FabricMamba: A fabric surface defect detection system based on large kernel attention and visual state space

Engineering Applications of Artificial Intelligence 162 (2025) 112558

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

FabricMamba: A fabric surface defect detection system based on large kernel attention and visual state space☆

Nengsheng Bao [a,b], Jiajun Lin [a,b], Yuchen Fan [c], Runxuan Bao [d], Alessandro Simeone [c,*]

[a] Department of Mechanical Engineering, Shantou University, 243 Daxue Road, Shantou, 515063, China
[b] Intelligent Manufacturing Key Laboratory of Ministry of Education, Shantou University, 243 Daxue Road, Shantou, 515063, China
[c] Department of Management and Production Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Turin, Italy
[d] Guangdong Technion Israel Institute of Technology, 241 Daxue Road, Shantou, Guangdong, 515063, China

ARTICLE INFO

Keywords:
Defect detection
Textile manufacturing
Process monitoring
You Only Look Once
State space model
Large kernel attention

ABSTRACT

In the textile manufacturing industry, fabric defect detection is essential for ensuring product quality. Traditional approaches based on Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) often encounter scalability issues, particularly due to the high computational complexity of the self-attention mechanism in ViTs. To address these limitations, this study introduces FabricMamba, a real-time defect detection framework built on the You Only Look Once version 8 (YOLOv8) CNN architecture. The model enhances detection precision and efficiency for complex fabric defects in high-resolution images while minimizing computational cost. YOLOv8 was selected as the base model due to its strong balance between accuracy and inference speed, which is critical in fast-paced textile production settings. FabricMamba extends YOLOv8 with several innovations: the Parallel Large Separable Kernel Attention (P-LSKA) mechanism for multi-scale perception, the Visual State Space Module (MVSS) for long-range dependency modeling, the lightweight DySample module for reduced resource usage, and Programmable Gradient Information (PGI) to optimize training without increasing inference complexity.

Extensive evaluations were conducted using a proprietary industrial fabric defect dataset and two public benchmarks, TILDA Textile Texture Database and FDDS Object Detection Dataset. FabricMamba achieved a mean Average Precision (mAP) of 90.0 %, 97.7 %, and 39.1 % on the respective datasets, outperforming the YOLOv8 baseline by 1.8 %, 2.3 %, and 2.0 %. Compared to Mamba-YOLO, FabricMamba reduced model size and computational requirements by 36.7 % and 33.1 %, respectively, with recall improving by 2.9 %, 1.4 %, and 4.0 %. These results confirm the model effectiveness and practical potential for industrial fabric inspection tasks.

1. Introduction

The textile industry constitutes a fundamental component of the global economy, encompassing a comprehensive sequence of processes from raw material preparation to the manufacturing of finished products. During the fabric production process, a range of surface defects, including holes, oil stains, and misweaves, may arise due to factors such as equipment degradation, operational inaccuracies, and environmental conditions. Therefore, accurate detection of fabric surface defects is essential for textile enterprises aiming to improve product quality and maintain competitiveness in the market (Guo et al., 2024).

Conventional methods for detecting cloth defects primarily rely on manual visual inspection, which is heavily dependent on the inspector experience and subjectivity (Rasheed et al., 2020). Such methods are often associated with inconsistent results and reduced detection accuracy, particularly when inspectors experience fatigue or distraction. Recent advances in computer vision technologies have driven the development of automated and intelligent systems for fabric defect detection (Bai and Jing, 2024; Zhao et al., 2025). Among these, deep learning-based object detection techniques have emerged as powerful tools, owing to their ability to extract high-level features and accurately recognize defects. Recent research in this domain has concentrated on enhancing the performance of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). For instance, (Bai and Jing, 2024), developed Mobile-Deeplab, a lightweight CNN-based model capable of real-time, end-to-end defect segmentation. Zhang et al. (2023)

☆ For more information, please refer to the relevant sections under submission guidelines for the journal in the Guide for Authors.

* Corresponding author.
E-mail address: alessandro.simeone@polito.it (A. Simeone).

https://doi.org/10.1016/j.engappai.2025.112558
Received 24 January 2025; Received in revised form 12 August 2025; Accepted 28 September 2025
Available online 3 October 2025

introduced a U-shaped Swin Transformer network utilizing a Quadtree attention mechanism for unsupervised detection of defects in yarn-dyed fabrics (Zhang et al., 2023). Additionally, Yang et al. (2024) proposed a hybrid architecture that combines the local feature extraction strengths of CNNs with the global contextual modeling capabilities of visual transformers (R. Yang et al., 2024). These developments address specific limitations inherent in each architecture: CNNs frequently struggle to capture long-range dependencies and subtle defect characteristics, while ViTs, although effective in modeling complex patterns, are constrained by high computational demands due to their quadratic complexity in relation to input size (Sampath et al., 2022).

Within this research context, the Mamba architecture, developed within the framework of State Space Models (SSMs), has received considerable attention for its capacity to manage long-range interactions with linear computational complexity (An et al., 2024). Mamba extends beyond the spatial limitations of CNNs by leveraging global receptive fields and dynamic weighting, while avoiding the computational overhead commonly associated with Transformer models. Its application in fabric defect detection offers the dual advantage of improved detection accuracy and increased processing efficiency, thereby facilitating rapid and reliable classification of complex defect types in real-world manufacturing environments. This study contributes to the advancement of fabric defect detection through the development of a novel framework, referred to as FabricMamba. This model integrates large kernel attention mechanisms with a visual state space architecture to achieve high-precision, real-time defect detection without increasing computational burden. A key component of the proposed framework is the Parallel Large Separable Kernel Attention module, which enhances the model ability to perceive variations in object scale and shape across diverse textile surfaces. Furthermore, the Multi-dimensional Visual State Space (MVSS) module enables the capture of long-range dependencies and global contextual information, thereby increasing the robustness of defect detection in complex background conditions.

The remainder of this manuscript is organized as follows. Section 2 provides a comprehensive review of recent developments in fabric defect detection using computer vision, with a focus on large-kernel attention mechanisms and state space modeling in visual applications. Section 3 presents the proposed automated fabric defect detection pipeline, including the detailed design of the FabricMamba model. Subsections 3.3 and 3.4 provide an in-depth analysis of the P-LSKA and MVSS modules, respectively. Section 4 evaluates the performance of the proposed model on an industrial dataset as well as two publicly available datasets, TILDA and FDDS, and compares the results with state-of-the-art approaches. Finally, Section 5 concludes the study with a discussion of the model performance, limitations, and implications for industrial deployment.

2. Literature review

With the rapid development of computer vision technology, the use of visual techniques for defect detection in fabric manufacturing has become a research hotspot (Ngan et al., 2011). Traditional methods, relying on predefined algorithms and features, may struggle to capture very subtle or complex patterns in images. In contrast, deep learning models, by constructing complex network structures, are capable of recognizing and processing more subtle and complex image features (Arshad and Shahzad, 2024; Jing et al., 2019). In recent years, an increasing number of researchers have adopted deep learning-based approaches for fabric defect detection. Among these, architectures based on the YOLO family have demonstrated promising results. However, alternative methods have also been explored, including the use of Faster R-CNN with attention mechanisms for high-precision detection, U-Net variants for pixel-level segmentation, and transformer-based architectures to enhance feature representation. Each of these approaches presents unique advantages and limitations in terms of detection accuracy, computational efficiency, and generalization capability. For instance (Yue et al., 2022), developed a fabric defect detection method based on YOLOv4, aiming to improve the detection of small-scale defects. The method utilized the k-means algorithm (Ikotun et al., 2023) to cluster ground truth bounding boxes and optimize anchor box selection. A convolutional block attention module was integrated into the backbone feature extractor, and the Complete Intersection over Union (CIOU) loss function was replaced with a Center-enhanced Intersection over Union (CEIOU) loss function (Zheng et al., 2022) improving both classification and localization performance.

Recent developments in YOLO-based architectures have significantly advanced the detection of small and irregular targets. Li et al. (2022) proposed an improved YOLOv4 algorithm that replaces the Mosaic augmentation with a small-target enhancement strategy and optimizes anchor box sizing, achieving a detection accuracy of 93.2 % for small objects (Q. Li et al., 2022). Tian et al. (2024) addressed class imbalance and weak feature extraction in YOLOv5 by integrating a coordinate attention mechanism into the backbone, resulting in improvements of 2 % in mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.50 (mAP50) and 3 % in mAP50–95 (Tian et al., 2024).

Peng et al. (2023) further enhanced YOLOv7 for the detection of small and irregularly shaped objects by incorporating a channel and spatial attention module, which increased mAP precision by 2.06 % compared to the baseline model (Peng et al., 2023). Mobile-YOLO, a lightweight three-stage model combining MobileNetV3-small with YOLOv8, enhances fiberglass fabric defect detection. Efficient Channel Attention replaces the SE module to reduce complexity, while Global and Similarity Attention Modules improve context modeling and feature fusion. The model achieves 99.4 % accuracy at 186 FPS, enabling real-time industrial use (Li and Kang, 2024).

Over the past decade, researchers have experimented with different kernel sizes in CNNs to improve their performance in image recognition tasks, but found that increasing the kernel size could lead to decreased performance (Chansong and Supratid, 2021; Tomen and van Gemert, 2021). In efforts to reduce the performance disparity between CNNs and hierarchical architectures such as Transformers, ConvNeXt (Todi et al., 2023) incorporated elements from Transformer designs into the ResNet framework. This approach demonstrated that moderate increases in kernel size, particularly up to 7×7, could improve performance, although further increases yielded limited additional gains. ReplkNet (Ding et al., 2022) extended this direction by employing significantly larger kernels, such as 31×31, within the MobileNetV2 framework. This adjustment expanded the effective receptive field and resulted in performance improvements in image classification and object detection tasks. Nonetheless, the elevated computational burden associated with such large kernels presents challenges for deployment in resource-sensitive applications, including image segmentation tasks. The Visual Attention Network (VAN) proposed by (Guo et al., 2023) combined CNN feature extraction with self-attention mechanisms, introducing a Large Kernel Attention (LKA) module. This module enhanced both local feature representation and the modeling of long-range dependencies. It achieved a larger receptive field through the use of dilated convolutions and included channel-wise adaptability. Despite its advantages, LKA parameter count increases quadratically with kernel size, leading to greater computational demands.

To mitigate this issue (Lau et al., 2024), introduced a Large Separable Kernel Attention (LSKA) mechanism. This strategy incorporated separable convolutional kernels, including depthwise and dilated depthwise convolutions (Wang et al., 2020), along with an attention mechanism. The LSKA design significantly reduced computational complexity while preserving the performance benefits of large receptive fields.

State Space Models (SSMs) offer a principled approach for modeling temporal dependencies and have recently gained attention in deep learning for their ability to capture long-range patterns in sequential data. Structured parameterizations have been shown to improve training efficiency and scalability, with successful applications in tasks such as speech synthesis, indicating strong potential for broader use in

sequence modeling (Kitamura and Itou, 2023). Rooted in continuous control theory and incorporating methods such as HiPPO initialization (Gu et al., 2020), SSMs offer a principled alternative to traditional sequence modeling techniques. The Linear State-Space Layer (LSSL) model highlighted the potential of SSMs for addressing long-range dependencies, though it was limited by high computational and memory demands. To mitigate these challenges, the Structured State Space for Sequence Modeling (S4) model was introduced, employing diagonal parameter structures and normalization to improve efficiency (Gu et al., 2021). Visual adaptations of State Space Models (SSMs) have been proposed to address the limitations of conventional convolutional and attention-based architectures in capturing long-range dependencies with computational efficiency. These models extend SSM principles to two-dimensional data, enabling global receptive fields with linear complexity. One example is the integration of Mamba into convolutional architectures for medical image segmentation. VMC-UNet, for instance, combines Mamba with CNNs to enhance feature representation, demonstrating improved accuracy in breast tumor segmentation tasks through effective fusion of local and global features (Wang et al., 2024).

In the context of fabric defect detection, visual recognition methods are typically based on CNNs and Vision Transformers (ViTs). Although ViTs offer improved global context modeling compared to CNNs, their practical scalability is constrained by the quadratic growth in computational complexity introduced by the self-attention mechanism as input size increases (R. Yang et al., 2024). To address this limitation, the present study adopts the high-performance and extensively validated YOLOv8 model as the foundational architecture. By leveraging the efficiency of the YOLO series in combination with the Mamba model, which offers effective long-range dependency modeling under linear computational complexity, and by incorporating large kernel attention through a streamlined architectural design, a novel real-time fabric surface defect detection model is introduced. Termed FabricMamba, this architecture achieves reduced computational overhead while preserving high detection accuracy. It enables accelerated inference and improves adaptability in complex textile environments, offering a practical and computationally efficient solution for industrial-scale fabric inspection tasks.

Recent work in real-time intelligent systems has demonstrated the utility of neural network-based methods for adaptive control and decision-making. Recurrent neural networks have improved trajectory tracking and robustness in quadrotor UAVs through model reference adaptive control and fuzzy PID architectures (Madebo et al., 2024; Wanore Madebo et al., 2024). Similarly, path planning approaches combining bidirectional rapidly-exploring random trees with adaptive Monte Carlo localization have enhanced navigation in uncertain environments (Ayalew et al., 2024). While these studies lie outside the fabric inspection domain, they share key priorities in real-time, adaptive, and resource-efficient computation systems. These priorities closely align with the computational requirements of high-performance fabric defect detection.

Existing visual inspection models such as VAN (Guo et al., 2023) and ConvNet (Liu et al., 2022) have attempted to address the performance disparity between CNNs and Transformer-based architectures. However, these approaches continue to encounter limitations in computational efficiency and feature expressiveness. FabricMamba addresses these concerns by achieving linear computational complexity, thereby overcoming the quadratic scaling challenges of ViTs. In contrast to the Large Kernel Attention (LKA) module employed in VAN, which scales quadratically with kernel size and consequently increases the number of parameters, FabricMamba employs a Parallel LSKA module. This design utilizes separable kernels to reduce computational resource requirements while maintaining feature extraction capabilities. Furthermore, while previous State Space Model (SSM)-based approaches, such as Mamba-UNet (X. Yang et al., 2024), utilize unidirectional modeling strategies, FabricMamba incorporates a Multi-directional Visual State Space Module. This module is capable of capturing defect features from multiple orientations, thereby enhancing the representation of complex fabric textures and improving defect localization. The integration of multi-directional modeling, separable attention mechanisms, and efficient architectural design collectively contributes to the robust performance and industrial applicability of the proposed model.

3. Research contributions

In light of the limitations of current methodologies and the ongoing need to improve both the accuracy and speed of real-time fabric defect detection in industrial applications, this study proposes a YOLO-Mamba model that combines the strengths of Convolutional Neural Networks (CNNs) and the Mamba State Space architecture. The primary contributions of this research are as follows:

1) Development of the FabricMamba Model: To address the challenges posed by diverse defect types, varying object scales, and complex textile backgrounds, this work introduces the FabricMamba model. By integrating large kernel attention mechanisms with the Mamba State Space framework and the YOLO detection backbone, this approach achieves real-time defect detection without increasing computational resource consumption. This integration represents the first application of large kernel attention within a visual state space model in the context of fabric defect detection, marking a significant advancement in the field.
2) Proposal of a Parallel LSKA Mechanism: To improve detection performance across varying defect scales and morphologies, a Parallel LSKA mechanism is proposed. This module enhances the Spatial Pyramid Pooling-Fast (SPPF) component by incorporating large separable convolutional kernels within a parallel architecture. The design enables the extraction of multi-scale spatial features, enhances information flow, and supports adaptive learning processes, thereby increasing the model sensitivity to irregular and fine-grained defects.
3) Introduction of the MVSS Module: In response to the difficulty of detecting defects against complex and noisy fabric backgrounds, the MVSS module is introduced. This component leverages a global receptive field and long-range dependency modeling capabilities inherent in state space models, and is further augmented with multilayer perceptrons (MLPs) to process high-dimensional feature representations. The module facilitates more robust feature extraction, enhancing the network ability to generalize across diverse textile patterns and improving performance on datasets with complex visual characteristics.

4. Methodology

4.1. Automatic defect detection system for garments

The garment manufacturing and defect detection process encompasses several key stages, from raw materials to the final product. After receiving a customer order, the factory first prepares the raw materials and conducts quality inspections to ensure that the fabric and accessories meet the required standards. The fabric is then produced through cutting and sewing processes. In the finished product stage, an automated fabric surface defect detection algorithm is used for a comprehensive visual inspection, with a focus on issues such as holes and oil stains, ensuring that the garment meets the quality standards for shipment. In this research, the proposed FabricMamba algorithm is applied during the fabric appearance inspection stage to achieve precise real-time defect detection. A complete flowchart of the automated garment production system is shown in Fig. 1.

4.2. Overall architecture of FabricMamba

The overall architecture of FabricMamba is illustrated in Fig. 2

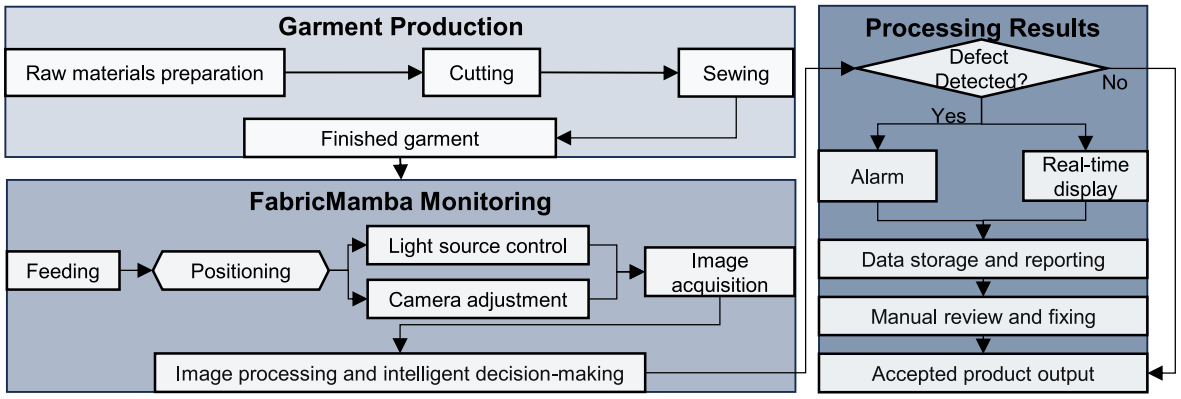

Fig. 1. Flow diagram of automatic defect detection system for garments.

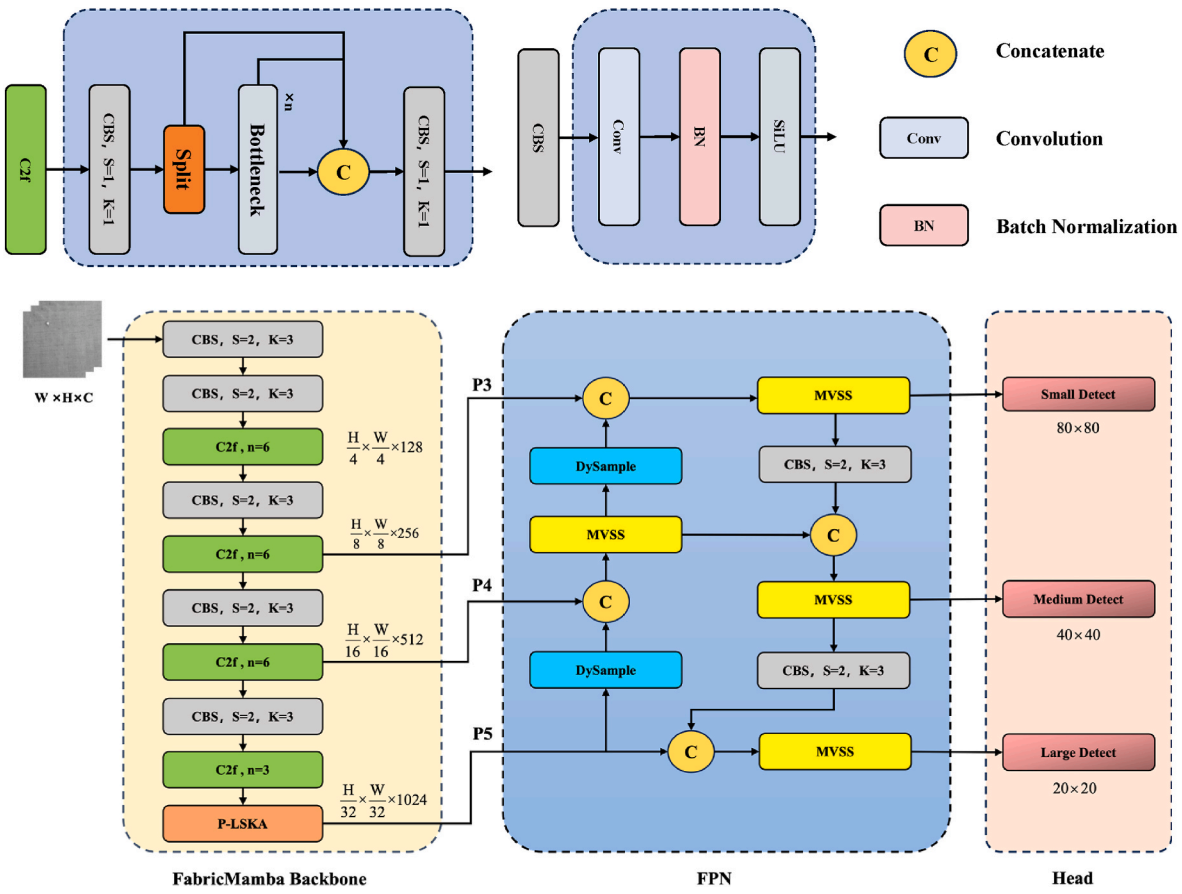

Fig. 2. Overall architecture of FabricMamba. Left: backbone for hierarchical feature extraction; middle: FPN for multi-scale feature fusion; right: detection head for defect classification and localization: top: detailed structure of the C2F and CBS modules.

primarily consisting of three core components: the FabricMamba Backbone, Feature Pyramid Networks (FPN), and the Head. The FabricMamba Backbone is mainly used for extracting preliminary features from the input image, converting the input into feature maps of three different scales that contain multi-level feature information. The YOLOv8 network utilizes the SPPF module at the end of the Backbone to further pool feature maps of different scales, thus accommodating targets of varying sizes. Based on LSKA, the FabricMamba model proposes a parallel LSKA architecture to further enhance the SPPF. LSKA captures extensive contextual information of the image through cascaded depthwise convolution and dilated depthwise convolution, subsequently generating attention maps. These attention maps are used to weight the original features, significantly enhancing the network focus on key features and boosting the overall performance of the model. Through the parallel architecture, each LSKA independently processes the multi-level outputs from the SPPF, capturing a broader range of contextual information, allowing each LSKA module to focus on extracting specific features, and then merging these features to improve the model feature representation capability.

In the FPN section, the FabricMamba model adopts the concept of PA-FPN (Jianqiang Zhang et al., 2024), achieving effective fusion between feature maps through a clever combination of upsampling and downsampling operations. Specifically, the FabricMamba model first utilizes upsampling techniques to enlarge the lower-level feature maps

to match the dimensions of the higher-level feature maps, thereby facilitating interaction between features of different scales. Concurrently, the FabricMamba model also employs downsampling methods to enrich the dimensionality and depth of feature fusion. The combination of these two branches not only enhances the expressiveness of the features but also improves the model ability to capture multi-scale information, thus maintaining detail while also capturing a broader range of contextual information. Additionally, the FabricMamba model introduces the MVSS module to replace C2f for capturing richer gradient information flow and employs a lightweight dynamic upsampler, DySample (Junyi Zhang et al., 2024), for upsampling, further reducing computational resource consumption. Through the collaborative operation of the FabricMamba Backbone and FPN, the network generates feature maps at three different scales, which are each fed into corresponding detection heads to perform classification and bounding box regression tasks, thereby effectively identifying defects in the images.

During the model training process, this research introduces the PGI module (Wang et al., 2024) to assist in training, thereby further enhancing training efficiency. During the inference stage, this module is removed to reduce computational overhead.

4.3. Parallel large kernel attention integration

In this research, to effectively address the multi-scale feature challenges in fabric defect detection, the proposed Parallel LSKA mechanism effectively combines the advantages of LSKA, aiming to significantly enhance the SPPF module perception of defects of various scales and shapes.

The Parallel Large Separable Kernel Attention (P-LSKA) module, as shown in Fig. 3, integrates the efficient attention mechanism of LSKA, applying the Parallel LSKA mechanism proposed in this study to SPPF. In this module, the input data first goes through a 1x1 convolution layer for preliminary feature extraction, keeping the channel count unchanged, and then these features are evenly split into two sub-inputs. Subsequently, the two branches of the P-LSKA module work in parallel, each combining the sub-input with the corresponding output from multi-level max pooling layers. This process ensures that each LSKA can independently focus on deep and shallow key features. Ultimately, by integrating the unique perspectives of deep and shallow features, the P-LSKA module not only enhances the expressiveness of the features but also strengthens the model ability to capture multi-scale information, thereby addressing the detection difficulties caused by the diverse sizes of fabric defects and significantly improving the overall robustness of the model.

The P-LSKA module receives an input feature map $X \in \mathbb{R}^{C_{in} \times H \times W}$ and outputs an enhanced feature map $Y \in \mathbb{R}^{C_{out} \times H \times W}$ through a series of processing steps. First, a 1x1 convolution reduces the number of channels, followed by multiple max pooling operations to generate multi-scale feature maps. These feature maps are then concatenated to form two distinct inputs, each processed by the LSKA.

In each LSKA, convolution operations use (1, 3) and (3, 1) kernels to extract features in horizontal and vertical directions, capturing edges and textures from different orientations. Subsequently, (1, 5) and (5, 1) kernels perform broader convolutions, with dilated convolutions expanding the receptive field and enhancing feature extraction capabilities. This multi-scale feature extraction method, combined with directional convolutions, effectively captures features of various scales and directions, enriching and improving feature extraction accuracy.

Finally, the outputs from the two LSKA processes are concatenated and passed through a 1x1 convolution to obtain the final output. This process effectively enhances the feature map expressive power, making it more suitable for subsequent tasks. Table 1 shows the pseudocode of the module.

4.4. Visual State Space Module

State space models can be viewed as linear time-invariant systems, where the core concept is to effectively map input stimuli $x(t) \in \mathbb{R}^L$ to output responses $y(t) \in \mathbb{R}^L$ through implicit latent states $h(t) \in \mathbb{C}^N$. The mapping relationship can be represented by a set of linear ordinary differential equations as illustrated in Eqs. (1) and (2) (Gu and Dao, 2024):

$$h'(t) = Ah(t) + Bx(t) \tag{1}$$

$$y(t) = Ch(t) + Dx(t) \tag{2}$$

where $A \in \mathbb{C}^{N \times N}$, $B, C \in \mathbb{C}^N$, and $D \in \mathbb{C}^1$ are the weighting parameter.

In the design of deep learning frameworks, models based on discrete time are commonly used to meet computational demands. This requires the transformation of continuous equations, which describe and analyze system dynamics, into discrete models. Specifically, it involves discretizing the system continuous-time ordinary differential equations

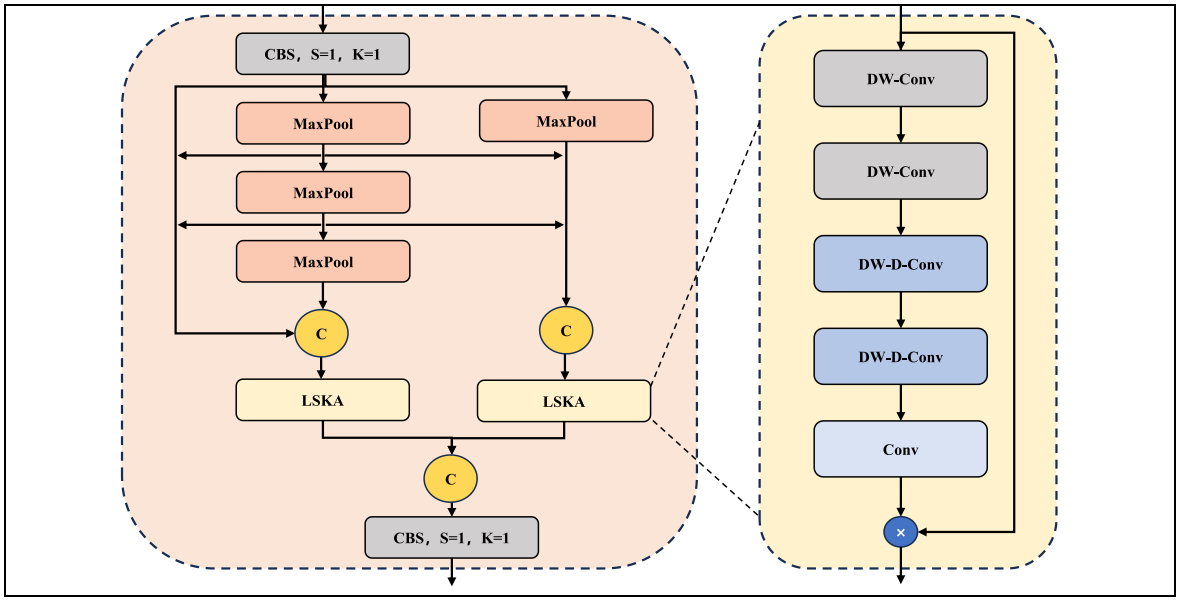

Fig. 3. Overview of the P-LSKA module. The left shows the full module structure; the right details the internal LSKA block.

Table 1
Pseudocode of the P-LSKA module.

Algorithm: P-LSKA Module	
Input: $X \in \mathbb{R}^{C_{in} \times H \times W}$	
Output: $Y \in \mathbb{R}^{C_{out} \times H \times W}$	
1: c ← C_{in} /2	▷ Step 1: Compute hidden channels
2: $x_{processed}$ ← Conv2D(X, c, 1 × 1)	▷ Step 2: 1x1 convolution to reduce channels
3: y_1 ← MaxPool2D($x_{processed}$, 5 × 5)	▷ Step 3: First max pooling
4: y_2 ← MaxPool2D(y_1, 5 × 5)	▷ Step 4: Second max pooling
5: $input_1$ ← Concat ($x_{processed}$, y_1, y_2, MaxPool2D(y_2))	▷ Step 5: Prepare first LSKA input
6: out_1 ← LSKA($input_1$, $c \times 4$)	▷ Step 6: First LSKA processing
7: $input_2$ ← Concat(($x_{processed}$, y_1, y_2, MaxPool2D($x_{processed}$))	▷ Step 7: Prepare second LSKA input
8: out_2 ← LSKA($input_2$, $c \times 4$)	▷ Step 8: Second LSKA processing
9: out ← Concat (out_1, out_2)	▷ Step 9: Concatenate the two LSKA outputs
10: Y ← Conv2D(out, C_{out}, 1 × 1)	▷ Step 10: 1x1 convolution to get final output
11: Return Y	▷ Step 11: Output the fused feature map
–	
Function: LSKA	
Input: $X \in \mathbb{R}^{dim \times H \times W}$	
Output: Attention-enhanced feature map	
1: u ← X	▷ Clone input feature map
2: attn ← DW-Conv2D(X, 1 × 3)	▷ Apply horizontal convolution
3: attn ← DW-Conv2D(attn, 3 × 1)	▷ Apply vertical convolution
4: attn ← DW-D-Conv2D(attn, 1 × 5, dilation = 2)	▷ Apply dilated horizontal convolution
5: attn ← DW-D-Conv2D(attn, 5 × 1, dilation = 2)	▷ Apply dilated vertical convolution
6: attn ← Conv2D(attn, dim, 1 × 1)	▷ Generate the attention map using a 1 × 1 convolution
7: Return u × attn	▷ Multiply the original feature map with the attention map element-wise

into an equivalent discrete-time representation. This transformation is typically achieved by applying a zero-order hold strategy to the input signal, thereby constructing a discrete-time state space model (SSM), as illustrated in Eqs. (3)–(8) (Liu et al., 2024). Here, the influence of the current control quantity on the output is considered negligible.

$$h_k = \overline{A}h_{k-1} + \overline{B}x_k \tag{3}$$

$$y_k = \overline{C}h_k + \overline{D}x_k \tag{4}$$

$$\overline{A} = e^{A\Delta t} \tag{5}$$

$$\overline{B} = \left(e^{A\Delta t} - I\right)A^{-1}B \tag{6}$$

$$\overline{C} = C \tag{7}$$

$$\overline{D} = D \tag{8}$$

Where, $A \in \mathbb{C}^{N\times N}$, $B, C \in \mathbb{C}^N$, and $D \in \mathbb{C}^1$ are the weighting parameter. Δt is the time step in the discretization process. $\overline{A}$ is a discretized state transition matrix calculated from the matrix exponent. $\overline{B}$ is the discretized input matrix considering zero-order holding. I is the identity matrix, which is used to subtract the identity matrix from the matrix exponent when calculating the discretized input matrix to calculate the delta of state transitions, with dimensions consistent with the system state matrix A. $\overline{C}$ is the output matrix and $\overline{D}$ is the direct transfer matrix, both of which usually remain constant during discretization.

The Mamba architecture incorporates a selective scanning mechanism, implemented in visual tasks through the two-dimensional Selective Scanning (SS2D) module (Shen et al., 2025).

This module enables deep integration with the Mamba framework via a multi-directional sequence reconstruction strategy. The input image is processed along four distinct directional paths, unfolding image blocks into sequences that preserve spatial continuity specific to each scanning orientation. These sequences are subsequently processed in parallel using enhanced S6 units, each of which adopts Mamba dynamic parameter generation mechanism. In this context, state-space parameters Δt, A, B, and C are derived adaptively during inference. Following parallel processing, the sequences are merged to reconstruct the output image. Unlike the original Mamba model, which employs a unidirectional parameter system, SS2D constructs independent parameter generation channels for each directional path. This design allows for the adaptive weighting of parameters based on directional characteristics, enabling each pixel to incorporate contextual information from all orientations.

To address the challenges posed by complex backgrounds in fabric defect detection, this study introduces the Visual State Space Module (MVSS). MVSS combines the long-range dependency modeling strengths of the Mamba framework with the efficient feature extraction capabilities of MLP. This integration enhances the network ability to capture critical visual features within high-dimensional data. The module enables dynamic attention to key regions in visually intricate textile patterns while maintaining low computational complexity, thereby supporting real-time applicability in industrial defect detection tasks.

The Visual State Space Module (MVSS), illustrated in Fig. 4, adopts a cascaded architecture specifically designed to address complex scenes and diverse defect types in fabric inspection tasks. The module receives input feature maps, which are initially processed through a 1 × 1 convolutional layer to extract preliminary representations. These feature maps are then evenly divided along the channel dimension. One portion is routed through a residual connection, while the other is forwarded to the MVSS Block for further processing. Within the MVSS Block, the input features are first standardized using a normalization layer to enhance data stability and consistency. The normalized features are then projected into a higher-dimensional space through a linear transformation and processed by the SS2D module, which performs selective scanning in four spatial directions using state space modeling techniques. These complementary directional traversals expand the global receptive field and enable the extraction of critical spatial features from multiple perspectives. A subsequent linear layer adjusts the output channel dimension to match that of the original input. An attention map is then generated and regularized via the DropPath method (Yang et al., 2022) which mitigates overfitting while preserving essential information. The resulting features are combined with the residual branch to complete the block.

To enhance feature representation, each MVSS block includes a multi-layer perceptron (MLP) module composed of layer normalization, linear projections, non-linear activation functions, and DropPath regularization. This design facilitates deep feature refinement by capturing complex interactions and latent structures across dimensions. A residual connection is applied after MLP processing to preserve gradient flow and promote model stability (Xie, 2024).

The outputs from multiple MVSS blocks are concatenated and aggregated using a learnable weighting mechanism, enabling comprehensive high-dimensional feature fusion. This approach leverages the representational power of MLPs to enhance feature extraction and apply dynamic spatial attention. As a result, the model demonstrates increased sensitivity to informative regions within the input. Additionally, the architecture supports efficient modeling of long-range dependencies with linear computational complexity, offering a balanced trade-off between accuracy and computational cost. The pseudocode for the MVSS module is presented in Table 2.

4.5. Programmable Gradient Information (PGI)

During the training process of deep neural networks, gradient information is crucial for guiding the update of network parameters. However, as the network depth increases, information loss inevitably occurs during the forward propagation of input data, making it difficult

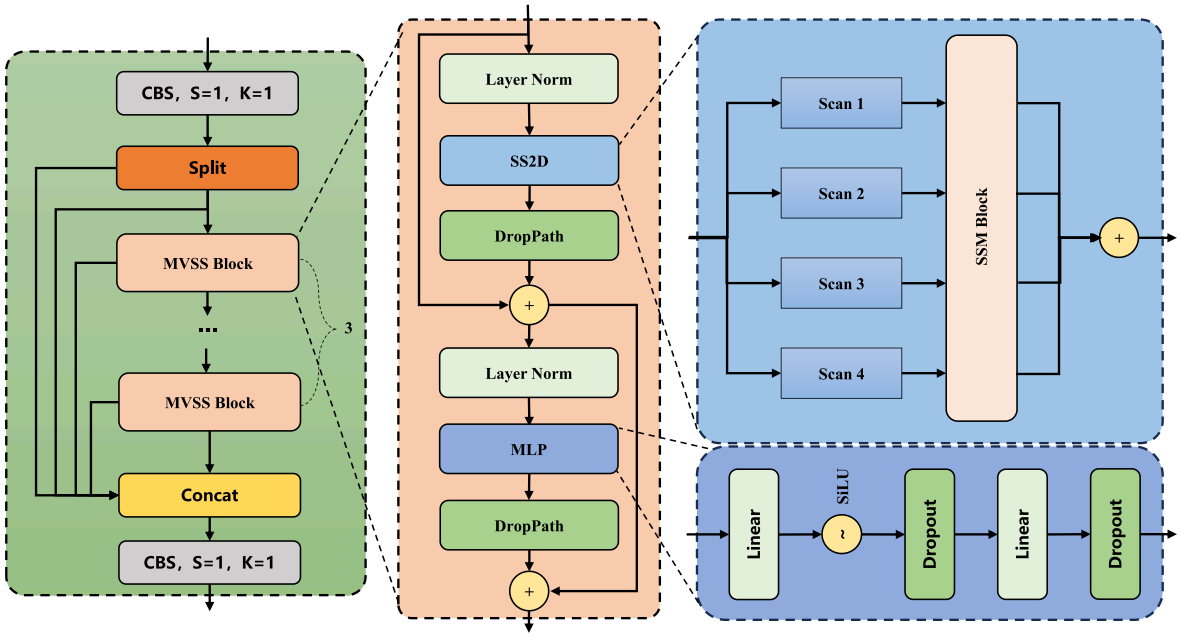

Fig. 4. Architecture of the MVSS module. The left panel shows the full MVSS structure; the middle illustrates the MVSS Block; the right details the SS2D and MLP components.

Table 2
Pseudocode of the MVSS module.

Algorithm: MVSS Module	
Input: $X \in \mathbb{R}^{C_{in} \times H \times W}$	
Output: $Y \in \mathbb{R}^{C_{out} \times H \times W}$	
1: Initialize hidden channels: c ← int(c_{out} * 0.5)	▷ Step 1: Apply 1 × 1 convolution to double hidden channels
2: cv_1 ← Conv(X, 2 * c, 1 × 1)	
3: Split cv_1 output into two parts: y_1, y_2 ← cv_1.chunk(2, 1)	
4: Initialize y ← [y_1, y_2]	
5: for each MVSS Block layer m in ModuleList: y_{new} ← m(y[-1]) Append y_{new} to y	▷ Step 2: Apply MVSS Block transformation
6: Concatenate all feature maps in y along channel dimension	
7: Y ← Conv(Concatenated_y, c_{out}, 1 × 1)	▷ Step 3: Fuse features using 1 × 1 convolution
8: Return Y	▷ Step 4: Output the final feature map
–	
Function: MVSS Block	
Input: $X \in \mathbb{R}^{hidden_{dim} \times H \times W}$	
Output: Enhanced feature map	
1: Permute input X to (0, 2, 3, 1) format	▷ Prepare input for processing
2: X_{norm} ← norm_layer(X)	▷ Apply normalization
3: X_{ssm} ← op(X_{norm})	▷ Apply SS2D operation
4: X ← X + drop_path(X_{ssm})	▷ Add residual connection with DropPath
5: X_{norm2} ← norm_layer(X)	▷ Apply second normalization
6: X_{mlp} ← mlp(X_{norm2})	▷ Apply MLP transformation
7: X ← X + drop_path(X_{mlp})	▷ Add residual connection with DropPath
8: Permute X back to original dimensions (0, 3, 1, 2)	▷ Restore original shape
9: Return X	▷ Output enhanced feature map

for deep features to retain key information from shallow features. This phenomenon is known as the information bottleneck problem (Tishby and Zaslavsky, 2015). To overcome this challenge, the FabricMamba model introduces the PGI module to assist training, as shown in Fig. 5. The core idea of this module is to add an auxiliary reversible branch and introduce auxiliary supervisory signals at different levels, thereby providing additional gradient information to the main network, effectively correcting and enhancing the original gradient flow. It is worth mentioning that the PGI module can be removed during the inference stage, thus it does not add any extra computational burden.

5. Experimental results and analysis

To validate the effectiveness of the proposed defect detection system, a case study is reported in which various textured fabric images with common industrial defects were selected for analysis. Due to their visual irregularities, these fabrics pose challenges to traditional defect detection techniques and are often difficult to accurately identify. Therefore, more advanced detection methods are applied to these complex textured fabrics to ensure the accurate identification and localization of various defects, thereby improving the efficiency and accuracy of quality control in the fabric manufacturing process.

5.1. Datasets and implementation details

In this research, a series of experimental tests was conducted on three datasets, including a clothing fabric defect dataset collected from a real factory setting, here referred to as ESQUEL dataset, as well as two publicly available datasets: TILDA (Andersen, 2021) and FDDS (Fabric, 2024). The ESQUEL dataset, collected from an industrial partner, provides a realistic representation of fabric defects encountered in actual manufacturing settings. Its inclusion ensures the proposed method practical applicability. The TILDA dataset, released by Irvin Andersen on the Roboflow platform, provides a diverse and annotated collection of fabric defect images, supporting comprehensive evaluation of defect detection models. The FDDS dataset from the Roboflow computer vision project complements these by simulating complex real-world fabric inspection scenarios. By testing on these datasets with varying characteristics, the study validates the proposed FabricMamba model robustness and generalization capabilities across different fabric types and defect patterns.

Data collection and annotation were conducted under different lighting conditions and for various styles of clothing. To preserve more details, this study uses a high-resolution MV-CS200-10UM industrial camera, adjusting the image resolution to 2560 × 2560 during model

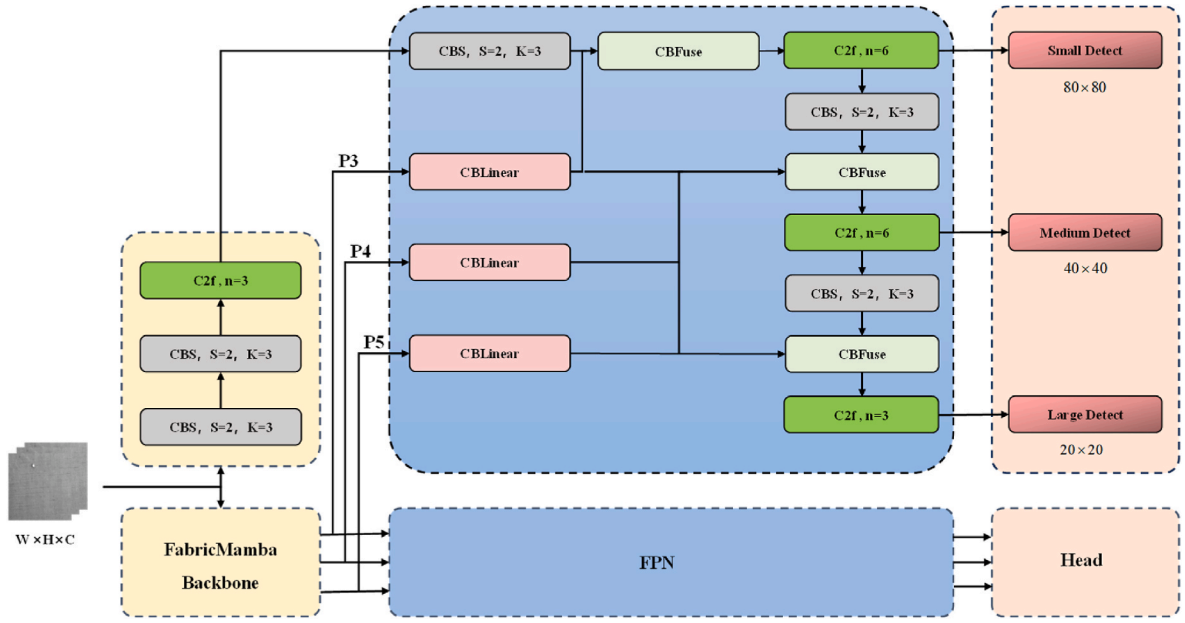

Fig. 5. Architecture of the PGI module; top: auxiliary reversible branch for feature refinement; bottom: main FabricMamba network for multi-scale defect detection.

training. Additionally, to better integrate with existing production lines, the YGV-2, commonly used in the garment defect inspection process in textile factories, is employed for backlighting, and image data is collected in the factory environment. Image augmentation techniques were applied to the ESQUEL dataset. such as rotation, flipping, and adding random noise, expanding the dataset to 1397 images, which includes 1116 training images and 281 validation images. This dataset contains three types of fabric defects: Hole, Float Thread, and Oil Stain. These three defects represent common types found in factories, with holes being the most frequent and unacceptable in textile factories, thus having the highest proportion. Table 3 shows the specific distribution of these defects in the ESQUEL dataset.

To further validate the robustness of the proposed system, additional experimental tests were carried out on two publicly available datasets. The TILDA dataset is a textile texture database sourced from the German Research Foundation. The dataset was augmented to 1286 images by flipping and scaling the fabric images, which includes 1028 training images and 258 test images. This dataset contains four types of fabric defects: Hole, Objects, Oil Spot, and Thread Error. The FDDS dataset is sourced from the Roboflow computer vision project and includes 1236 training images and 285 validation images, without any data augmentation techniques, accurately simulating the complex scenarios encountered in real-world fabric defect detection tasks. This dataset includes four types of fabric defects: Cut, Hole, Stain, and Thread Error.

The experimental tests were carried out on a specific hardware server, using the Pytorch deep learning framework for algorithm development and model training. The hardware specifications include a 14th generation Core i7-14700KF processor, 32 GB of memory, and an RTX4090D 24 GB graphics card, ensuring an efficient computing environment. In terms of training settings, all models uniformly used the SGD optimization algorithm with a batch size of 64, enabled mixed precision training, and used the LambdaLR learning rate scheduler with an initial learning rate of 0.01 and a weight decay of 0.0005. Mosaic data augmentation was employed during the training process and was turned off in the last ten rounds.

Table 3
Distribution of defect categories in the training and validation sets of the ESQUEL dataset.

Classes	Train	Val	Total
Hole	704	171	875
Float Thread	255	48	303
Oil Stain	330	62	392

5.2. Evaluation metrics

The performance of the fabric defect detection system is evaluated using the key object detection metrics described in Eqs. 9–14 (Padilla et al., 2020):

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$AR = \frac{1}{N}\sum_{i=1}^{N} Recall_i \tag{11}$$

$$AP = \int_0^1 Precision(r)dr \tag{12}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{13}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{14}$$

Where TP (True Positives) represents the number of correct predictions of defects, and FP (False Positives) represents the number of incorrect predictions of defects as present when they are not. $Precision_i$ represents the precision for the i-th category, and N is the total number of categories. FN represents False Negatives, which are the actual positive samples that were incorrectly predicted as negative. $Recall_i$ represents the recall of the i-th category. AR is the average of the recall rates across all categories. AP_i represents the Average Precision score for the i-th category, measuring the accuracy of the model predictions by calculating the area under the precision-recall curve for that specific category. mAP is a comprehensive metric for evaluating the localization and classification performance of object detection models across multiple categories. mAP50 refers to the AP at an IoU threshold of 0.5, while

mAP50-95 represents the AP across IoU thresholds from 0.5 to 0.95. The F1 score is the harmonic mean of precision and recall, used to evaluate the overall performance of a model.

5.3. Comparison with state-of-the-art methods

Table 4 reports the experimental results of the FabricMamba model on the ESQUEL dataset. Tables 5 and 6 present the comparative evaluation of FabricMamba against state-of-the-art methods on the TILDA and FDDS datasets, respectively. To account for the effects of training variability and to ensure statistical reliability, each model was trained and evaluated multiple times using different random seeds. The results in each table include mean values derived from these repeated trials. The highest-performing metrics across all methods are indicated in bold for clarity.

As shown in Tables 1–3, the FabricMamba model achieved the highest mAP scores across all three datasets, recording 90.0 % on ESQUEL, 97.7 % on TILDA, and 39.1 % on FDDS. These results represent a substantial improvement over the baseline YOLOv8 model. In comparison with other state-of-the-art detectors in the YOLO series, the proposed method consistently maintained high detection accuracy, particularly for critical defect categories such as “Hole,” which are considered unacceptable in industrial fabric production. For instance, on the ESQUEL dataset, FabricMamba achieved an Average Precision (AP) of 90.0 % for the “Hole” category and 89.7 % for “Oil Spot.” On the TILDA dataset, the corresponding AP values were 98.5 % and 99.5 %, respectively. In the FDDS dataset, FabricMamba achieved an AP of 39.6 % for the “Cut” category. These values consistently surpassed those of competing models. The results indicate that conventional YOLO-based models, constrained by local receptive fields, have limited ability to detect subtle or small-scale defects within complex fabric textures. Although Mamba-YOLO introduces a state-space model to address this limitation, its unidirectional sequence modeling restricts the capacity to capture multi-directional defect features.

To assess the robustness and statistical significance of the performance improvements, a five-fold cross-validation procedure was conducted on the ESQUEL dataset. Additionally, a paired samples *t*-test was employed to compare the mAP50 performance of FabricMamba against the YOLOv8 baseline. As reported in Table 7, the observed performance improvement was statistically significant ($p < 0.05$). Furthermore, the standard deviation of results was reduced by 47.3 %, indicating greater model robustness and reduced sensitivity to data partitioning.

The bar chart in Fig. 6 demonstrates that YOLOv9 and YOLOv10 exhibit notable efficiency in terms of computational resources, with YOLOv9 utilizing 2.0 million parameters and 7.7 GFLOPs, and YOLOv10 employing 2.7 million parameters and 8.2 GFLOPs. FabricMamba, by comparison, operates with 3.8 million parameters and 9.1 GFLOPs, indicating a modest increase in computational complexity. Despite the slightly higher resource consumption, FabricMamba achieves superior detection performance. On the ESQUEL dataset, FabricMamba outperformed YOLOv9 and YOLOv10 by 3.0 % and 1.9 %, respectively, in terms of mAP. On the TILDA dataset, the improvement over YOLOv9 and YOLOv10 was 1.9 % and 1.6 %, respectively. In the FDDS dataset, the corresponding gains were 0.6 % and 4.9 %.

When compared to Mamba-YOLO, FabricMamba demonstrated enhanced efficiency, with reductions of 36.7 % in model size and 33.1 % in computational demand, while maintaining superior accuracy. These findings suggest that FabricMamba achieves a balanced trade-off between model complexity and computational efficiency. While YOLOv9 and YOLOv10 exhibit lower resource requirements, this comes at the cost of reduced classification accuracy. In contrast, FabricMamba integrates two key architectural enhancements that contribute to its improved performance. The P-LSKA module enhances multi-scale feature representation, and the MVSS module increases the global receptive field, enabling more effective differentiation of subtle defects within complex fabric patterns. These architectural improvements directly address the specific challenges in fabric defect detection while maintaining computational efficiency for real-time industrial applications.

5.4. Ablation experiments

Table 8 describes the individual contributions of the various modules integrated in the FabricMamba model evaluated on the ESQUEL, TILDA and FDDS datasets respectively. These modules include S- The F1 score serves as a balanced metric for evaluating model performance, rewarding models that successfully minimize false positives while maximizing true positives. P-LSKA, MVSS, DySample, and PGI. A high F1 score reflects an effective balance between precision and recall.

The mean Average Precision (mAP) measures the average precision across all categories and incorporates both precision and recall, along with model performance at varying confidence thresholds. In object detection tasks, mAP is the most widely adopted evaluation metric, offering a comprehensive assessment of a model detection capabilities across multiple categories and confidence levels.

As presented in Table 5, the integration of the P-LSKA, MVSS, DySample, and PGI modules results in the highest performance across all three evaluated datasets. The inclusion of these components leads to consistent improvements in both the average F1 score (Avg_F1) and mAP, indicating the effectiveness of the modular architecture in enhancing detection performance on benchmark fabric defect datasets. It is notable that even in the absence of PGI-assisted training, the FabricMamba model maintains a high mAP, reflecting its inherent robustness.

An analysis of computational overhead reveals that the P-LSKA module increases the number of parameters by approximately 18.8 % and raising FLOPs by 6.9 %. In contrast, the MVSS module introduces no additional computational cost or parameter increase compared to the baseline. Despite the modest rise in resource consumption, the full FabricMamba model demonstrates clear performance gains relative to the YOLOv8 baseline. These results confirm that the proposed architecture achieves an effective balance between computational efficiency and detection accuracy, making it suitable for real-time industrial deployment.

Table 4
Comparison of the performance of different methods on ESQUEL dataset.

Method	Year	Hole	AP (%) Float Thread	Oil Stain	mAP50 (%)	mAP50-95 (%)	AR(%)	Dataset
YOLOv3 (Redmon and Farhadi, 2018)	2018	87.8	**91.7**	84.6	88.0	56.4	82.3	ESQUEL
YOLOv5	2020	88.1	86.9	89.1	88.0	55.7	82.6	
YOLOv6 (C. Li et al., 2022)	2020	83.1	89.7	89.1	87.3	53.3	83.0	
YOLOv7	2022	83.3	90.8	87.3	87.2	53.0	83.1	
YOLOv8	2023	87.0	89.1	88.4	88.2	58.4	81.0	
YOLOv9	2024	85.5	90.4	84.9	87.0	55.4	80.4	
YOLOv10 (Wang et al., 2024)	2024	90.0	88.5	85.7	88.1	58.5	81.3	
Mamba YOLO (Wang et al., 2024)	2024	86.8	87.8	87.6	87.4	57.0	82.9	
FabricMamba	2024	**90.0**	90.2	**89.7**	**90.0**	60.1	83.9	

Table 5
Comparison of the performance of different methods on the TILDA dataset.

Method	Year	AP (%)				mAP50 (%)	mAP50-95 (%)	AR(%)	Dataset
		Hole	Objects	Oil Spot	Thread Error				
YOLOv3	2018	88.7	91.3	99.4	93.6	93.2	77.3	90.5	TILDA
YOLOv5	2020	95.5	96.0	99.3	94.6	96.3	79.6	92.4	
YOLOv6	2020	89.3	95.4	99.5	93.2	94.4	75.7	90.7	
YOLOv7	2022	90.8	92.9	98.9	96.8	94.9	69.8	90.4	
YOLOv8	2023	92.4	94.3	99.5	95.5	95.4	76.1	93.0	
YOLOv9	2024	92.1	94.4	99.5	97.2	95.8	81.7	91.8	
YOLOv10	2024	84.7	95.1	99.4	95.3	96.1	81.8	90.3	
Mamba YOLO	2024	93.9	**97.1**	99.5	**97.9**	97.1	83.9	95.2	
FabricMamba	2024	**98.5**	96.8	**99.5**	95.8	**97.7**	84.1	94.4	

Table 6
Comparison of the performance of different methods on the FDDS dataset.

Method	Year	AP (%)				mAP50(%)	mAP50-95(%)	AR(%)	Dataset
		Cut	Hole	Stain	Thread Error				
YOLOv3	2018	29.7	**36.0**	**40.6**	35.4	35.4	14.9	39.4	FDDS
YOLOv5	2020	33.4	32.5	31	41.1	34.5	14.3	42.8	
YOLOv6	2020	37.2	28.7	32.5	32.3	32.7	13.9	39.1	
YOLOv7	2022	28.6	33.9	39.2	20.9	30.7	13.5	33.0	
YOLOv8	2023	35.3	32.4	39.9	40.8	37.1	16.5	40.9	
YOLOv9	2024	37.0	36.4	33.2	47.7	38.5	16.4	42.0	
YOLOv10	2024	35.4	23.5	33.4	44.6	34.2	14.5	43.3	
Mamba YOLO	2024	35.2	35.1	36.5	**48.1**	38.7	16.4	42.0	
FabricMamba	2024	**39.6**	35.5	38.0	43.2	**39.1**	16.9	44.9	

Table 7
Paired *t*-test results of YOLOv8 and FabricMamba on mAP50.

Evaluation Method	YOLOv8 mAP50	FabricMamba mAP50	Improvement	t-value	p-value
Original	88.2	90.0	+1.80	–	–
Re-evaluation	96.52 ± 0.85	97.76 ± 0.43	+1.24	−4.032	0.016

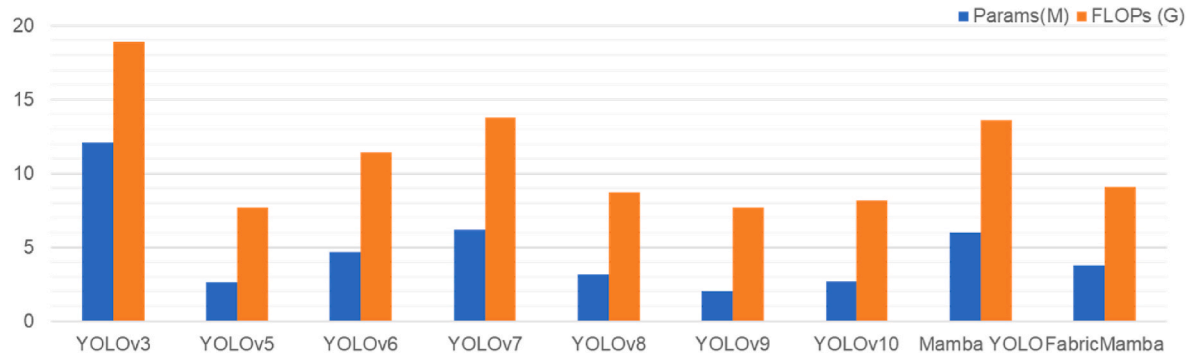

Fig. 6. Comparison of different network models in terms of parameters (M) and FLOPs (G).

5.5. Visualization of detection results

Figs. 7–9 show the results of the FabricMamba network model used on three datasets for fabric surface defect detection, along with a heatmap visualization. The display for each dataset includes three rows: the first row shows the original test images, the second row displays the detection results with different defects marked in various colors and annotated with defect categories and confidence levels. The third row shows the corresponding heatmaps, in which the red areas represent the defect regions of interest.

This research utilizes the HiResCAM (Wu et al., 2024) visualization method to demonstrate the model performance in the backbone network. Regardless of the size variations of the target objects, FabricMamba is able to allocate higher weights to the detection areas and maintains high sensitivity to the locations of objects. Thanks to the design of P-LSKA and MVSS, the detector effectively captures the local and global fusion features of position information for objects of various sizes.

This visualization highlights FabricMamba focus on environmental and defect target areas. It demonstrates the model ability to accurately locate fabric defects within complex backgrounds and effectively extract critical features from visual noise, achieving precise identification and annotation. Additionally, it accounts for environmental heat loss, ensuring minimal interference from surrounding areas. These results validate the model effectiveness in capturing diverse surface characteristics under real-world conditions, showcasing its robust detection capabilities, making it well-suited for real-time fabric defect detection in industrial manufacturing settings.

Table 8
Performance of different components in the FabricMamba network.

P-LSKA	MVSS	DySample	PGI	F1	AP (%)	AR (%)	mAP50 (%)	mAP50 -95(%)	Dataset
✔			✔	0.95	95.0	95.8	97.1	83.6	TILDA
	✔		✔	0.95	95.9	93.8	96.4	82.7	
		✔	✔	0.96	97.3	94.8	96.8	82.5	
✔	✔	✔		0.95	96.0	94.4	97.4	82.7	
✔	✔	✔	✔	**0.96**	96.8	94.4	**97.7**	84.1	
✔			✔	0.41	37.5	45.0	35.9	15.6	FDDS
	✔		✔	0.43	41.3	46.5	35.3	14.5	
		✔	✔	0.43	39.5	47.3	38.6	15.9	
✔	✔	✔		0.43	42.6	43.6	37.7	15.9	
✔	✔	✔	✔	**0.44**	43.0	44.9	**39.1**	16.9	
✔			✔	0.86	91.7	81.5	88.4	57.4	ESQUEL
	✔		✔	**0.89**	95.1	83.1	88.1	56.0	
		✔	✔	0.86	94.6	79.0	88.4	58.2	
✔	✔	✔		0.86	92.6	80.0	88.2	57.8	
✔	✔	✔	✔	0.87	94.5	83.9	**90.0**	60.1	

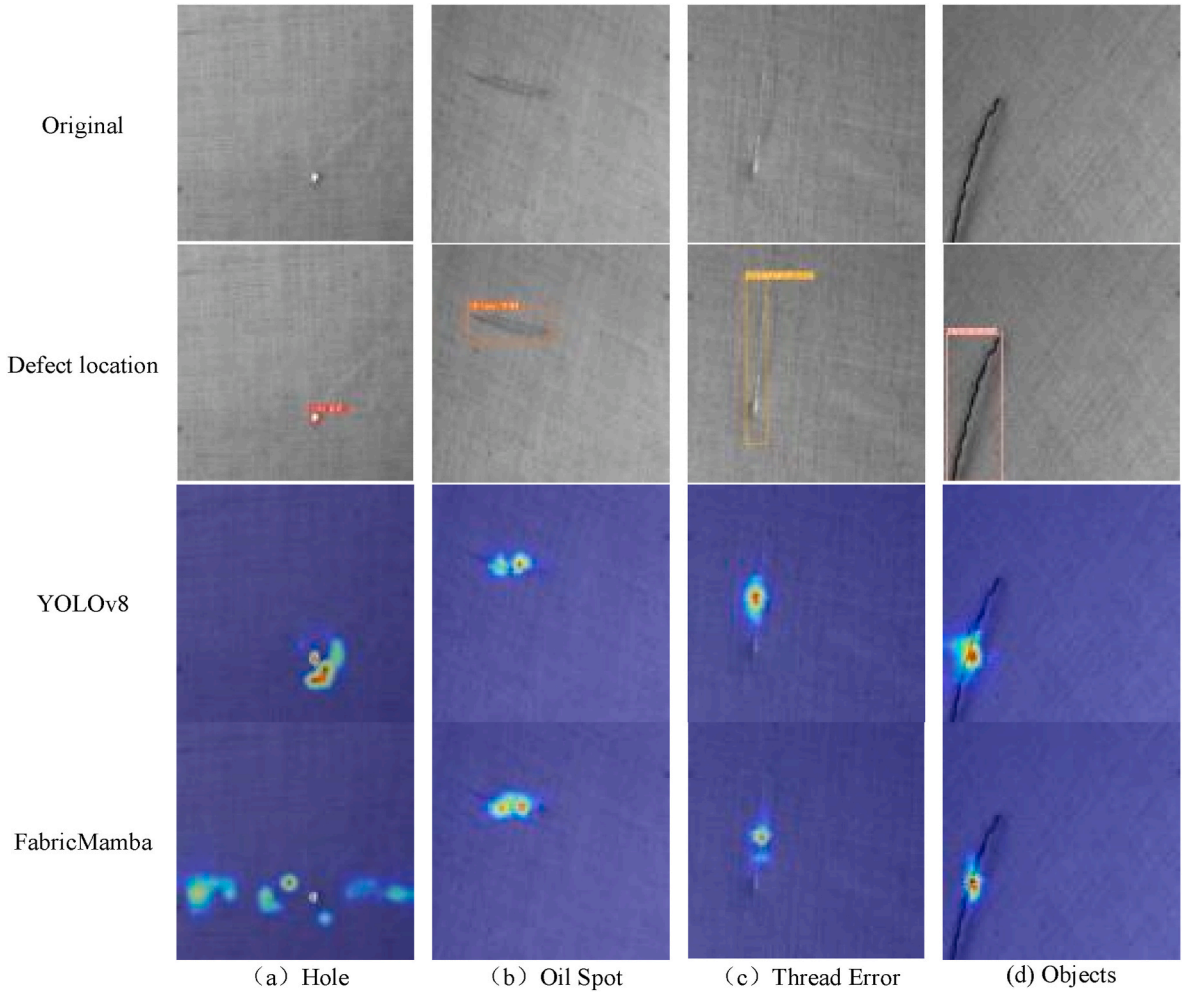

Fig. 7. Detection results of test samples on the TILDA dataset and corresponding heat maps; red areas in the heatmaps indicate defects regions of interest. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

6. Conclusions

This study introduced state space models into the domain of fabric defect detection, presenting a novel detector that combines large-kernel attention with visual state space modeling. The proposed architecture integrated the efficiency of separable large-kernel attention for local feature extraction with the long-range dependency modeling capabilities of state space representations. The model, named FabricMamba, was evaluated on one proprietary industrial dataset and two public datasets (TILDA and FDDS), achieving the highest mAP scores of 90.0 %, 97.7 %, and 39.1 %, respectively. These results surpassed YOLOv8 by 1.8 %, 2.3 %, and 2.0 %, and recall improvements of 2.9 %, 1.4 %, and 4.0 % were also observed. FabricMamba consistently outperformed Mamba-YOLO across all datasets. These findings demonstrate that the proposed model delivers improved robustness, generalization, and efficiency, making a meaningful contribution to the development of high-

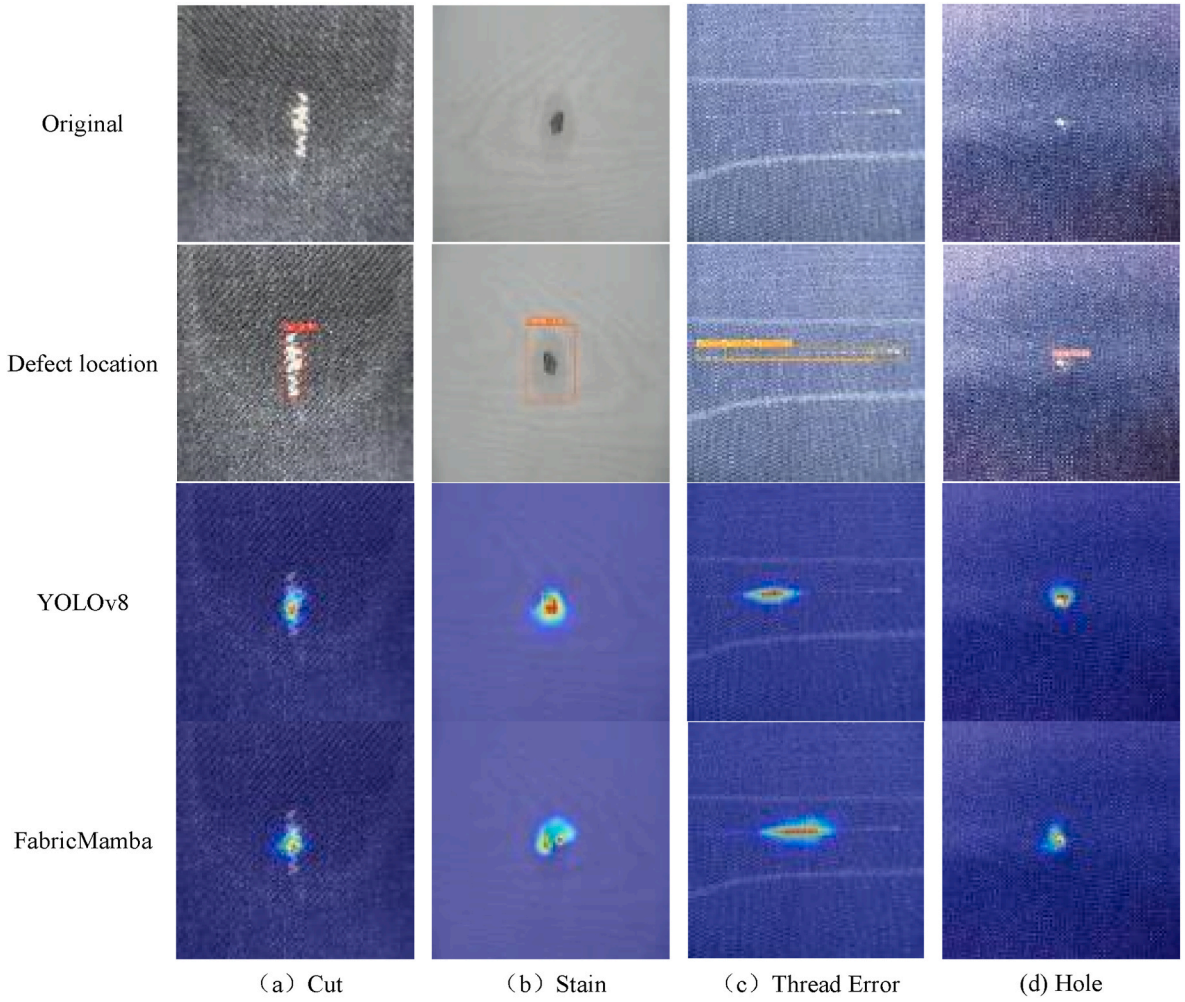

Fig. 8. Detection results of test samples on the FDDS dataset and corresponding heat maps; red areas in the heatmaps indicate defects regions of interest. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

performance defect detection systems for textile manufacturing. The model architecture, with its global receptive field and ability to capture long-range dependencies, suggests strong potential for generalization to novel fabric materials and unseen defect patterns. This adaptability stems from FabricMamba ability to extract fundamental structural anomalies rather than focusing on specific appearance characteristics, although additional testing across a wider range of fabric types would further validate this capability.

The FabricMamba method developed in this research already features a relatively compact model size, but there remains scope for further optimization. Future research will explore model compression techniques to reduce the model footprint and enhance its suitability for deployment on resource-constrained platforms, including mobile and embedded systems. On-site testing conducted in real factory settings demonstrated that the proposed system substantially improves production efficiency, reduces manual labor requirements and error rates, and lowers both operational costs and the frequency of rework. In addition, by minimizing material waste and defect occurrence, the automated detection process contributes to more sustainable manufacturing practices, supporting industry goals for eco-efficiency and reduced environmental impact.

Looking ahead, the FabricMamba architecture could be extended to other industrial domains such as metal surface inspection, plastic component evaluation, and quality assurance in electronics manufacturing. Furthermore, future research may incorporate reinforcement learning methods to enable adaptive detection systems capable of continuous performance refinement through real-time production feedback. These developments would further enhance the model generalization ability and reduce the need for retraining when applied to new materials or previously unseen defect types, reinforcing its value in dynamic and heterogeneous production environments.

CRediT authorship contribution statement

Nengsheng Bao: Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Jiajun Lin:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Yuchen Fan:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Runxuan Bao:** Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Alessandro Simeone:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used ChatGPT and DEEPL in order to improve syntax and grammar. After using these tools, the authors reviewed and edited the content as needed and assume full responsibility for the content of the manuscript.

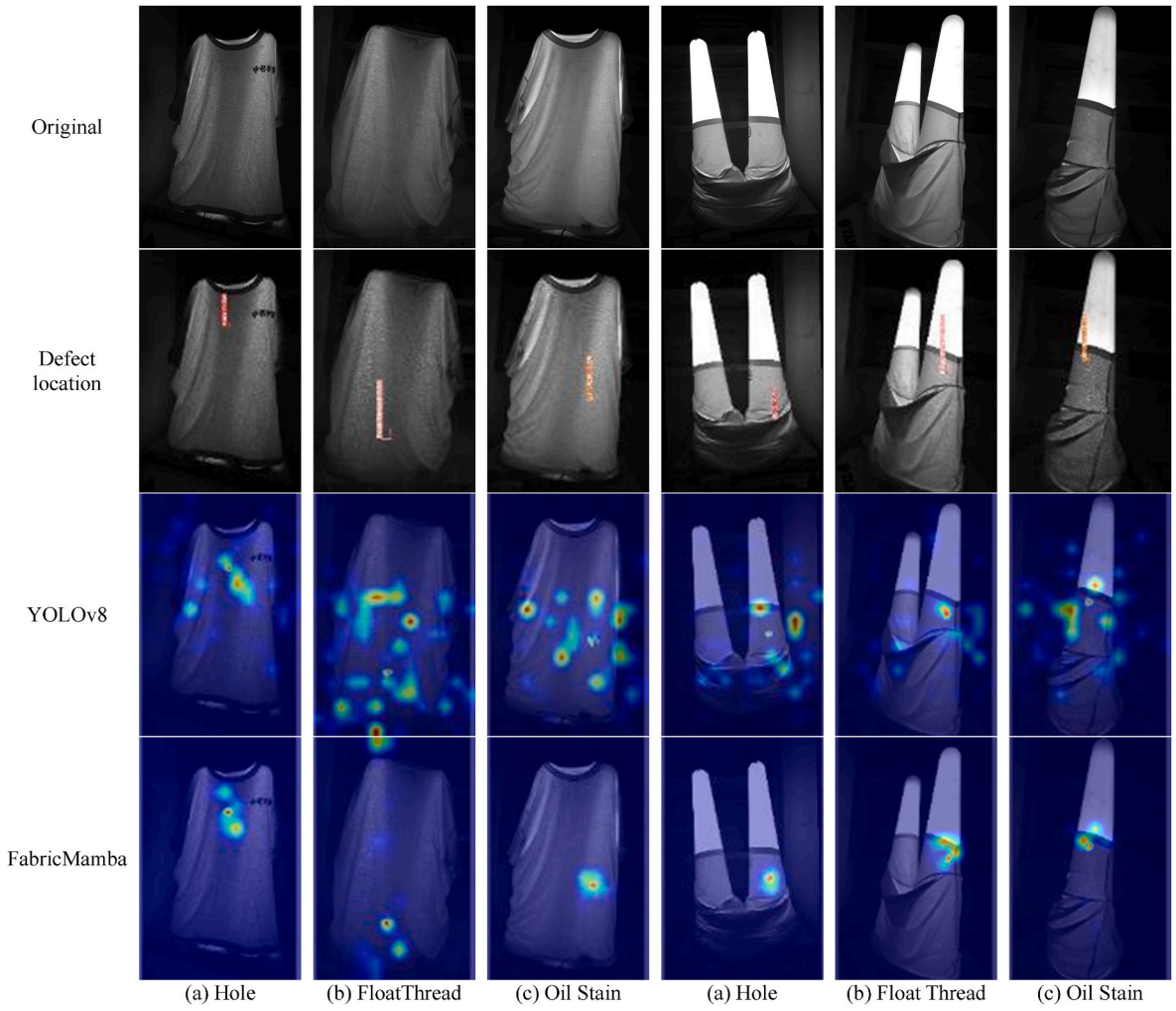

Fig. 9. ESQUEL sample detections with heat maps; red areas in the heatmaps indicate defects regions of interest. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to express their sincere gratitude to Guangdong Esquel Textiles Co., Ltd. and Shantou JunGuo Mechanical & Electrical Technology Co., Ltd. for their valuable support and contributions to this work. This research was also supported by the Guangdong Province Quality Engineering Project: Construction of Shantou University Intelligent Manufacturing Industry College (Grant No. GDEDU20230111023).

Data availability

The datasets generated during this study are not publicly available due to industrial restrictions and the use of proprietary, self-built data.

References

An, G., He, A., Wang, Y., Guo, J., 2024. UWMamba: UnderWater image enhancement with state space model. IEEE Signal Process. Lett. 31, 2725–2729.

Andersen, I., 2021. TILDA fabric dataset [WWW Document]. Roboflow Universe, Open Source Dataset, visited on 2025-04-16. URL. https://universe.roboflow.com/irvin-andersen/tilda-fabric.

Arshad, S.R., Shahzad, M.K., 2024. Deep learning based fabric defect detection. Res. Rep. Computer Sci.

Ayalew, W., Menebo, M., Merga, C., Negash, L., 2024. Optimal path planning using bidirectional rapidly-exploring random tree star-dynamic window approach (BRRT*-DWA) with adaptive Monte Carlo localization (AMCL) for mobile robot. Eng. Res. Express 6, 035212.

Bai, Z., Jing, J., 2024. Mobile-Deeplab: a lightweight pixel segmentation-based method for fabric defect detection. J. Intell. Manuf. 35, 3315–3330.

Chansong, D., Supratid, S., 2021. Impacts of kernel size on different resized images in object recognition based on convolutional neural network. In: 2021 9th International Electrical Engineering Congress (IEECON). IEEE, pp. 448–451.

Ding, X., Zhang, X., Han, J., Ding, G., 2022. Scaling up your kernels to 31×31: revisiting large kernel design in CNNs. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 11953–11965.

Fabric, 2024. FDDS dataset [WWW Document]. Roboflow Universe, Open Source Dataset, visited on 2025-04-16. URL. https://universe.roboflow.com/fabric-9fmcc/fdds.

Gu, A., Dao, T., 2024. Mamba: Linear-Time Sequence Modeling with Selective State Spaces.

Gu, A., Dao, T., Ermon, S., Rudra, A., Re, C., 2020. HiPPO: recurrent memory with optimal polynomial projections. Preprint ArXiv.

Gu, A., Johnson, I., Goel, K., Saab, K., Dao, T., Rudra, A., Ré, C., 2021. Combining recurrent, convolutional, and continuous-time models with linear state space layers. Adv. Neural Inf. Process. Syst. 34, 572–585.

Guo, M.-H., Lu, C.-Z., Liu, Z.-N., Cheng, M.-M., Hu, S.-M., 2023. Visual attention network. Comput Vis Media (Beijing) 9, 733–752.

Guo, P., Liu, Y., Wu, Y., Gong, R.H., Li, Y., 2024. Intelligent quality control of surface defects in fabrics: a comprehensive research progress. IEEE Access 12, 63777–63808.

Ikotun, A.M., Ezugwu, A.E., Abualigah, L., Abuhaija, B., Heming, J., 2023. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. Inf. Sci. 622, 178–210.
Jing, J., Ma, H., Zhang, H., 2019. Automatic fabric defect detection using a deep convolutional neural network. Color. Technol. 135, 213–223.
Kitamura, K., Itou, K., 2023. Binaural audio synthesis with the structured state space sequence model. In: 2023 9th International Conference on Computer and Communications (ICCC). IEEE, pp. 1505–1509.
Lau, K.W., Po, L.-M., Rehman, Y.A.U., 2024. Large separable kernel attention: rethinking the large kernel attention design in CNN. Expert Syst. Appl. 236, 121352.
Li, C., Li, Lulu, Jiang, H., Weng, K., Geng, Y., Li, Liang, Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, Xiaoming, Wei, Xiaolin, 2022. YOLOv6: a single-stage object detection framework for industrial applications. ArXiv Preprint.
Li, J., Kang, X., 2024. Mobile-YOLO: an accurate and efficient three-stage cascaded network for online fiberglass fabric defect detection. Eng. Appl. Artif. Intell. 134, 108690.
Li, Q., Zhu, L., Li, B., Luo, J., Liu, D., Chen, M., 2022. Small target detection algorithm based on YOLOv4. In: 2022 International Seminar on Computer Science and Engineering Technology (SCSET). IEEE, pp. 76–79.
Liu, J., Yang, H., Zhou, H.-Y., Xi, Y., Yu, L., Li, C., Liang, Y., Shi, G., Yu, Y., Zhang, S., Zheng, H., Wang, S., 2024. Swin-UMamba: Mamba-Based UNet with ImageNet-Based Pretraining, pp. 615–625.
Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A ConvNet for the 2020s. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 11966–11976.
Madebo, M.M., Abdissa, C.M., Lemma, L.N., Negash, D.S., 2024. Robust tracking control for quadrotor UAV with external disturbances and uncertainties using neural network based MRAC. IEEE Access 12, 36183–36201.
Ngan, H.Y.T., Pang, G.K.H., Yung, N.H.C., 2011. Automated fabric defect detection—a review. Image Vis Comput. 29, 442–458.
Padilla, R., Netto, S.L., da Silva, E.A.B., 2020. A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE, pp. 237–242.
Peng, M., Zhang, W., Li, F., Xue, Q., Yuan, J., An, P., 2023. Weed detection with improved Yolov 7. EAI Endorsed Transac. Internet Things 9, e1.
Rasheed, Aqsa, Zafar, B., Rasheed, Amina, Ali, N., Sajid, M., Dar, S.H., Habib, U., Shehryar, T., Mahmood, M.T., 2020. Fabric defect detection using computer vision techniques: a comprehensive review. Math. Probl Eng. 2020, 1–24.
Redmon, J., Farhadi, A., 2018. YOLOv3: an incremental improvement, 2018. In: Redmon, J., Farhadi, A. (Eds.), YOLOv3: an Incremental Improvement. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794–7803. Salt Lake City, UT, 18-22 June 2018.
Sampath, V., Maurtua, I., Martin, J.J.A., Iriondo, A., Lluvia, I., Rivera, A., 2022. Vision Transformer based knowledge distillation for fasteners defect detection. In: 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET). IEEE, pp. 1–6.
Shen, Y., Yao, S., Qiang, Z., Pei, G., 2025. SD-Mamba: a lightweight synthetic-decompression network for cross-modal flood change detection. Int. J. Appl. Earth Obs. Geoinf. 136, 104409.
Tian, L., Guan, Y., Peng, J., Chen, X., 2024. An improved YOLOv5s target detection algorithm. In: Bilas Pachori, R., Chen, L. (Eds.), International Conference on Remote Sensing, Mapping, and Image Processing (RSMIP 2024). SPIE, p. 140.
Tishby, N., Zaslavsky, N., 2015. Deep learning and the information bottleneck principle. In: 2015 IEEE Information Theory Workshop (ITW). IEEE, pp. 1–5.
Todi, A., Narula, N., Sharma, M., Gupta, U., 2023. ConvNext: a contemporary architecture for convolutional neural networks for image classification. In: 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT). IEEE, pp. 1–6.
Tomen, N., van Gemert, J.C., 2021. Spectral leakage and rethinking the kernel size in CNNs. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, pp. 5118–5127.
Wang, D., Zhao, W., Cui, K., Zhu, Y., 2024. VMC-UNet: a vision Mamba-CNN U-Net for tumor segmentation in breast ultrasound image. Int. J. Imag. Syst. Technol. 34.
Wang, X., Lv, R., Zhao, Y., Yang, T., Ruan, Q., 2020. Multi-scale context aggregation network with attention-guided for crowd counting. In: 2020 15th IEEE International Conference on Signal Processing (ICSP). IEEE, pp. 240–245.
Wanore Madebo, N., Merga Abdissa, C., Negash Lemma, L., 2024. Enhanced trajectory control of quadrotor UAV using fuzzy PID based recurrent neural network controller. IEEE Access 12, 190454–190469.
Wu, F., Song, S., Hu, J., 2024. Interpretable roadside 2D object detection based on axial attention. In: 2024 IEEE/ACIS 24th International Conference on Computer and Information Science (ICIS). IEEE, pp. 145–151.
Xie, R., 2024. The analysis of time series forecasting based on MLP models. Transac. Computer Sci. Intell. Syst. Res. 7, 326–332.
Yang, R., Guo, N., Tian, B., Wang, J., Liu, S., Yu, M., 2024. Fabric defect detection via saliency model based on adjacent context coordination and transformer. J. Eng. Fiber Fabr. 19.
Yang, X., Luo, Z., Wu, Y., Xie, X., Nan, L., Li, T., 2024. TMU: Transmission-Enhanced Mamba-UNet for Medical Image Segmentation, pp. 428–438.
Yang, Y., Xu, H., Lu, L., 2022. ASD-ISSPA:Adaptive stochastic droppath and interactive slow semi-polarized attention based on FPN for object detection. In: 2022 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.
Yue, X., Wang, Q., He, L., Li, Y., Tang, D., 2022. Research on tiny target detection technology of fabric defects based on improved YOLO. Appl. Sci. 12, 6823.
Zhang, H., Xiong, W., Lu, S., Chen, M., Yao, L., 2023. QA-USTNet: yarn-dyed fabric defect detection via U-shaped Swin transformer network based on quadtree attention. Textil. Res. J. 93, 3492–3508.
Zhang, Junyi, Chen, R., Liu, F., Liu, H., Zheng, B., Hu, C., 2024. DC-mamba: a novel network for enhanced remote sensing change detection in difficult cases. Remote Sens. (Basel) 16, 4186.
Zhang, Jianqiang, Hou, J., He, Q., Yuan, Z., Xue, H., 2024. MambaPose: a human pose estimation based on gated feedforward network and Mamba. Sensors 24, 8158.
Zhao, S., Zhong, R.Y., Xu, C., Wang, J., Zhang, J., 2025. A dynamic inference network (DI-Net) for online fabric defect detection in smart manufacturing. J. Intell. Manuf. 36, 2881–2896.
Zheng, Z., Wang, P., Ren, D., Liu, W., Ye, R., Hu, Q., Zuo, W., 2022. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Trans. Cybern. 52, 8574–8586.