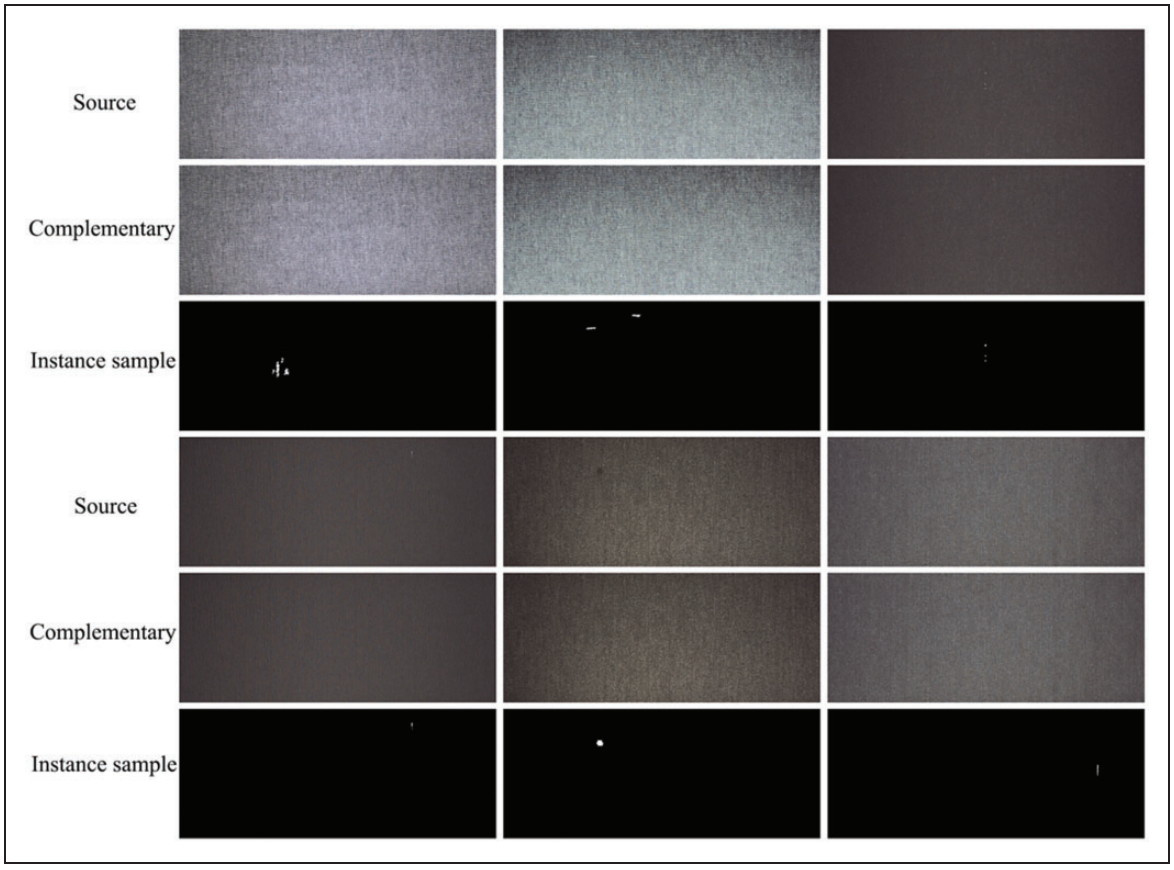

Figure 14. Complementary defect-free images and pixel-level defect semantic segmentations are visually accurate.

and high aspect ratio samples. The small sample size in the left-hand column causes the position drift in all methods. However, the proposed method has the most accurate boxes. In the right-hand column, our method retains the shape of the defect target localization box with high aspect ratios compared to the inadequate length of the detection box's long edges from previous methods.

Further, Figure 12 provides detection results for scenarios with multiple complex defect samples. The localization boxes in the left-hand column from various methods have varying errors, but our method is the most accurate. The high density of defect samples in the right-hand column images led to misclassification by previous methods. The cross-warp is mistaken as weft shrinkage in the RCNN, and one more weft shrinkage is detected erroneously in the RCNN and the C-RCNN.

Overall, our approach surpasses existing advanced approaches in qualitative and quantitative comparisons. Thus, the proposed method is a highly accurate real-time algorithm that can be deployed in production lines. On the other hand, the weaknesses for the proposed model are obvious: (1) the independent two-stage networks require separate training; (2) the model training requires numerous samples, implying that it is difficult to adapt to few-shot datasets.

Failure cases

Potential drawbacks can be explored in the spatial cloze strategy, which is mined from the detection failure cases in Figure 13. As shown in the left-hand column, we consider our localization is more precise compared to the ground-truth box, which is an ambiguity that is even difficult to distinguish for humans. In the right-hand column, the proposed detection method incorrectly recognizes the cross-warp as a knot, despite the precise localization. For such difficult samples, we expect to improve the classification accuracy with strict contrast loss in future research.

Segmentation results

With the rapid development of high-performance computing devices, high-precision models have evolved faster. The proposed method may be replaced in a few years, but the high-quality annotation of the dataset can consistently contribute to advancing the defect detection field.

To capture instance-level segmentation, we set a threshold of 0.2 in the comparison stage, that is, the area is considered to be an anomaly if the pixel value difference is greater than 50 after gray-scaling. To confirm the validity of the semantic segmentation results, we conduct a laborious manual check on the instance-level results. The 73 faulty images out of 4758 segmentation images, implying a 1.53% error probability, require further manual corrections. Despite the imperfect results from the model, it greatly alleviates the laborious instance-level segmentation work.

As shown in Figure 14, complementary defect-free images and pixel-level defect segmentation are visually accurate. Model-generated complementary images and semantic segmentation contribute to create an enhanced dataset, aiming to augment the original samples and replace crude box annotations.

Conclusion

To alleviate the scarcity and imbalance of defect samples, we propose a deep-learning method that applies the spatial cloze strategy to fit complex production scenarios. This slice-completion strategy avoids importing defect information as it only learns from defective-free image slices. A progressive fusion mechanism conducts a multi-channel spatial convolution operation as the input layer of the C-RCNN, which is the backbone of the detection module. Extensive experiments have demonstrated that the proposed fabric defect detection framework achieves state-of-the-art performance on the Tianchi fabric defect dataset, outperforming existing advanced methods. Moreover, we provided typical failure cases from our method, aiming to point out further research directions. Furthermore, we annotated the Tianchi fabric defect dataset with instance-level semantic segmentation and published it online, which is a key contribution in the field of fabric detection.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant number KYCX20_1891).

ORCID iD

Zhengyang Lu https://orcid.org/0000-0002-1540-0678

References

1. Xu X, Cao D, Zhou Y, et al. Application of neural network algorithm in fault diagnosis of mechanical intelligence. *Mech Syst Sign Proc* 2020; 141: 106625.
2. Yu G, Wang S, Cai Z, et al. Cloze test helps: effective video anomaly detection via learning to complete video

events. In: *proceedings of the 28th ACM international conference on multimedia*, October 12-16, 2020, Association for Computing Machinery, New York, USA, pp.583–591.
3. Tianchi. Smart diagnosis of cloth flaw dataset. 2020.
4. Kumar PS and Hafedh H. Detection of defects in knitted fabric images using Eigen values. *Int J Comput Sci Eng* 2013; 2: 7–10.
5. Song L, Li R and Chen S. Fabric defect detection based on membership degree of regions. *IEEE Acc* 2020; 8: 48752–48760.
6. Ben Gharsallah M and Ben Braiek E. A visual attention system based anisotropic diffusion method for an effective textile defect detection. *J Text Inst* 2020; 112: 1–15.
7. Ngan HYT, Pang GKH, Yung S-P, et al. Wavelet based methods on patterned fabric defect detection. *Patt Recognit* 2005; 38: 559–576.
8. Abouelela A, Abbas HM, Eldeeb H, et al. Automated vision system for localizing structural defects in textile fabrics. *Patt Recognit Lett* 2005; 26: 1435–1443.
9. Li P, Zhang H, Jing J, et al. Fabric defect detection based on multi-scale wavelet transform and Gaussian mixture model method. *J Text Inst* 2015; 106: 587–592.
10. Rebhi A, Benmhammed I, Abid S, et al. Fabric defect detection using local homogeneity analysis and neural network. *J Photon* 2015; 2015: 1–9. Hindawi.
11. Ngan HYT, Pang GKH and Yung NHC. Ellipsoidal decision regions for motif-based patterned fabric defect detection. *Patt Recognit* 2010; 43: 2132–2144.
12. Bissi L, Baruffa G, Placidi P, et al. Automated defect detection in uniform and structured fabrics using Gabor filters and PCA. *J Vis Commun Image Represent* 2013; 24: 838–845.
13. Schulz-Mirbach H. *Ein referenzdatensatz zur evaluierung von sichtprüfungsverfahren für textiloberflächen*. Hamburg-Harburg: TU, 1996.
14. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. In: *European conference on computer vision*, October 8-16, 2016, Amsterdam, Netherlands, pp.21–37. Springer.
15. Liu Z, Liu S, Li C, et al. Fabric defects detection based on SSD. In: *proceedings of the 2nd international conference on graphics and signal processing*, October 6-8, 2018, Association for Computing Machinery, New York, USA, pp.74–78. Sydney, Australia.
16. Ouyang W, Xu B, Hou J, et al. Fabric defect detection using activation layer embedded convolutional neural network. *IEEE Acc* 2019; 7: 70130–70140.
17. Li Y, Zhang D and Lee D-J. Automatic fabric defect detection with a wide-and-compact network. *Neurocomputing* 2019; 329: 329–338.
18. Zhou T, Zhang J, Su H, et al. EDDs: a series of Efficient Defect Detectors for fabric quality inspection. *Measurement* 2021; 172: 108885.
19. Jing J, Wang Z, Rätsch M, et al. Mobile-Unet: an efficient convolutional neural network for fabric defect detection. *Text Res J* 2022; 92: 30–42.
20. Girshick R. Fast R-CNN. In: *proceedings of the IEEE international conference on computer vision*, December 7-13, 2015, Santiago, Chile, pp.1440–1448. IEEE, New York, USA.
21. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *proceedings of the IEEE conference on computer vision and pattern recognition*, June 24-27, 2014, Columbus, Ohio, pp.580–587. IEEE, New York, USA.
22. Wu J, Le J, Xiao Z, et al. Automatic fabric defect detection using a wide-and-light network. *Appl Intell* 2021; 51: 4945–4961.
23. Li F and Li F. Bag of tricks for fabric defect detection based on Cascade R-CNN. *Text Res J* 2021; 91: 599–612.
24. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Communications of the ACM* 2020; 63: 139–144.
25. Liu J, Wang C, Su H, et al. Multistage GAN for fabric defect detection. *IEEE Trans Image Proc* 2019; 29: 3388–3400.
26. Le X, Mei J, Zhang H, et al. A learning-based approach for surface defect detection using small image datasets. *Neurocomputing* 2020; 408: 112–120.
27. Zhou Q, Mei J, Zhang Q, et al. Semi-supervised fabric defect detection based on image reconstruction and density estimation. *Text Res J* 2021; 91: 962–972.
28. Dlamini S, Kao C-Y, Su S-L, et al. Development of a real-time machine vision system for functional textile fabric defect detection using a deep YOLOv4 model. *Text Res J* 2022; 92: 675–690.
29. Bochkovskiy A, Wang C-Y and Liao H-YM. Yolov4: optimal speed and accuracy of object detection. *arXiv preprint arXiv:200410934* 2020.
30. Redmon J and Farhadi A. Yolov3: an incremental improvement. *arXiv preprint arXiv:180402767* 2018.
31. Cai Z and Vasconcelos N. Cascade R-CNN: delving into high quality object detection. In: *proceedings of the IEEE conference on computer vision and pattern recognition*, June 19-21, 2018, Salt Lake City, UT, USA, pp.6154–6162. IEEE, New York, USA.
32. Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Proc* 2004; 13: 600–612.
33. Lin T-Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context. *In: European conference on computer vision*, September 6-12, 2014, Zurich, Switzerland, pp.740–755. Springer.
34. Ronneberger O, Fischer P and Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *international conference on medical image computing and computer-assisted intervention*, October 5-9, 2015, Munich, Germany, pp.234–241. Springer.
35. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *proceedings of the IEEE conference on computer vision and pattern recognition*, June 27-30, 2016, Las Vegas, NV, USA, pp.770–778. IEEE, New York, USA.
36. Amari S-i. Backpropagation and stochastic gradient descent method. *Neurocomputing* 1993; 5: 185–196.