processes

MDPI

Article

Research on a Lightweight Textile Defect Detection Algorithm Based on WSF-RTDETR

Jun Chen [1,*], Shubo Zhang [1], Yingying Yang [2], Weiqian Li [1] and Gangfeng Wang [3,*]

[1] Shaanxi Key Laboratory of Clothing Intelligence and State-Province Joint Engineering, Research Center of Advanced Networking and Intelligent Information Services, School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China; zhangshubo@stu.xpu.edu.cn (S.Z.); wqli@xpu.edu.cn (W.L.)
[2] School of Economics and Statistics, Xingzhi College of Xi'an University of Finance and Economics, Xi'an 710038, China; yangyyxa@outlook.com
[3] Key Laboratory of Road Construction Technology and Equipment of MOE, School of Construction Machinery, Chang'an University, Xi'an 710064, China
* Correspondence: chenjun@xpu.edu.cn (J.C.); wanggf@chd.edu.cn (G.W.)

Abstract

Textile defect detection technology has become a core component of industrial quality control. With the advancement of artificial intelligence technologies, an increasing number of intelligent recognition methods are being actively researched and deployed in the textile defect detection. To further improve detection accuracy and quality, we propose a new lightweight process named WSF-RTDETR with reduced computational resources. Firstly, we integrated WTConv convolution with residual blocks to form a lightweight WTConv-Block module, which could enhance the capability of capturing detailed features of tiny defective targets while reducing computational overhead. Subsequently, a lightweight slimneck-SSFF feature fusion architecture was constructed to enhance the feature fusion performance. In addition, the Focaler–MPDIoU loss function was presented by incorporating dynamic weight adjustment and multi-scale perception mechanism, which could improve the detection accuracy and convergence speed for tiny defective targets. Finally, we conducted experiments on a textile defect dataset to further validate the effectiveness of the WSF-RTDETR model. The results demonstrate that the model improves mean average precision (mAP50) by 4.71% while reducing GFLOPs and the number of parameters by 24.39% and 31.11%, respectively. The improvements in both detection performance and computational efficiency would provide an effective and reliable solution for industrial textile defect detection.

Keywords: textile defect; WSF-RTDETR; lightweight structure; slimneck; Focaler–MPDIoU; tiny defective targets

check for updates

Academic Editor: Yo-Ping Huang

Received: 13 August 2025
Revised: 2 September 2025
Accepted: 4 September 2025
Published: 5 September 2025

Citation: Chen, J.; Zhang, S.; Yang, Y.; Li, W.; Wang, G. Research on a Lightweight Textile Defect Detection Algorithm Based on WSF-RTDETR. *Processes* **2025**, *13*, 2851. https://doi.org/10.3390/pr13092851

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Textile defect detection [1] plays a crucial role in textile production and quality control. It is widely applied in product inspection, automated sorting, and intelligent manufacturing processes. In product inspection, real-time monitoring and sensing technologies are used to detect defects during weaving, dyeing, and finishing processes: in automated sorting, big data and barcode recognition technologies enable rapid retrieval of product information; in intelligent manufacturing, Internet of Things technologies allow remote monitoring of inspection equipment and intelligent warning systems. The core task involves accurately locating and classifying defects in fabric images, such as warp breaks, weft breaks, color

differences, and stains. However, textile defects are characterized by tiny size, multi-scale, and dense distribution, which still poses great challenges for multi-scene-oriented textile defect detection [2].

Traditional textile defect detection methods primarily rely on manual visual inspection and conventional image processing techniques [3]. However, with the diversification of textile materials and the increasing demand for higher production efficiency and product quality, these methods face several limitations. They are often time-consuming and highly subjective, and their performance is poor when detecting tiny, densely distributed, or multi-scale defects. In addition, misdetections and missed detections become more pronounced when dealing with complex patterns, colored fabrics, or varying lighting conditions [4]. These limitations make it difficult to meet the modern textile industry's demand for efficient and precise inspection.

With the rapid development of deep learning, detection methods based on the deep convolutional neural network (CNN) [5] have gradually made progress. There are usually two main types of methods for textile defect detection: one-stage detection methods and two-stage detection methods. One-stage methods predict the location and class of the target directly on the image, which has higher detection speed. Common one-stage detection methods include SSD [6] and YOLO series [7]. In contrast, two-stage detection methods firstly generate candidate regions by region proposition network (RPN) and then perform classification and bounding box regression. Common two-stage detection methods include R-CNN [8], Faster R-CNN [9], and Mask R-CNN [10]. Although CNN has demonstrated promising accuracy and speed in textile defect detection, they inherently rely on local convolutional kernels to capture local features, which leads to limitations in modeling global context and remote dependencies.

In recent years, the self-attention mechanism has gradually gained widespread attention, which could overcome the limitations of short-range dependencies in CNN. Transformer [11] can directly capture long-range dependencies and global information through the self-attention mechanism, enabling more flexible modeling of complex relationships. In 2020, the Facebook AI Research team first proposed DETR [12], achieving a fully end-to-end object detection framework. It treats the object detection task as a set prediction problem, using a Transformer encoder–decoder architecture to directly predict object bounding boxes and categories. It abandons traditional methods such as anchor boxes and NMS post-processing. By leveraging a global self-attention mechanism, the model can capture global information in the image, enabling better understanding of the scene context [13]. In 2023, Zhao et al. proposed the RT-DETR detection model [14], which first employs a convolutional backbone to extract multi-scale features, followed by feature fusion to preserve both fine-grained spatial details and high-level semantic information. The fused features are then fed into a Transformer encoder–decoder, where object queries interact with the encoded features via self-attention to directly predict bounding boxes and object categories. While retaining the end-to-end advantages of the Transformer, it significantly improves detection speed and accuracy for small targets, making it a representative method for efficient real-time detection.

Although the RT-DETR model performs well in efficient end-to-end detection, it still faces challenges in textile defect detection scenarios characterized by densely distributed small targets with significant scale variations. While RT-DETR can detect defects in dense fabrics, its backbone mainly relies on standard convolutions and residual structures, which may limit its ability to extract fine-grained features for tiny targets. Meanwhile, its feature fusion strategy has certain limitations in integrating multi-scale information, potentially reducing the detection efficiency and accuracy for dense small defects. In addition, existing IoU-based bounding box regression loss functions mostly use fixed weights and lack

dynamic adaptability to targets of different scales, thereby affecting the convergence speed and accuracy of tiny target regression. Aiming at the above problems, this paper proposes a WSF-RTDETR model based on the RT-DETR model to improve the performance of target detection in textile defect detection scenarios. Our main contributions are as follows:

1. The paper combined WTConv convolution with residual blocks to form a lightweight WTConv-Block module, which could improve the feature extraction capability while maintaining low computational complexity;
2. The lightweight slimneck-SSFF architecture was constructed by integrating the Sequence Feature Fusion (SSFF), GSConv and VoVGSCSP, which could promote the detection efficiency of tiny defects while reducing computational cost;
3. With incorporating dynamic weighting adjustment and multi-scale perception mechanism, the Focaler–MPDIoU loss function was proposed to balance the regression errors of targets in the different scales, thereby improving the detection accuracy.

The rest of the paper is organized as follows: Section 2 introduces related work in the field of textile defect detection; Section 3 outlines the framework and details of the proposed WSF-RTDETR model; Section 4 describes the experimental setup and analysis of the results; Section 5 concludes the paper.

2. Related Work

2.1. Traditional Algorithms for Textile Defect Detection

Traditional target detection algorithms are mainly divided into three stages, which are candidate region generation, feature extraction, and target classification and identification. Firstly, candidate regions are generated by methods such as selective search. Then, the visual features of the target are extracted using manually designed feature extraction techniques such as HOG [15] and SIFT (Scale Invariant Feature Transform) [16]. Finally, the extracted features are inputted into an SVM (Support Vector Machine) [17] classifier for target recognition. However, the performance of these methods is not satisfactory in practical applications, and it is difficult to achieve the desired results in complex image scenes.

2.2. Deep Learning Methods for Textile Defect Detection

With CNNs demonstrating excellent performance in tasks such as image segmentation [18] and object detection [19], they have also been widely applied in textile defect detection. For example, Ouyang et al. [20] proposed a convolutional neural network based on activation layer embedding for fabric defect detection, which enhanced the feature extraction capability for tiny multi-scale defects and reduced the false detection rate caused by complex texture interference. Chen et al. [21] developed an improved Faster R-CNN method based on Gabor filters and genetic algorithm optimization to optimize its parameters using genetic algorithms, thereby improving the accuracy and efficiency of textile defect detection.

To improve the real-time performance of textile defect detection, many researchers have proposed single-stage detection methods. H. Xie et al. [22] proposed an improved textile defect detection method based on SSD, introducing multi-scale feature fusion and local enhancement strategies, which significantly improved detection accuracy and real-time performance. Liu et al. [23] proposed a lightweight model for fabric defect detection based on YOLOv7, incorporating new convolution operators and an expanded efficient layer aggregation network to enhance feature extraction efficiency. They combined receptive field blocks and content-aware up-sampling to strengthen feature fusion and context awareness, used the HardSwish activation function to accelerate convergence, and optimized bounding box regression with the Wise-IOU v3 loss function, while employing data augmentation techniques to improve generalization. Li et al. [24] proposed an enhanced fabric defect

detection network based on YOLOv8 by adding multi-scale detection channels that preserve low-level texture features to improve the detection of tiny defects. They introduced the Powerful-IoU adaptive anchor guidance mechanism to enhance localization accuracy for defects with extreme aspect ratios and used the SlideLoss loss function to effectively alleviate the issue of imbalanced easy and hard sample distributions. Dau et al. [25] proposed an automatic fabric defect classification system that integrated advanced optical imaging with deep learning. By optimizing dark-field and backlight imaging techniques and combining diverse deep learning detection models, the system achieved high-accuracy, real-time detection, and classification of small, blurry, and overlapping fabric defects.

In recent years, with the rapid development of the transformer model, it has also been studied more deeply in the field of textile defect detection. Zhang et al. [26] proposed an unsupervised fabric defect detection method based on a U-shaped Swin Transformer with a quadtree attention framework, which achieved efficient feature learning and defect localization by constructing a U-shaped codec architecture for multi-scale feature interaction. Li et al. [27] proposed a multi-scale insulator defect detection method based on DETR, introducing self-attention up-sampling (SAU) and the defect IDIoU loss function to enhance the detection of small targets. Zhang et al. [28] proposed a surface defect detection algorithm framework that enhanced Transformer sensitivity to input positional information and scale variations, which significantly improved object recognition performance. Fang et al. [29] proposed an RT-DETR fabric defect detection method based on the introduction of a noise suppression module and a dynamic snake convolutional encoder as well as a hybrid convolutional module, thus significantly improving the detection accuracy and real-time performance in filtering complex background noise and enhancing detailed feature extraction.

In summary, these methods have leveraged deep learning networks to achieve precise detection and localization of textile defects. However, in scenarios with complex backgrounds (such as colored or patterned fabrics) and small target defects (such as loosely woven fabrics with larger apertures or small fabric fragments), the stability of these algorithms still needs improvement to further reduce missed and false detections. Therefore, this paper fully exploits the advantages of existing models and proposes the WSF-RTDETR method. By enhancing fine-grained feature extraction and multi-scale feature fusion capabilities, it can more accurately locate and recognize tiny defects in fabrics, thereby further improving the accuracy and efficiency of textile defect detection.

3. Methods

In this section, we elaborate in detail on the proposed textile defect detection method, including the architecture of the WSF-RTDETR network, improvements to the feature extraction network, cross-scale feature fusion, and optimization of the loss function.

3.1. WSF-RTDETR Model Architecture

To resolve the challenges of high-precision and real-time processing in textile defect detection with complex backgrounds and small defect scenarios, we propose a lightweight WSF-RTDETR textile defect detection model. The model mainly consists of image preprocessing, backbone network, neck network, and transformer decoder with auxiliary prediction head. The details are shown in Figure 1.

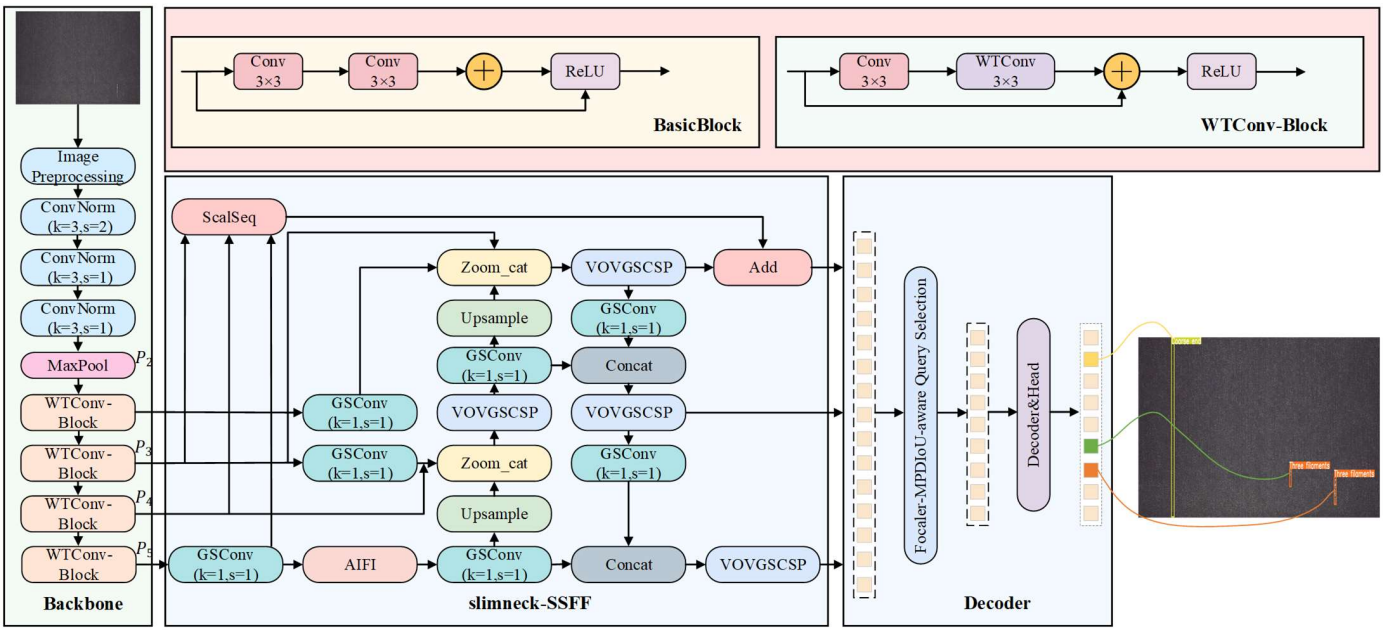


Figure 1. Network architecture of WSF-RTDETR.

Firstly, to improve the detection effect of defect information and feature extraction efficiency, the input image is preprocessed using Gaussian filtering, binarization, and gray scale scaling operations.

Secondly, in the backbone network, this paper constructs the WTConv-Block module based on WTConv [30] with residual structure to optimize the traditional convolutional operation, so as to be able to effectively expand the model sensory field with reduced parameters, and better extract the global information of the image. Meanwhile, this module can further improve the feature extraction efficiency of the backbone network and reduce the model computational overhead.

Then, the slimneck-SSFF feature fusion module is constructed to combine the scale sequence feature fusion framework [31] with the slimneck design in the neck network, and the GSConv and VoVGSCSP modules [32] are used to enhance the identification of small targets and the overall accelerated inference process.

Finally, the Focaler–MPDIoU module constructed based on Focaler-IoU [33] with MPDIoU [34] is used for transformer decoder with auxiliary prediction header to achieve optimization of the loss function, which could improve the efficiency of bounding box regression.

These efficient feature extraction methods and lightweight design not only improve the detection of textile defects in small targets but also optimize the computational performance to enable efficient operation on resource-constrained hardware.

3.2. Improvement of Feature Extraction Network

In the fabric defect detection task, complex models are often accompanied by additional computational redundancy when dealing with simple tasks, which may lead to a decrease in detection speed. Thus, in this paper, we adopt the lightweight ResNet-18 [35] as the baseline of the backbone network. In addition, we introduce WTConv into the original residual blocks in order to form a WTConv-Block module that could enhance the backbone network. This approach not only improves feature extraction but also expands the receptive field by utilizing the wavelet transform [36], while reducing the computational and memory overhead of the model to make the model suitable for deployment on edge devices.

WTConv under two levels of WT transformation is illustrated in Figure 2. WTConv decomposes the input by wavelet transform and divides it into low-frequency and high-frequency sub-bands to achieve independent modeling of different frequency information. The input signal is filtered and down-sampled using the wavelet transform to extract the feature representations at different frequencies, and small-scale deep convolution is applied to each sub-band separately to enhance the perception of local features. Finally, the processed features are reconstructed by Inverse Wavelet Transform (IWT) to generate the final output. The structure of the process can be described by the following formula:

$$Y = IWT(Conv(W, WT(X))) \quad (1)$$

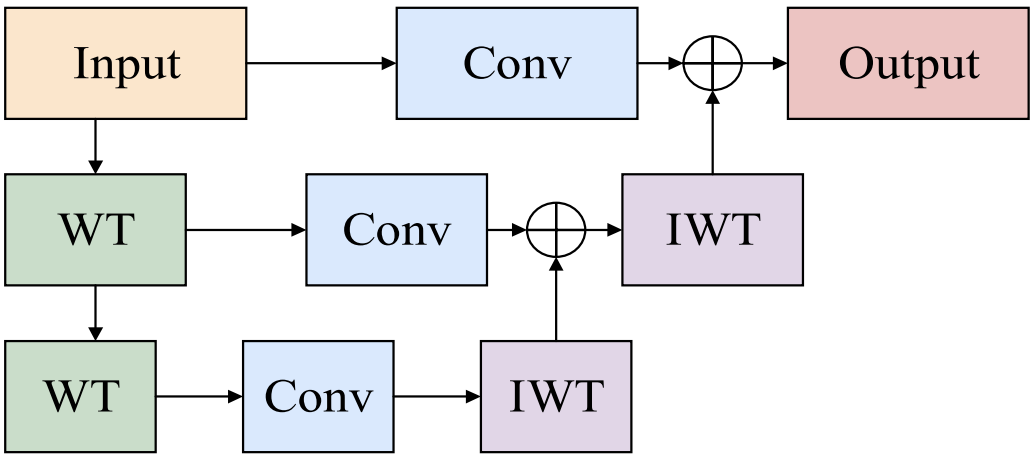


Figure 2. Example of WTConv operation using two-stage wavelet decomposition on a single channel.

In the formula, X denotes the input tensor and W denotes the convolution kernel weights. $Conv$, WT, and IWT denote the convolution operation, wavelet transform, and its inverse function, respectively.

To provide a more intuitive illustration of the specific implementation and structural layout of WTConv within the backbone network, we integrate the WTConv into the residual blocks of the feature extraction network. This design not only simplifies the backbone architecture, but also efficiently enhances the model's inference speed. Figure 3 displays the backbone structure of the proposed WTConv-Block module.

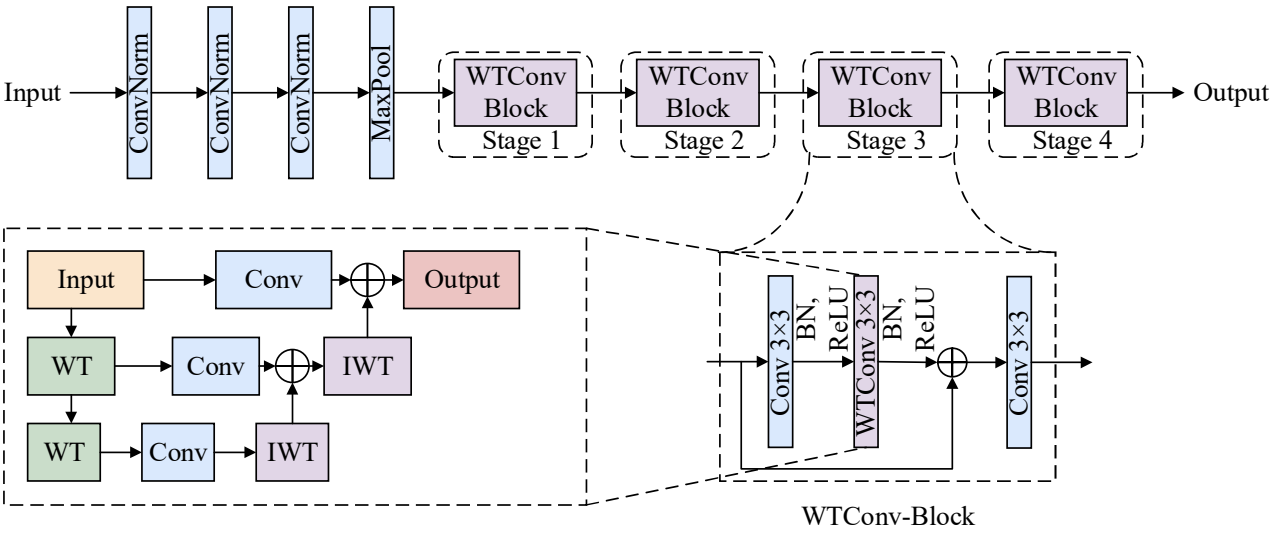


Figure 3. Lightweight feature extraction backbone network structure incorporating WTConv.

3.3. Optimization of Cross-Scale Feature Fusion Network

The CCFM in the RT-DETR neck enhances the detection of multi-scale targets through cross-scale feature fusion, which improves the expressiveness of the model for targets at different scales. However, CCFM may bring larger model parameters and computational effort, which affects the detection accuracy of small target imperfections.

To address these issues, we constructed a new slimneck-SSFF feature fusion module that introduces a GSConv module at the neck to enhance the feature fusion capability by employing GSConv convolution to receive feature mapping from the backbone network.

The VoVGSCSP module continues to be introduced on top of GSConv to enable the up-sampling and down-sampling operations of image features. Moreover, the SSFF module is used to improve the ability of the network to capture different proportions of features. The GSConv module is shown in Figure 4.

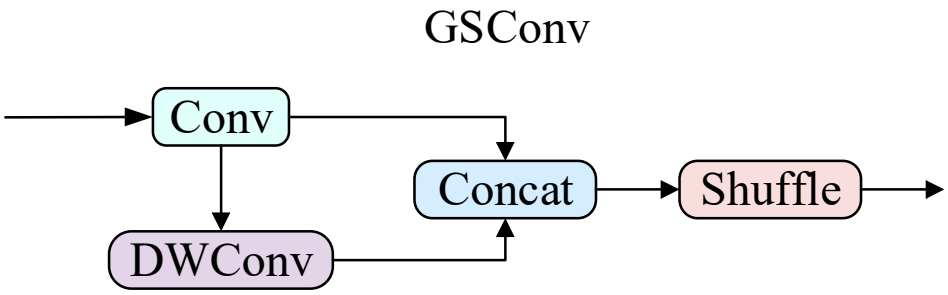


Figure 4. Structure of the GSConv module.

GSConv reduces the amount of computation by dividing the feature extraction module, while using multi-channel dense convolution to preserve the hidden connections among the channels in order to reduce the loss of semantic information. After combining the information generated by depthwise convolution(DWConv) [37] and standard convolution, GSConv uses Shuffle operation to exchange information between different channels and finally fuses the features to obtain the output results.

VoVGSCSP has achieved high stability with less redundant information after the feature mapping has been processed by GSConv. VoVGSCSP takes full advantage of GSConv for up-sampling and down-sampling, which subtly facilitates the transfer of strong semantic features, thus reducing model parameters and floating-point computation. The VoVGSCSP module is shown in Figure 5.

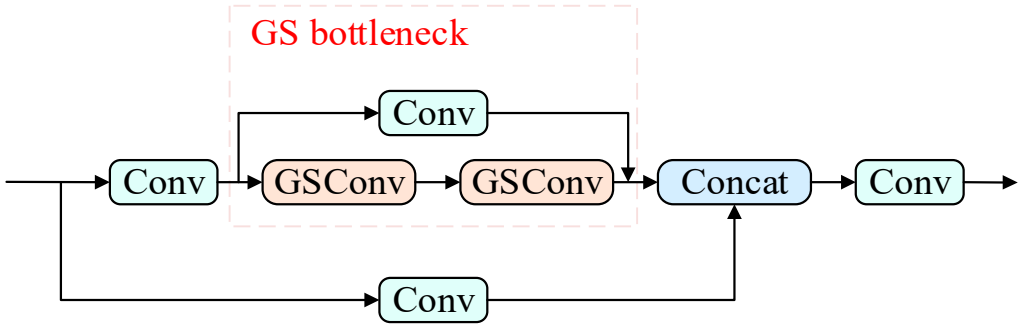


Figure 5. Structure of VoVGSCSP module.

To accurately detect textile imperfections of different sizes, we adopt SSFF module to better capture relationships across different scales. The feature representation capability is promoted by optimizing the multi-scale feature fusion so as to improve the target detection performance. The structure of SSFF module is shown in Figure 6.

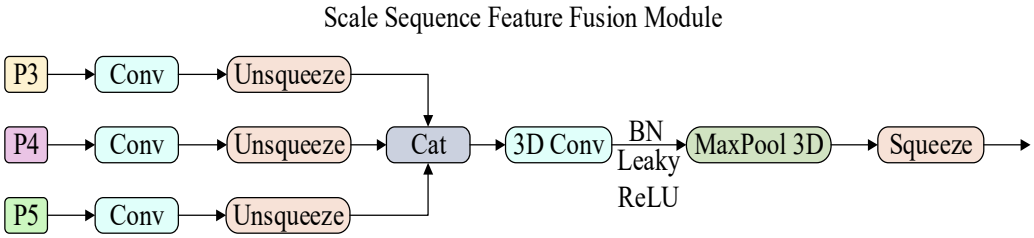


Figure 6. The Scale Sequence Feature Fusion (SSFF) module structure diagram.

3.4. Optimization of the Loss Function

The RT-DETR model utilizes the GIoU [38] loss function to optimize bounding box regression to solve the problems of vanishing gradient caused by the traditional IoU when there is no overlap; however, with imbalanced simple and complex samples, it still struggles to achieve precise localization, consequently degrading regression performance. In order to improve the detection accuracy of dense and small target defects, this paper introduced the concepts of Focaler-IoU and MPDIoU loss. Afterwards, the Focaler−MPDIoU is proposed.

Focaler-IoU uses dynamically tuned parameters to mitigate textile sample imbalance and enhance detection accuracy for specific sample types. MPDIoU directly optimizes the geometric alignment of the bounding box by minimizing the distance between the predicted box and the vertices of the real box, improving localization accuracy and accelerating convergence. In this way, the model could be able to detect textile imperfections of smaller size or lower contrast more effectively while maintaining a high level of detection and localization accuracy against complex texture backgrounds. The formulation of MPDIoU is as follows:

$$MPDIoU = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \tag{2}$$

$$L_{MPDIoU} = 1 - MPDIoU = 1 + \frac{d_1^2}{d^2} + \frac{d_2^2}{d^2} \tag{3}$$

where A denotes the real box, B denotes the predicted box, w and h denote the width and height of A and B, as well as d_1^2 denotes the squared distance from the upper-left corner of B to A, and d_2^2 denotes the squared distance from the bottom-right corner of B to A.

The Focaler-IoU formula is as follows:

$$IoU^{focaler} = \begin{cases} 0, & IoU < d \\ \frac{IoU-d}{u-d}, & d \ll IoU \ll u \\ 1, & IoU > u \end{cases} \tag{4}$$

Through adjusting the parameters d and u, Focaler-IoU can focus on a variety of regression samples, whose loss function formula is as follows:

$$L_{Focaler\text{-}IoU} = 1 - IoU^{focaler} \tag{5}$$

In an effort to further improve the accuracy of the model, we combine Focaler-IoU and MPDIoU to make it more accurate to focus on different samples, whose loss function formula is as follows:

$$L_{Focaler-MPDIoU} = L_{MPDIoU} + IoU - IoU^{Focaler} \tag{6}$$

4. Experimental Results and Analyses

4.1. Experimental Environment

To verify the validity of the proposed method, the whole experiment is implemented on an experimental platform with Windows 10.0 operating system, PyTorch 1.13.1 deep

learning framework. The RT-DETR is used as the benchmark model. The experimental environment configuration is shown in Table 1.

Table 1. Experimental environment.

Environmental Parameters	Value
Operating system	Windows 10.0
Deep Learning Framework	PyTorch 1.13.1
Programming language	Python 3.8.0
CPU	Intel(R) Core(TM) i5-11400F
GPU	GeForce RTX 3060

The same hyperparameters were maintained throughout the training phase of the experiment, and the parameter settings are shown in Table 2.

Table 2. Experimental parameters.

Parameters	Value
Learning Rate	0.0001
Image Size	640 × 640
Momentum	0.9
Optimizer Type	Adamw
Weight Decay	0.0001
Batch Size	4
Epoch	100

4.2. Datasets

In this study, the public dataset of Ali Tianchi fabric defect detection (https://tianchi.aliyun.com/dataset/79336, accessed on 22 November 2024) is used for experiments to verify the effectiveness of the proposed method in practical applications. From this dataset, we screened and processed eight types of defective, including spots, namely, hole, stain, three filaments, skipping, rough end, stacked end, pulp spot, and floating, which contains a total of 2420 images and the resolution of each image is 2446 × 1000 pixels. This dataset includes two types of materials, plain fabrics, and patterned fabrics, and all defects occur locally. The number of images in the dataset is shown in Figure 7.

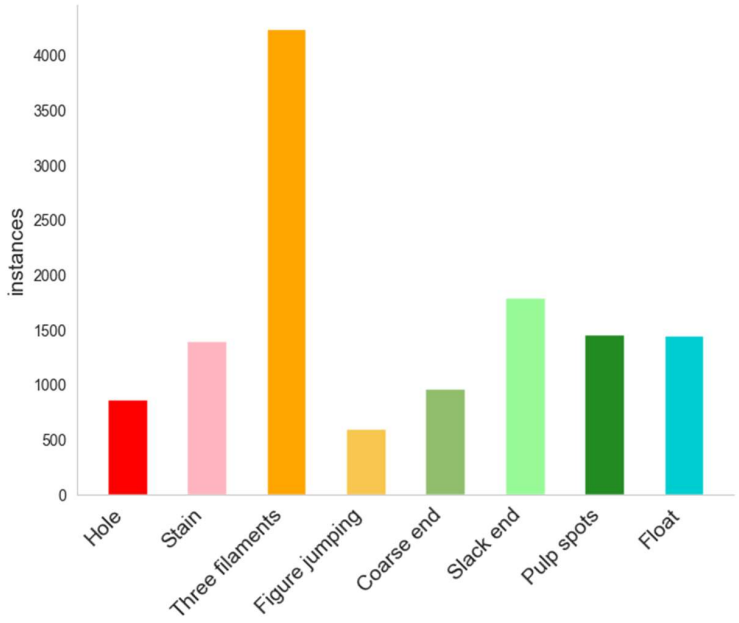


Figure 7. Number of images in the dataset.

To improve the generalization ability of the model and simulate various detection scenarios (shooting angle, light, and noise interference scenarios, etc.), we expanded the original dataset to 12,100 sheets by data enhancement such as flipping, adjusting the brightness, adding noise, etc., and then which were divided according to the ratio of 8:2 between the training set and the validation set. The number of labels for each type of textile defect is shown in Table 3.

Table 3. Number of defects in various types of textiles.

Number	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Classifications	Hole	Stain	Three Filaments	Figure jumping	Coarse end	Stack end	Pulp spots	Float
Quantities	876	1409	4251	606	973	1796	1462	1454

4.3. Evaluation Indicators

To evaluate the detection performance of the models, this study compares the detection differences between the models before and after the improvement under the same experimental setup. The evaluation indicators used include Precision (P), Recall (R), Mean Average Precision (mAP), F1 − scores, and GFLOPs.

Precision indicates the ratio of the number of correctly predicted positive samples to the number of all samples predicted positive. Recall indicates the proportion of all true targets that are correctly detected. Average Precision (AP) is calculated by averaging the highest precision for each category at different recall rates. Mean Average Precision (mAP) represents the average of the APs of so categories and is used to measure the overall performance of the model.F1 − scores are the reconciled average of precision and recall. GFLOPs measure the computational complexity of the model.

The formulae for the above indicators are as follows:

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$AP = \int_0^1 P(R)dR \tag{9}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{10}$$

$$F1 = \left(\frac{2}{Recall^{-1} + Precision^{-1}} \right) = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{11}$$

where TP denotes correctly predicted positive samples, FP denotes incorrectly predicted negative samples of positive samples, and FN denotes incorrectly predicted positive samples of negative samples. n denotes the total number of categories.

4.4. Comparative Experiments

To verify the superiority of the WSF-RTDETR model in textile defective target detection, we compare it with several existing representative object detectors on the Tianchi public dataset, including Faster R-CNN,SSD,YOLOv7 [39], YOLOv8 [40], YOLOv11 [41], and RT-DETR series. The results are shown in Table 4.

Table 4. Comparative experiments with different models.

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP50:95 (%)	F1 (%)	FPS (f/s)
Faster R-CNN	78.62	68.35	71.46	41.21	73.14	48
SSD	80.21	68.52	71.82	42.05	73.91	102
YOLOv7	79.64	74.76	77.68	46.27	77.12	98
YOLOv8	82.37	79.23	78.92	48.13	80.77	118
YOLOv11	81.50	78.11	78.52	47.81	79.77	120
RT-DETR	86.04	75.11	76.69	51.20	80.20	112
RT-DETR-L	84.11	77.56	78.38	49.49	80.70	110
RT-DETR-R34	81.75	74.06	72.10	48.95	77.72	115
RT-DETR-R50	86.23	77.08	77.62	51.10	80.95	108
WSF-RTDETR	89.39	77.16	80.30	52.81	81.51	128

From Table 4 and Figure 8, compared with multiple object detection methods including Faster R-CNN, SSD, YOLOv7, YOLOv8, YOLOv11, RT-DETR, RT-DETR-L, RT-DETR-R34, and RT-DETR-R50, the WSF-RTDETR model demonstrates significant improvements in mAP50, achieving performance gains of 12.37%, 11.81%, 3.37%, 1.75%, 2.27%, 4.71%, 2.45%, 11.37%, and 3.45%, respectively.

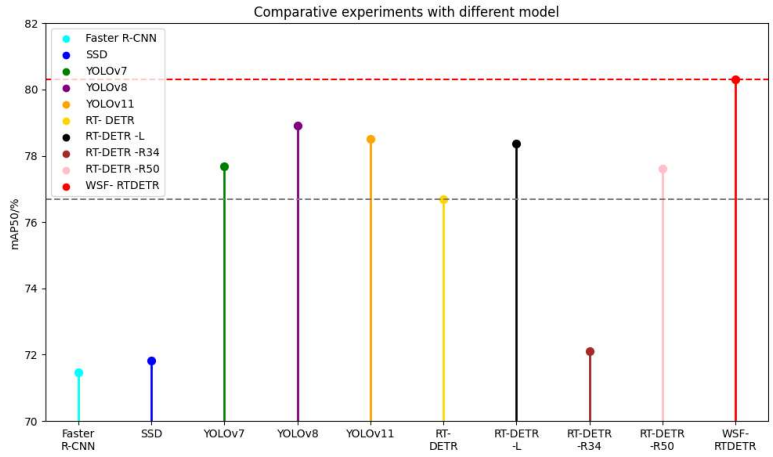


Figure 8. Comparative experiments with different models.

Faster R-CNN and SSD have high accuracy on textile defect detection tasks, but their complex processes and slow inference speeds limit their application in scenarios with high real-time requirements.

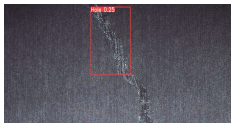
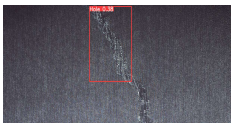
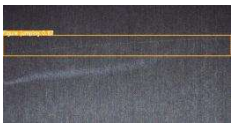
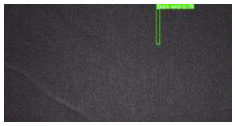
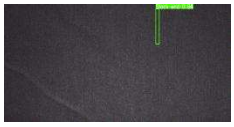
Meanwhile, YOLOv7, YOLOv8, and YOLOv11 have improved the performance of small target detection compared to the previous two algorithms, but the overall detection effect still suffers from certain misdetection and omission problems in complex backgrounds.

In RT-DETR series (RT-DETR-R34, RT-DETR-R50, RT-DETR-L), by introducing query mechanism and Transformer architecture, it improves the efficiency of target detection and small target sensing ability, and simplifies the detection process, but it is a struggle to take care of both accuracy and efficiency at the same time.

In contrast, the WSF-RTDETR model has a high accuracy of 80.3 and a good F1 score, and outperforms the above methods in several other aspects, improving the detection performance overall.

Table 5 shows some of the comparison results of textile defect detection results between the original model and the WSF-RTDETR model.

Table 5. Comparison of RT-DETR and WSF-RTDETR detection results.

Model	Defect Category			
	Hole	**Stain**	**Three Filaments**	**Figure Jumping**
RT-DETR				
WSF-RTDETR				
Model	**Defect Category**			
	Coarse End	**Slack End**	**Pulp Spots**	**Float**
RT-DETR				
WSF-RTDETR				

From Table 5, although the RT-DETR model can detect most defects in categories such as Hole, Figure Jumping, Coarse End, Slack End, and Float, the improved WSF-RTDETR model demonstrates superior accuracy in fabric defect detection. The WSF-RTDETR model leverages a lightweight feature extraction network and multi-scale feature fusion to enhance its ability to capture fine-grained features. This enables better perception of small and densely distributed defects, thereby effectively improving the performance of small target detection. Meanwhile, we show the confusion matrix [42] of the WSF-RTDETR model under the best performance in Figure 9, where the x-axis represents the real results and the y-axis represents the predicted results, which can further validate the effectiveness of the method.

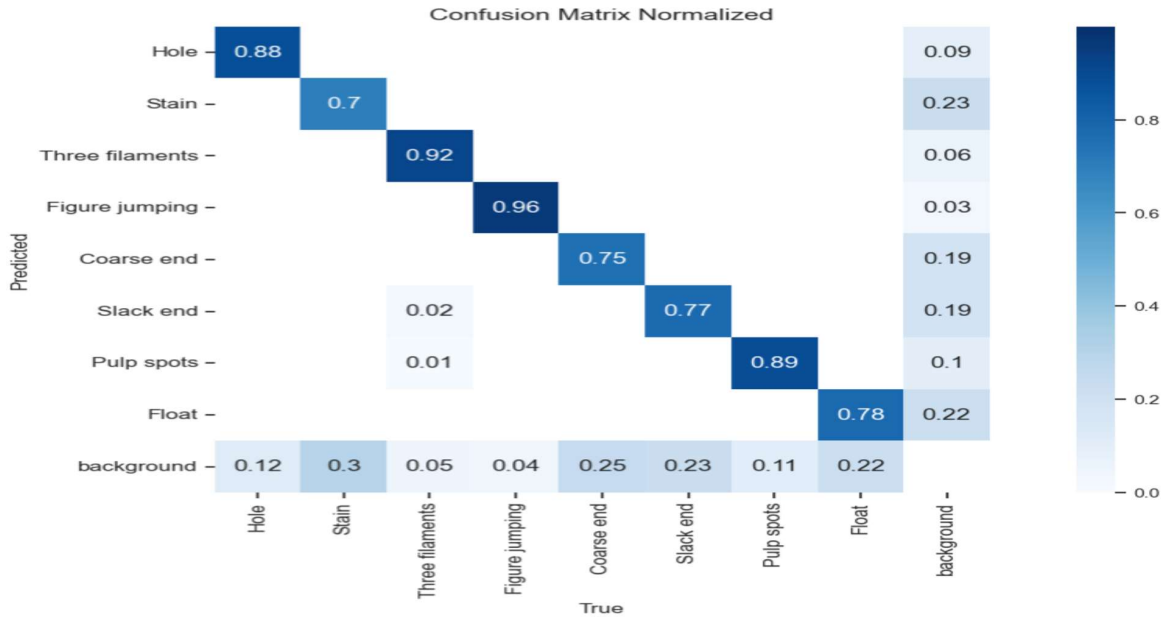


Figure 9. Confusion matrix.

As shown in Figure 9, the classification performance and easily confused categories of the WSF-RTDETR model are presented. The model achieves high recognition accuracy across most categories, with the recognition accuracies for the "Figure jumping" and "Three filaments" categories reach 96% and 92%, respectively, demonstrating the outstanding performance of the model.

In summary, it can be seen that the WSF-RTDETR model with improved loss function and combined with slimneck-SSFF structure can focus on the defective region more efficiently and improve the localization accuracy of tiny targets, while the introduction of WConv-Block achieves the lightweight design of the model. The above comparative experimental results fully verify the feasibility and advantages of the proposed method in terms of accuracy and efficiency.

4.5. Ablation Experiments

To further evaluate the performance of the proposed model, we performed ablation experiments on WSF-RTDETR using the same experimental environment, and the results are shown in Table 6.

Table 6. Results of ablation experiments.

ID	RT-DETR	WTConv-Block	Slimneck-SSFF	Focaler-MPDIoU	Params (M)	GFLOPs (G)	mAP50 (%)	mAP50:95 (%)	FPS (f/s)
1	√				19.83	57.0	76.69	51.20	112
2	√	√			12.83	40.3	77.74	51.81	122
3	√		√		19.62	57.6	78.33	52.19	115
4	√	√	√		13.66	43.1	79.27	53.47	124
5	√	√	√	√	13.66	43.1	80.30	52.81	128

In this study, five sets of ablation experiments were designed to step-by-step validate and analyze the advantages and disadvantages of each improvement method, so as to assess the effectiveness and applicability of different improvement methods more clearly.

As shown in Table 6, the first experiment is the RT-DETR baseline model, which achieves a mAP50 index of 76.69 in the textile defect detection task.

In the second experiment, by integrating a lightweight WTConv-Block module into the backbone network, the model improved by 1.37% and 1.19% on the mAP50 and mAP50:95 metrics, respectively. This improvement enhanced the model's detection accuracy and reduced the GFLOPs and number of parameters by 29.30% and 35.30%, respectively.

In the third experiment, the results indicate that the slimneck-SSFF framework enhances cross-scale feature fusion, improving mAP50 by 2.14% and mAP50:95 by 1.93% without compromising other evaluation metrics, which could validate its capability in effectively fusing and representing features of small-target defects too.

Then, in the fourth set of experiments, we adopt the baseline model combined with WTConv-Block and slimneck-SSFF module. Experimental data show that by focusing on defect localization and reducing model parameters, mAP50 increased by 3.36%, while the number of parameters decreased by 31.11%. This validates the critical role of the collaborative design between the backbone and neck, which could enhance overall detection performance.

Finally, the fifth set of experiments further improved the model accuracy effectively by combining the loss function optimization. The proposed WSF-RTDETR achieves improvements of 4.71% in mAP50 and 3.14% in mAP50:95, while reducing computational cost (GFLOPs) and model parameters by 24.39% and 31.11%, respectively. These results validate its enhanced effectiveness in textile defect detection.

In summary, it can be seen that the WSF-RTDETR improves the detection accuracy of tiny defective targets by slimneck-SSFF feature fusion and Focaler–MPDIoU, while WTConv-Block can effectively reduce the parameters for model lightweighting.

In addition, Figure 10 demonstrates the training curves of WSF-RTDETR compared with the baseline model RT-DETR. It can be found that there is not much difference in terms of precision, but WSF-RTDETR performs better in terms of recall, mAP50, and loss value, which further validates its effectiveness.

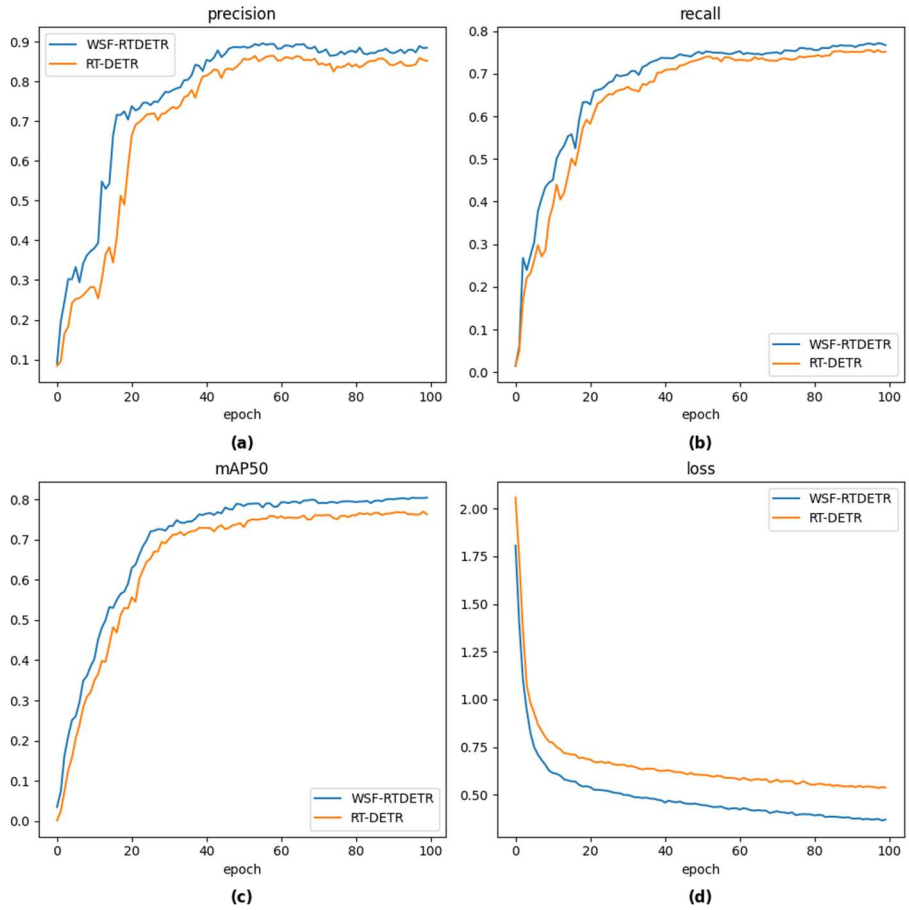


Figure 10. (**a**) Precision comparison curve; (**b**) Recall comparison curve; (**c**) mAP50 comparison curve; (**d**) Loss comparison curve.

Finally, the model proposed in this paper effectively enhances the ability to locate and identify tiny textile defects by improving cross-scale feature fusion, optimizing the loss function and lightweight design. The recognition ability of textile tiny defects is enhanced while improving the accuracy and reducing the number of parameters.

5. Conclusions

At present, the textile defect detection confronts with the challenges of multi-scale defects and complex textile backgrounds. Aiming to further enhance the accuracy and efficiency of textile defect detection, the WSF-RTDETR model is proposed in the paper. This method uses the lightweight WTConv-Block module to optimize the backbone feature extraction network so as to reduce the computational overhead and parameters of the model. Then, with combining the GSConv and VoVGSCSP modules, the slimneck-SSFF cross-scale feature fusion module is constructed to reduce computational overhead while generating more detailed semantic information. In addition, the original loss function exhibits limited sensitivity to subtle defective targets, thus the Focaler–MPDIoU loss function is presented

to enhance detection accuracy and accelerate convergence by dynamically adjusting weight allocation and the auxiliary bounding box mechanism.

To validate the performance of the WSF-RTDETR model, we conducted comparative experiments with existing detection models on Tianchi's publicly available textile defect dataset, such as the Faster R-CNN, SSD, and YOLO series. The WSF-RTDETR model exhibits significant advancements in lightweight backbone design, cross-scale feature fusion mechanisms, and the improved loss function, which together achieve higher computational efficiency and superior detection performance for multi-scale tiny defects. The results show that the WSF-RTDETR has higher detection accuracy and lower parameters. Specifically, the WSF-RTDETR model achieves improvements of 4.71% in mAP50 and 3.14% in mAP50:95, while reducing GFLOPs and parameter count by 24.39% and 31.11%.

In the future, we will continue our research in the field of textile defect detection, focusing primarily on continuous model optimization, dataset expansion, lightweight model deployment, and the impact of material size and surface area differences.

Author Contributions: J.C. and S.Z. conducted the project planning; Y.Y. was in charge of the investigation; S.Z. and J.C. performed the experimental work and data analysis; S.Z. wrote the original draft; J.C. and W.L. reviewed and edited the manuscript; G.W. provided supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2025JC-YBMS-796); in part by the Education Department of the Shaanxi Provincial Government (Program No. 23JS027); in part by the Key Laboratory of Expressway Construction Machinery of Shaanxi Province and the Fundamental Research Funds for the Central Universities, CHD (Chang'an University, No. 300102250510); and in part by the Research Foundation of Xi'an Polytechnic University (No. BS201847).

Data Availability Statement: This dataset analyzed during the current study are available in the Ali Tianchi cloth defect detection public dataset repository, https://tianchi.aliyun.com/dataset/79336 (accessed on 22 November 2024).

Acknowledgments: The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ngan, H.Y.T.; Pang, G.K.H.; Yung, N.H.C. Automated fabric defect detection—A Review. *Image Vis. Comput.* **2011**, *29*, 442–458. [CrossRef]
2. Ha, Y.-S.; Oh, M.; Pham, M.-V.; Lee, J.-S.; Kim, Y.-T. Enhancements in image quality and block detection performance for Reinforced Soil-Retaining Walls under various illuminance conditions. *Adv. Eng. Softw.* **2024**, *195*, 103713. [CrossRef]
3. Shahrabadi, S.; Castilla, Y.; Guevara, M.; Magalhães, L.G.; Gonzalez, D.; Adão, T. Defect detection in the textile industry using image-based machine learning methods: A brief review. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2022; Volume 2224, p. 012010.
4. Wei, B.; Hao, K.; Gao, L.; Tang, X.-S. Detecting textile micro-defects: A novel and efficient method based on visual gain mechanism. *Inf. Sci.* **2020**, *541*, 60–74. [CrossRef]
5. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*; Pereira, F., Burges, C.J., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; ISBN 9781627480031.
6. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37. [CrossRef]
7. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

8. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
9. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]
10. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
11. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
12. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In *European Conference on Computer Vision*; Springer International Publishing: Cham, Switzerland, 2020; pp. 213–229. [CrossRef]
13. Dang, L.M.; Wang, H.; Li, Y.; Nguyen, T.N.; Moon, H. DefectTR: End-to-end defect detection for sewage networks using a transformer. *Constr. Build. Mater.* **2022**, *325*, 126584. [CrossRef]
14. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Detrs beat yolos on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 16965–16974. [CrossRef]
15. Wei, Y.; Tian, Q.; Guo, J.; Huang, W.; Cao, J. Multi-vehicle detection algorithm through combining Harr and HOG features. *Math. Comput. Simul.* **2019**, *155*, 130–145. [CrossRef]
16. Low, D.G. Distinctive image features from scale-invariant keypoints. *J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
17. Tang, Y. Deep learning using linear support vector machines. *arXiv* **2013**, arXiv:1306.0239. [CrossRef]
18. Lüddecke, T.; Ecker, A. Image segmentation using text and image prompts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 7086–7096.
19. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 3520–3529.
20. Ouyang, W.; Xu, B.; Hou, J.; Yuan, X. Fabric defect detection using activation layer embedded convolutional neural network. *IEEE Access* **2019**, *7*, 70130–70140. [CrossRef]
21. Chen, M.; Yu, L.; Zhi, C.; Sun, R.; Zhu, S.; Gao, Z.; Ke, Z.; Zhu, M.; Zhang, Y. Improved faster R-CNN for fabric defect detection based on Gabor filter with Genetic Algorithm optimization. *Comput. Ind.* **2022**, *134*, 103551. [CrossRef]
22. Xie, H.; Zhang, Y.; Wu, Z. An improved fabric defect detection method based on SSD. *AATCC J. Res.* **2021**, *8* (Suppl. 1), 181–190. [CrossRef]
23. Liu, B.; Wang, H.; Cao, Z.; Wang, Y.; Tao, L.; Yang, J.; Zhang, K. PRC-Light YOLO: An efficient lightweight model for fabric defect detection. *Appl. Sci.* **2024**, *14*, 938. [CrossRef]
24. Li, C.; Huang, Z.; Yu, M.; Tang, S. Enhanced fabric detection network with retained low-level texture features and adaptive anchors. *J. Ind. Text.* **2025**, *55*, 15280837251330532. [CrossRef]
25. Sy, H.D.; Thi, P.D.; Gia, H.V.; An, K.L.N. Automated fabric defect classification in textile manufacturing using advanced optical and deep learning techniques. *Int. J. Adv. Manuf. Technol.* **2025**, *137*, 2963–2977. [CrossRef]
26. Zhang, H.; Xiong, W.; Lu, S.; Chen, M.; Yao, L. QA-USTNet: Yarn-dyed fabric defect detection via U-shaped Swin Transformer Network based on Quadtree Attention. *Text. Res. J.* **2023**, *93*, 3492–3508. [CrossRef]
27. Li, D.; Yang, P.; Zou, Y. Optimizing insulator defect detection with improved DETR models. *Mathematics* **2024**, *12*, 1507. [CrossRef]
28. Zhang, L.; Yan, S.-F.; Hong, J.; Xie, Q.; Zhou, F.; Ran, S.-L. An improved defect recognition framework for casting based on DETR algorithm. *J. Iron Steel Res. Int.* **2023**, *30*, 949–959. [CrossRef]
29. Fang, H.; Lin, S.; Hu, J.; Chen, J.; He, Z. CPF-DETR: An End-to-End DETR Model for Detecting Complex Patterned Fabric Defects. *Fibers Polym.* **2025**, *26*, 369–382. [CrossRef]
30. Finder, S.E.; Amoyal, R.; Treister, E.; Freifeld, O. Wavelet convolutions for large receptive fields. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2024; pp. 363–380. [CrossRef]
31. Kang, M.; Ting, C.-M.; Ting, F.F.; Phan, R.C.-W. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *Image Vis. Comput.* **2024**, *147*, 105057. [CrossRef]
32. Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhang, Z.; Ren, Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv* **2022**, arXiv:2206.02424. [CrossRef]
33. Zhang, H.; Zhang, S. Focaler-iou: More focused intersection over union loss. *arXiv* **2024**, arXiv:2401.10525. [CrossRef]
34. Ma, S.; Xu, Y. Mpdiou: A loss for efficient and accurate bounding box regression. *arXiv* **2023**, arXiv:2307.07662. [CrossRef]
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
36. Zhang, D. Wavelet transform. In *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval*; Springer International Publishing: Cham, Switzerland, 2019; pp. 35–44. [CrossRef]

37. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [CrossRef]
38. Rezatofighi, H.; Tsoi, N.; Gwak, J.Y.; Sadeghian, A.; Reid, A.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
39. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475. [CrossRef]
40. Wang, C.Y.; Yeh, I.H.; Mark Liao, H.Y. Yolov9: Learning what you want to learn using programmable gradient information. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2024; pp. 1–21. [CrossRef]
41. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
42. Room, C. Confusion matrix. *Mach. Learn* **2019**, *6*, 27.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.