textiles

MDPI

Article

FAD-RNet: A Reverse Distillation Network with Frequency-Decoupled Feature Fusion for Unsupervised Fabric Defect Localization

Shuheng Li, Jun Liu *, Jiuzhen Liang and Hao Liu

School of Computer Science and Artificial Intelligence, Changzhou University, Yanzheng West Avenue, Changzhou 213164, China; lsh155127@gmail.com (S.L.); jzliang@cczu.edu.cn (J.L.); helenliuhao@cczu.edu.cn (H.L.)

* Correspondence: liu_jun@cczu.edu.cn

Abstract

Unsupervised anomaly detection in industrial fabric inspection remains a formidable challenge due to the complexity of background textures and the subtle, irregular nature of real-world defects. Although the teacher-student distillation paradigm has demonstrated promising performance without reliance on anomalous data, existing methods still struggle in the presence of complex textures, largely due to limited semantic guidance, insufficient frequency modeling, and inadequate multi-scale representation. To address these limitations, we propose a novel reverse distillation framework tailored for fabric defect detection. The core of our method is the frequency decoupling Feature fusion module (FDFM), which achieves frequency domain alignment between teacher and student features through spatially adaptive and learnable filter banks, namely the adaptive high-pass filter (AHPF) and the adaptive low-pass filter (ALPF). Specifically: (1) the high-frequency pathway employs deconvolutional residual enhancement to emphasize boundary details; (2) the low-frequency pathway leverages the CARAFE operator to Handle these normal fluctuations to prevent the model from mistakenly identifying background changes as abnormal areas. This design not only maintains a lightweight architecture but also significantly improves sensitivity to fine-grained anomalies. Furthermore, we introduce a cross-layer residual alignment mechanism that guides the student network in reconstructing deep semantic representations from the teacher-student feature pairs. To balance detection accuracy and deployment efficiency, we develop two model variants: a high-capacity version optimized for precision, and a lightweight version tailored for real-time industrial applications. Compared with other methods from recent years, the experimental results of FAD-RNet validate its superiority in relevant metrics. It should be noted that this study is conducted based on the data organization and processing protocol of the ZJU-Leaper dataset, which may introduce certain dataset-specific characteristics.

Keywords: unsupervised anomaly detection; Reverse Knowledge Distillation; Frequency-Decoupled Feature Fusion; Fabric Defect Localization

Check for updates

Academic Editor: David E. Breen

Received: 7 February 2026
Revised: 27 March 2026
Accepted: 7 April 2026
Published: 11 May 2026

Copyright: © 2026 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.

1. Introduction

Computer vision plays a vital role in automated quality inspection. As a key intermediate product in the textile industry, fabrics are prone to surface defects such as yarn breaks, oil stains, and specks during production. Undetected defects compromise downstream product quality and manufacturer reputation [1]. However, detecting fabric defects is

challenging due to complex textures, diverse defect morphologies, small sizes, blurred boundaries, and low contrast,especially in repetitive or non-structured textures where ambiguous boundaries hinder accurate localization [2,3].

Traditional supervised methods rely on large amounts of labeled defect data, limiting their scalability in real-world scenarios with rare and imbalanced anomalies. Consequently, research has shifted toward unsupervised anomaly detection, which mainly follows three directions: (1) Reconstruction-based methods (e.g., AnoGAN [4], MemAE [5], VT-ADL [6]) use autoencoders or GANs trained on normal images to detect discrepancies, but often fail under high-frequency interference or subtle defects. (2) Synthetic anomaly-based methods (e.g., DRAEM [7], CutPaste [8], NSA [9]) improve robustness by introducing pseudo-anomalies, though limited by the realism and diversity of these synthetic patterns. (3) Knowledge distillation-based methods (e.g., ST [10], STFPM [11], FAST-DAD [12]) adopt a teacher-student structure to locate anomalies via feature differences, offering a balance of accuracy and efficiency.

Despite progress, three key challenges remain in complex textured environments:

1. Limited cross-layer semantic guidance restricts fine-grained reconstruction and subtle anomaly localization.

2.Lack of frequency-domain modeling leads to misclassification due to overlooked frequency separability.

3. Insufficient multi-scale modeling makes it difficult to detect anomalies of varying sizes in dense textures.

In fabric texture analysis, the minimal structural unit—often referred to as a "single grain"—varies across different weave patterns and textures. Rather than manually defining a fixed grain size, an effective detection system should adaptively capture structural units at multiple scales. Shallow layers can focus on fine-grained details such as individual yarns or small weave points, while deeper layers encode broader pattern structures. This multi-scale design is essential for eliminating the need for explicit grain size specification while ensuring sensitivity to defects of varying scales.

To address these, we propose FAD-RNet, as illustrated in Figure 1c, an unsupervised reverse distillation framework that exploits teacher-student feature reconstruction discrepancies for anomaly detection. The framework processes input images through a pretrained teacher encoder to extract multi-scale features. Our One-Class Bottleneck Embedding (OCBE) module compresses these features into compact latent representations aligned with the student decoder. During decoding, a reverse-symmetric student progressively recovers spatial details, guided by our FDFM. FDFM employs adaptive frequency filtering and cross-layer residuals to explicitly separate, reconstruct high and low-frequency components, enhancing sensitivity to subtle structural anomalies in complex backgrounds. During inference, significant feature discrepancies (measured via normalized cross-correlation) between teacher and student networks localize defects where low similarity indicates anomalies.

1. FDFM enhances perception of weak or blurred anomalies by separating high- and low-frequency components using spatially adaptive filters, a CARAFE-based operator [13], and attention-guided fusion.

2. A frequency-guided cross-layer residual alignment mechanism that integrates teacher-student interaction directly into the student's backbone, improving semantic consistency and structural representation.

3. Two model variants: a high-capacity version for maximum accuracy and a lightweight version for real-time deployment.

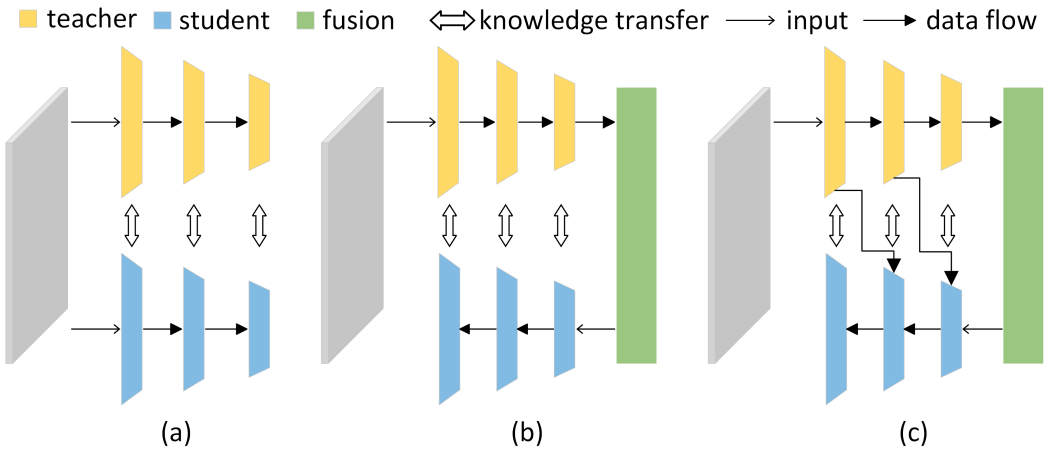


Figure 1. T-S models and data flow in (**a**) conventional KD framework, (**b**) traditional reverse distillation model and (**c**) our FAD-RNet.

Extensive experiments on the ZJU-Leaper dataset demonstrate that FAD-RNet achieves state-of-the-art performance across diverse fabric textures, maintaining high accuracy and efficiency under challenging conditions, making it well-suited for practical industrial inspection.

It is important to clarify that the proposed framework is developed and evaluated based on the data acquisition and preprocessing protocol of the ZJU-Leaper dataset. While the methodology is general in design, certain implementation details may be influenced by dataset-specific characteristics.

2. Related Work

2.1. Feature Reconstruction Method

Feature reconstruction is a dominant strategy in unsupervised anomaly detection. Models learn normal structural distributions from defect-free samples, detecting anomalies via reconstruction discrepancies. Since anomalous regions are unseen during training, they yield prominent residuals for scoring.

Early approaches focused on image-space reconstruction. Subsequent autoencoder-based frameworks gained popularity due to stability. Hybrid efforts like AE-SSIM [14] aimed to preserve details. Despite improvements, these methods remain constrained by pixel-space limitations.

Image-space reconstruction faces two core challenges in textured fabrics. First, dense backgrounds cause false regeneration of defects as normal. Second, conventional loss functions such as pixel-wise L1 or L2 lack sensitivity to structural anomalies.

Consequently, research shifted to feature-space reconstruction, leveraging pre-trained network features for richer anomaly representation. PaDiM [15] models normal feature distributions via multivariate Gaussians, using Mahalanobis distance for scoring. SPADE [16] detects anomalies via feature map similarity. These approaches demonstrate superior robustness in complex textures.

Feature-space reconstruction offers greater discriminative power than image-based methods, particularly for textured backgrounds. Future gains may arise from multi-scale architectures, attention mechanisms, and frequency-domain representations to enhance boundary anomaly perception.

2.2. Knowledge Distillation

Knowledge distillation has emerged to address the limitations of traditional reconstruction methods in unsupervised anomaly detection, especially for complex textures and subtle defects. This approach adopts a teacher-student architecture: a pre-trained teacher

network offers stable feature representations for normal samples, while a lightweight student network learns to mimic them, modeling the normal semantic distribution. During inference, anomalies lead to discrepancies in student features, enabling localization.

The foundational Student-Teacher framework [10] employs a ResNet-18 [17] teacher and a symmetric student to reconstruct intermediate features, with anomaly heatmaps derived from layer-wise Euclidean distances. STFPM [11] improves this using feature pyramid matching and multi-scale cosine similarity.

Later studies advanced both architecture and modeling strategies. FAST-DAD [12], for instance, introduces frequency-domain alignment, applying Fourier transforms and selecting dominant frequency bands to guide anomaly-aware reconstruction. RDMS [18] adopts an asymmetric reverse decoder, enabling the student to reconstruct deep semantic features and improving fine-grained anomaly detection.

Efforts have also targeted semantic consistency and robust reconstruction. Multi-student architectures model structured anomalies across spatial regions, while cross-layer residual guidance helps the student progressively align with the teacher's deep features, enhancing localization for low-contrast or blurred anomalies.

Despite these gains, current distillation methods face key challenges. Most adopt fixed-layer alignment, lacking adaptive mechanisms to select the most discriminative semantic layers. Additionally, they focus primarily on spatial-domain modeling, overlooking the frequency domain's value for capturing fine-grained cues. Finally, the absence of robust multi-scale feature fusion limits performance on anomalies with diverse shapes and scales.

Thus, integrating frequency-aware modeling, cross-scale feature fusion, and adaptive semantic selection into distillation frameworks is essential for advancing unsupervised anomaly detection,particularly for industrial inspection tasks with complex textures and subtle defects, such as fabric surface analysis.

In addition to distillation-based approaches, memory-bank-based methods such as PatchCore [19] and CFA [20] have shown strong performance on standard anomaly detection benchmarks. These methods leverage pre-trained feature extractors and store representative normal features for inference-time comparison. While effective, their reliance on large memory banks and distance-based scoring differs from our end-to-end distillation framework. We refer readers to these works for complementary perspectives on unsupervised anomaly detection.

2.3. Frequency Domain Feature Modeling

Spatial-domain anomaly detection often struggles with noise and repetitive textures, leading to poor performance on low-contrast or blurred anomalies. In contrast, frequency-domain modeling offers a robust alternative by separating low- and high-frequency components, thereby enhancing anomaly localization.

Recent studies highlight a spectral bias in deep networks, where models tend to learn low-frequency features more readily. This underscores the importance of high-frequency components in capturing fine-grained details critical for defect detection. Consequently, adaptive frequency filters, both low-pass and high-pass, have been introduced to improve boundary sensitivity and reduce aliasing.

In anomaly detection, frequency modeling has gained traction. FAST-DAD [12] aligns teacher-student features in the frequency domain to suppress high-frequency background interference. Similarly, the Frequency-Aware model [21] constrains feature fusion through frequency responses to emphasize structural variation in defects. Other methods integrate frequency-aware filters into convolutional modules, enabling multi-scale frequency extraction for complex textures.

However, existing methods often suffer from shallow integration and limited cross-scale guidance, with increased computational cost hindering practical deployment.

To address these issues, we propose a lightweight, deeply integrated frequency mechanism that enhances high-frequency anomaly perception while suppressing structural noise, without significantly increasing complexity. This improves robustness and accuracy in textured defect analysis.

3. Method

To address the challenges of fabric defect detection-such as subtle anomaly details, diverse defect morphologies, and the scarcity of labeled samples-we propose a novel Reverse Distillation Framework for unsupervised anomaly detection. This approach integrates unsupervised feature learning, frequency-domain modeling, and multi-scale semantic enhancement strategies, enabling effective extraction, reconstruction, and identification of anomalous regions without reliance on defect annotations.

As illustrated in Figure 2, the proposed framework consists of four key components: a pretrained teacher encoder (Encoder, E), a OCBE module,a student decoder (Decoder, D), and a FDFM.

Together, these modules form a cohesive architecture designed to distill rich semantic features from the teacher network, guide the student network's reconstruction process, and enhance anomaly localization through frequency-aware, multi-scale representation.

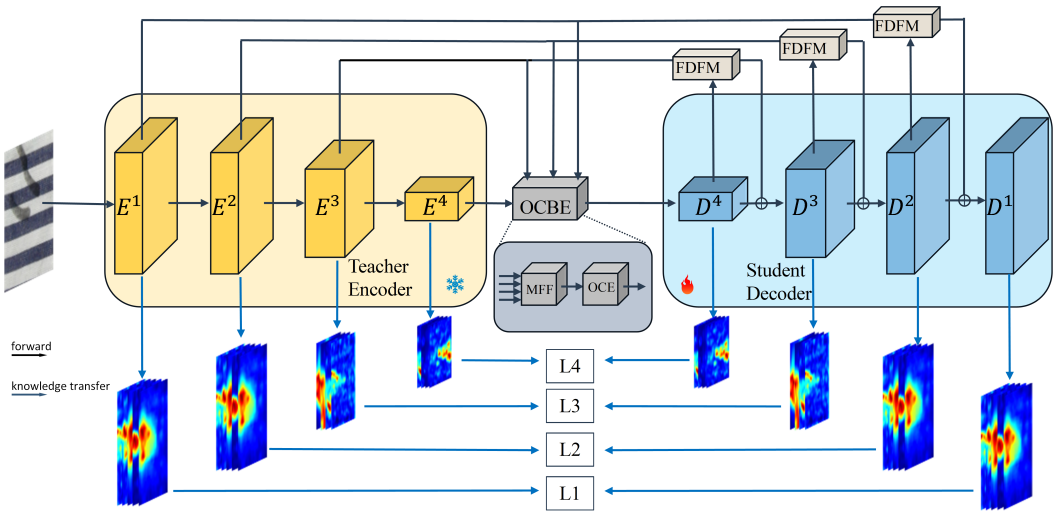


Figure 2. The architecture comprises a pretrained teacher, a trainable multi-level fusion module (OCBE), a Frequency-Decoupled Feature Fusion Module (FDFM), and a student decoder.

3.1. Reverse Distillation

The proposed reverse distillation mechanism aims to model the implicit distribution of normal samples by establishing an asymmetric feature mapping between a teacher and a student network. Unlike conventional knowledge distillation paradigms, which typically adopt symmetric pathways, FAD-RNet is based on an encoder-decoder architecture that introduces an asymmetric design of forward encoding and reverse decoding, thereby enhancing the network's sensitivity to anomalous regions.

In traditional distillation frameworks, both teacher and student networks usually share a similar bottom-up feature extraction path, progressively aggregating semantic information from low-level features. In contrast, FAD-RNet approach preserves the conventional bottom-up multi-scale encoding in the teacher network, while the student network performs top-down decoding from high-level teacher features to reconstruct low-level

feature representations. This design encourages the student to focus more on the recovery of fine-grained texture details, which is crucial for preserving subtle defect cues.

The overall objective can be formulated as a minimization function:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{feat}}\left(F_t^l, F_s^l\right) + \lambda \mathcal{L}_{\text{recon}}(x, \hat{x}) \quad (1)$$

where F_t^l and F_s^l denote the multi-scale features extracted by the teacher and the reconstructed features produced by the student at layer l, respectively. $\mathcal{L}_{\text{feat}}$ represents the feature alignment loss to measure their similarity, $\mathcal{L}_{\text{recon}}$ is the pixel-wise reconstruction loss, and λ are the corresponding weighting factors.

For the teacher encoder, we use WideResNet-50 [22] as the backbone. Its parameters are kept constant during training to avoid semantic drift. To evaluate the robustness of FAD-RNet across different model capacities, we also construct a lightweight variant using ResNet-34 [17] as the backbone. Both variants are trained using ImageNet pre-trained models as teacher models within the same distillation framework.

The teacher's multi-scale features are compressed via OCBE module into a compact representation, which serves as the input to the student decoder D. The student network performs hierarchical reconstruction in reverse order, with each decoding layer updated as:

$$F_s^l = \text{Deconv}_{3\times3}(F_s^{l+1} \oplus \mathcal{F}_{\text{FDFM}}(F_t^l, F_s^{l+1})) \quad (2)$$

where $\mathcal{F}_{\text{FDFM}}$ denotes the Frequency-Decoupled Feature Fusion module, elaborated in Section 3.3.

To further enhance the student's ability to model normal patterns, we introduce two additional losses: a feature matching loss and a structure-aware reconstruction loss. The feature matching loss enforces consistency between the normalized reconstructed features and the teacher's features:

$$\mathcal{L}_{\text{align}} = \frac{1}{CHW}\sum_{c,h,w}\left\| \frac{F_t^{(c,h,w)}}{\|F_t\|} - \frac{F_s^{(c,h,w)}}{\|F_s\|} \right\|^2 \quad (3)$$

where c, h, w denote the number of channels, height, and width of the feature maps. By calculating the differences pixel by pixel for the normalized features, it ensures that the student network models normal samples accurately in the feature space.

In addition, to improve the student's capability in modeling structural context, we introduce the SSPCAB module, which combines masked convolutions, dilated convolutions, and channel attention mechanisms to formulate a self-supervised reconstruction task. Given the input feature X, the SSPCAB generates a reconstruction $\widehat{X}$, optimized via a mean squared error loss:

$$\mathcal{L}_{\text{recon}} = \frac{1}{C \cdot H \cdot W}\sum_{c,h,w}(X_{c,h,w} - \widehat{X}_{c,h,w})^2 \quad (4)$$

This design improves anomaly detection in two complementary ways: (1) the reverse decoding path forces the student to explicitly learn a nonlinear inverse mapping from semantics to texture; (2) the cross-layer residual connections inject high-frequency information, which enhances the network's sensitivity to subtle anomalies.

3.2. OCBE

Within the reverse distillation framework, the OCBE module serves as a pivotal bridge between the teacher encoder and the student decoder. Its primary objective is to mitigate the reconstruction difficulty arising from semantic disparities between the two networks. Since the high-level features extracted by the teacher encoder are semantically abstract and sparse

in spatial detail, directly feeding them into the student decoder may lead to inadequate recovery of low-level textures, thereby compromising anomaly localization accuracy.

To address this issue, the OCBE is designed as a feature compression structure that integrates both multi-scale fusion and semantic enhancement, providing intermediate representations that are more suitable for decoder reconstruction. As illustrated in Figure 3a, OCBE comprises two key submodules: a MFF block and a CSA module.

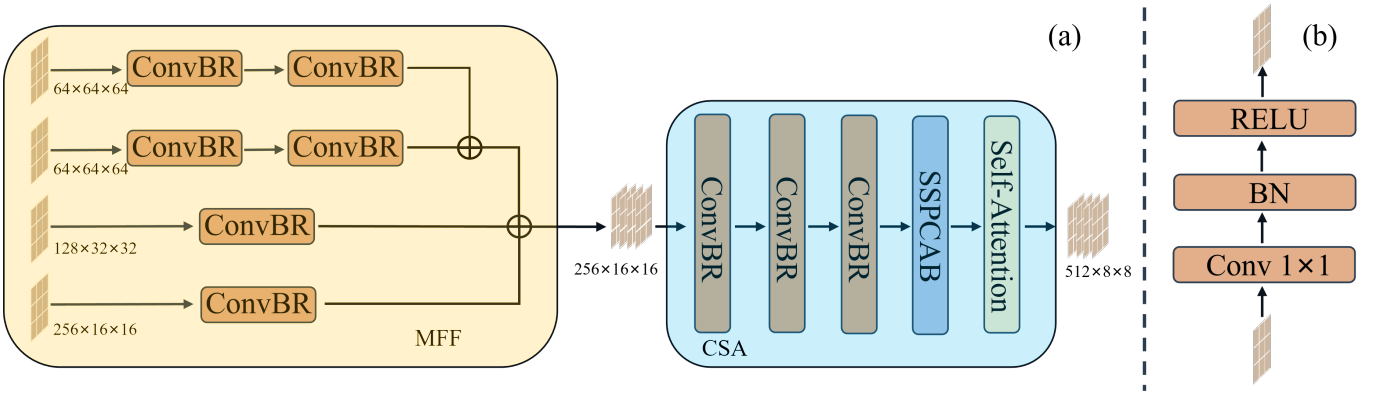


Figure 3. (**a**) is a schematic diagram of the OCBE module.Comprises trainable MFF and CSA blocks. MFF aligns teacher E's output with multi-scale features, which CSA compresses into compact bottleneck features. (**b**) is a schematic diagram of the ConvBR module, consisting of a 1 × 1 convolution, normalization, and relu activation function.

The MFF module integrates multi-level features from the teacher encoder to generate a bottleneck representation that balances spatial precision with semantic richness. Specifically, it receives feature outputs from four stages of the teacher encoder: the lower-level features preserve fine-grained textures and higher-level features encapsulate abstract semantics. Due to inconsistencies in spatial resolution and channel dimensions across these stages, To address spatial and channel inconsistencies. MFF employs ConvBR blocks-each composed of a 1 × 1 convolution, Batch Normalization and ReLU activation-as shown in Figure 3b,The first two inputs are downsampled and channel-aligned using two sequential ConvBR, while the latter two are processed by a single ConvBR. $\{F_t^l\}_{l=1}^{L}$ denote the selected teacher features. The aligned features are aggregated using channel-wise weighted summation:

$$F_{mff} = \sum_{l=1}^{4} Conv_{3\times3}^{BN+ReLU}(F_t^l) \tag{5}$$

where $Conv_{3\times3}^{BN+ReLU}(\cdot)$ denotes the convolutional transformation applied to each input path.

The fused feature is then fed into the CSA module for semantic enhancement. The CSA consists of three sequential ConvBR blocks, followed by a Self-Supervised Predictive Context-Aware Block (SSPCAB) and a lightweight self-attention mechanism.

The SSPCAB incorporates two key components: masked convolution and channel attention. The masked convolution applies a random binary mask $M \in \{0,1\}^{H\times W}$ to the input features F_{mff}, simulating occluded contexts by selectively zeroing out spatial regions. This encourages the model to learn context-aware reconstruction from the visible regions:

$$F_{mask} = F_{mff} \odot M \tag{6}$$

where $\odot$ denotes element-wise multiplication. This operation enhances the network's ability to infer spatially missing structures, which is essential for modeling local context and detecting subtle anomalies. Following this, a Squeeze-and-Excitation (SE) mechanism is employed to recalibrate semantic responses across channels. Each channel undergoes global average pooling, followed by two fully connected (FC) layers and a sigmoid activation to produce attention weights:

$$\alpha = \mathrm{F}_{\mathrm{mask}} \cdot \sigma(W_2 \cdot \delta(W_1 \cdot GAP(\mathrm{F}_{mask}))) \tag{7}$$

where W_1, W_2 are the FC layer weights, δ is ReLU, σ is the sigmoid function, and GAP denotes global average pooling.

To further calibrate global semantics and reinforce abnormal region activation, we introduce a lightweight self-attention mechanism after the SSPCAB. This attention block employs two consecutive 3×3 convolutions and a ReLU activation to generate a spatial attention map, which is then used to reweight the features. Despite its simplicity, this module effectively complements SSPCAB's local modeling by emphasizing discriminative spatial regions.

3.3. FDFM

3.3.1. Residual Feature Injection

In traditional decoder architectures, shallow-layer features are prone to semantic degradation due to repeated upsampling and convolution operations. This issue is exacerbated in reverse distillation frameworks, where the student network lacks strong high-level semantic guidance, making it difficult to recover low-level texture details accurately. To address this challenge, we propose a FDFM module, as shown in Figure 4, which introduces a residual feature injection mechanism. This mechanism builds cross-layer connections between the teacher encoder and the student decoder to mitigate information loss and enhance fine-grained reconstruction.

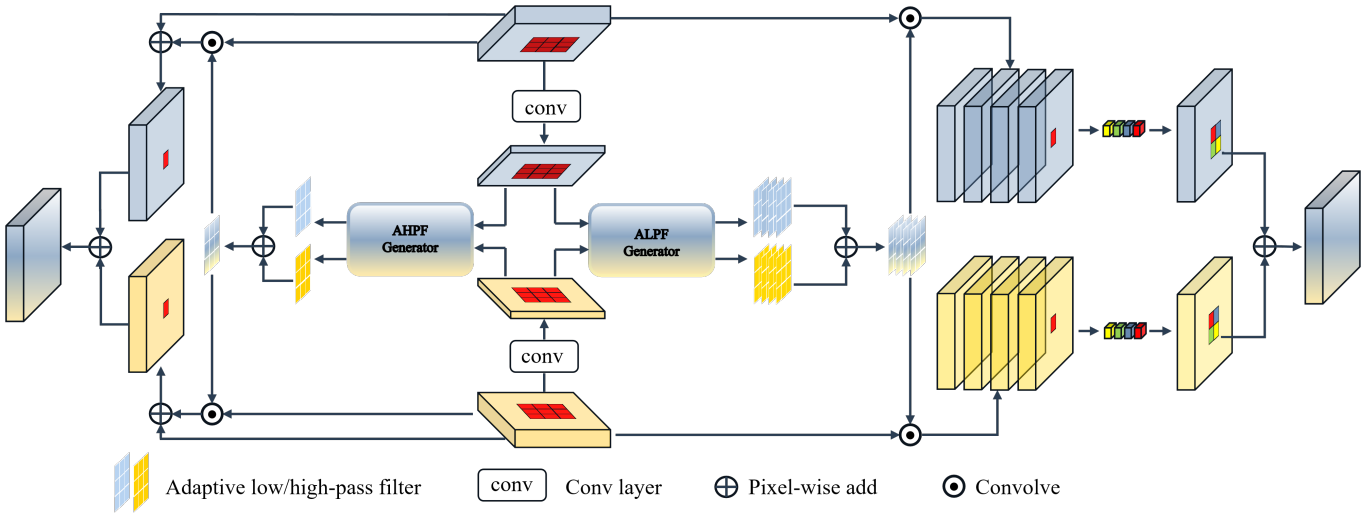


Figure 4. The FDFM module performs frequency-aware feature enhancement through adaptive filtering.

Specifically, for each decoding stage of the student network, we fuse the current decoding features with the corresponding teacher features from other semantic layers to generate the input for the next decoder layer. This design not only supplements the student with structurally consistent semantic context but also introduces a residual supervision path anchored on teacher features, enhancing sensitivity to anomalies in fine textures.

Before fusion, if the teacher features and student features differ in spatial resolution, we first downsample the teacher features to match the size of the student features:

$$\mathrm{Y} = \mathrm{Downsample}(\mathrm{Y}) \quad \text{if } \mathrm{size}(\mathrm{Y}) \neq \mathrm{size}(\mathrm{X}) \tag{8}$$

Next, both aligned teacher and student features undergo channel compression and frequency decomposition to extract high- and low-frequency components:

$$\begin{aligned} \text{high_fiff} &= \mathcal{F}^{hp}(Conv_{1\times1}(\mathrm{X}+\mathrm{Y})) \\ \text{low_fiff} &= \mathcal{F}^{lp}(Conv_{1\times1}(\mathrm{X}+\mathrm{Y})) \end{aligned} \tag{9}$$

where $Conv_{1\times1}$ denotes a channel compression layer, and $\mathcal{F}^{\mathrm{lp}}$ and $\mathcal{F}^{\mathrm{hp}}$ represent the low-frequency and high-frequency feature extractors, where their filter kernel sizes are 3×3 and 5×5, respectively.

To preserve structural consistency while emphasizing subtle variations, we perform channel-wise normalization and apply frequency-guided enhancement:

$$\begin{aligned} \mathrm{X}_{\mathrm{high}} &= \mathrm{X} + \mathrm{X} - \mathrm{CARAFE}(\mathrm{X}, \mathrm{Norm}(\text{high_fiff})) \\ \mathrm{X}_{\mathrm{low}} &= \mathrm{CARAFE}(X, \mathrm{Norm}(\text{low_fiff})) \\ \mathrm{Y}_{\mathrm{high}} &= \mathrm{Y} + \mathrm{Y} - \mathrm{CARAFE}(\mathrm{Y}, \mathrm{Norm}(\text{high_fiff})) \\ \mathrm{Y}_{\mathrm{low}} &= \mathrm{CARAFE}(Y, \mathrm{Norm}(\text{low_fiff})) \end{aligned} \tag{10}$$

where Norm represents normalization, CARAFE is a content-aware upsampling operator. Its kernel weights are dynamically predicted by the input features, $\mathrm{X}_{\mathrm{high}}$ and $\mathrm{Y}_{\mathrm{high}}$ represents the teacher features and student features after high-frequency enhancement, $\mathrm{X}_{\mathrm{low}}$ and $\mathrm{Y}_{\mathrm{low}}$ represents the teacher features and student features after low-frequency compensation.

Ultimately, the enhanced feature representation is generated through the attention mechanism:

$$Z = \sigma(\mathrm{X}_{\mathrm{high}} + \mathrm{Y}_{\mathrm{high}} + \mathrm{X}_{\mathrm{low}} + \mathrm{Y}_{\mathrm{low}}) \tag{11}$$

This dual-path frequency fusion greatly improves the student network's ability to distinguish subtle texture differences between normal and abnormal regions, especially in complex or repetitive background settings.

Notably, the FDFM module is instantiated at multiple decoder stages, with each stage employing an independent FDFM tailored to the corresponding semantic level. The parameters of FDFM are not shared across scales, allowing each instance to specialize in fusing features of different resolutions and channel depths. This hierarchical design facilitates progressive enhancement of multi-scale anomaly sensitivity while maintaining architectural consistency.

Figure 5 illustrates the feature transformation process within our proposed FDFM module. From left to right, each column respectively shows: (1) the original teacher features extracted at three different levels, (2) the student features from corresponding cross-level layers, (3) the frequency-enhanced features produced by FDFM via high-pass and low-pass filtering, and (4) the final fused features obtained by adding the student features with the frequency-enhanced outputs.

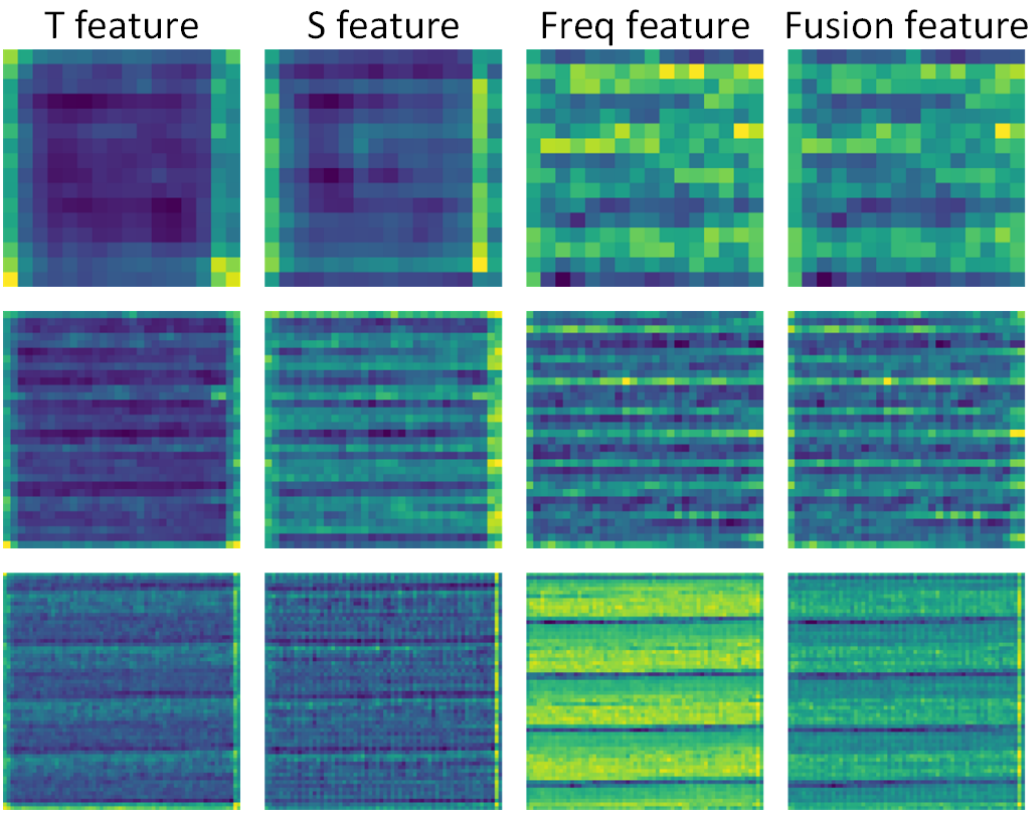


Figure 5. Feature visualization of FDFM fusion across layers. From left to right, each column shows the teacher features, student features, frequency-enhanced features, and fused features, respectively. The feature maps are visualized using a color map, where higher activation values are represented by warmer colors and lower values by cooler colors.

3.3.2. Frequency Domain Processing Theory

To further enhance the representational quality of fused features and boost the model's sensitivity to anomalous regions, we incorporate a frequency-domain analysis mechanism into the FDFM module. Inspired by Nyquist sampling theory and multi-scale image analysis, this design leverages adaptive spatial filtering to decouple and control frequency components, aiming to compensate for frequency information loss caused by downsampling operations in convolutional neural networks.

In CNNs, repeated downsampling progressively eliminates high-frequency details. According to the Nyquist theorem, when the sampling rate is halved, frequency components above the Nyquist frequency will alias into the lower spectrum, making the high-frequency content irreversibly lost. This aliasing leads to blurred boundaries and inaccurate localization of anomalies, especially under high-texture or low-contrast conditions.

To verify this issue, we perform discrete Fourier transform on intermediate feature maps to analyze their spectral distribution, defined as:

$$\mathrm{F}(u,v) = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} f(x,y) \cdot e^{-j2\pi\left(\frac{ux}{H}+\frac{vy}{W}\right)} \tag{12}$$

where H and W are the height and width of the feature map, and (u, v) denotes frequency coordinates. Our analysis reveals that deep semantic features tend to capture low-frequency, smooth components, while lacking sensitivity to edges and textures, which significantly limits their ability to detect fine-grained or boundary-related anomalies-common in industrial defect detection tasks.

To suppress high-frequency noise in semantic features while retaining smooth structure in low-frequency areas, we design an ALPF Generator that dynamically predicts spatially varying smoothing kernels for each location. This enables region-aware filtering that adapts to local content variations in the feature map.

Specifically, ALPF applies a 3×3 convolution layer to generate kernel response maps $\widehat{W}_{init} \in \mathbb{R}^{H \times W \times k^2}$ with $k = 3$ (i.e., $k^2 = 9$), followed by a softmax normalization to construct position-wise adaptive low-pass kernels :

$$\overline{W}_{p,q}(i,j) = \frac{\exp(\widehat{W}_{init}(p,q,i,j))}{\Sigma_{(i',j') \in \mathcal{N}} \exp(\widehat{W}_{init}(p,q,i',j'))} \tag{13}$$

The smoothed feature map is then obtained through kernel-weighted convolution:

$$\mathrm{F}_{\mathrm{low}}(p,q) = \sum_{(i,j) \in \mathcal{N}} \overline{W}_{p,q}(i,j) \cdot \mathrm{F}(p+i, q+j) \tag{14}$$

To preserve spatial details after low-pass filtering, ALPF employs sub-pixel upsampling, which expands the channel dimension and rearranges spatial positions to reconstruct high-resolution representations efficiently. While low-pass filtering suppresses noise effectively, it may also blur edge structures critical for anomaly detection. To mitigate this, we introduce an AHPF Generator to recover high-frequency details lost during downsampling.

AHPF shares a similar architecture: it applies a 5×5 convolutional layer followed by softmax to predict low-pass kernels $\overline{W}_{p,q}$, the number of output channels is consistent with the input feature dimension (C), and both AHPF and ALPF are implemented as lightweight convolutional layers trained end-to-end via backpropagation and computes high-pass kernels by subtracting them from a normalized identity kernel E with all weights set to $[[1,0,0],[0,1,0],[0,0,1]]$:

$$\widehat{W}_{p,q} = E - \overline{W}_{p,q} \tag{15}$$

Then, the high-frequency enhanced feature is obtained as:

$$\mathrm{F}_{\mathrm{high}}(p,q) = \sum_{(i,j) \in \mathcal{N}} \widehat{W}_{p,q}(i,j) \cdot \mathrm{F}(p+i, q+j) \tag{16}$$

Finally, the frequency-aware fused feature is obtained by aggregating the low- and high-frequency responses:

$$\mathrm{F}_{\mathrm{enhanced}} = \mathrm{F}_{\mathrm{high}} + \mathrm{F}_{\mathrm{low}} \tag{17}$$

Both the ALPF and AHPF modules are implemented as lightweight learnable components and optimized end-to-end within the overall distillation framework. By combining ALPF and AHPF, FDFM simultaneously performs low-frequency smoothing and high-frequency enhancement, significantly improving the network's capacity to capture subtle textures and edge anomalies. Owing to its spatial adaptivity, this dual-filtering mechanism dynamically adjusts filter responses based on local image content, providing precise and structure-aware features to support the generation of accurate pixel-level anomaly heatmaps.

During inference, anomaly regions are defined as areas that significantly deviate from the learned normal feature distribution. This deviation is quantified by the reconstruction discrepancy between teacher and student features, where lower similarity indicates higher anomaly likelihood. The frequency-aware mechanism further suppresses regular structural variations, improving robustness in complex textures.

4. Experiment

4.1. Experimental Setup

4.1.1. Dataset

According to the official ZJU-Leaper protocol, the 19 texture categories are divided into five groups based on structural complexity and visual similarity: the first group contains simple patterns (e.g., solid colors, stripes); the second group contains small repeating structures (e.g., dots, squares); the third and fourth groups contain more complex pattern arrangements (e.g., checkered patterns, houndstooth, irregular prints); the fifth group was excluded from our experiments due to low data quality (inconsistent lighting and annotation noise). Therefore, we focus on the first four groups, which contain a total of 15 texture-rich categories, to evaluate the robustness of our method under complex conditions. We use the ZJU-Leaper dataset [23] to evaluate our method. This dataset is a large-scale benchmark dataset for industrial fabric defects containing 19 texture categories. All experiments were conducted on the official version without any modifications and strictly followed standard evaluation procedures. The classification and ranking in the experiments strictly followed the original dataset documentation, where samples were grouped by texture type and sorted in ascending order of pattern complexity. All samples were from real production environments and included pixel-level masks and image-level annotations. To test the robustness of the model under complex conditions, we focused on 15 textured categories exhibiting a wide range of structural features, including plain weave, twill weave, multi-arm weave, and jacquard fabrics. Patterns range from simple periodic patterns (e.g., pinstripes, checks) to highly irregular patterns (e.g., knot patterns, floral prints), and color patterns range from solid colors to stripes, checkerboards, and multi-color prints. While detailed fiber composition information is not publicly available, this structural diversity collectively constitutes the challenge of this benchmark.

These defect types represent typical real-world anomalies commonly encountered in industrial textile inspection scenarios, including stains, damage (e.g., holes and broken yarns), wrinkles, and oil stains, as shown in Figure 6. These defects exhibit a variety of visual characteristics, such as irregular shapes, slight structural deformations, and low contrast with complex textures.



Figure 6. Representative defect types in the ZJU-Leaper dataset, including stain, damage (e.g., holes or broken yarns), wrinkle, and oil contamination. These defects exhibit significant variability in texture, shape, and contrast.

An example of a normal fabric sample for each category is shown in Figure 7.



Figure 7. Sample images of 15 classes in the ZJU-leaper dataset.

Images are resized to 256 × 256, enhance data by rotation, flipping and mirroring, and split 8:2 for training and testing. Models train for 200 epochs (Adam, lr = 0.005, batch = 64). Final anomaly maps are smoothed via Gaussian filtering.

4.1.2. Experimental Environment Configuration

All experiments run on an Ubuntu 20.04 system with an Intel Xeon E5-2678 v3 CPU, NVIDIA Tesla V100 (32 GB VRAM), and 256 GB RAM. Models are implemented in PyTorch 2.1.2 (Python 3.8, CUDA 11.8) within isolated Conda environments, with fixed random seeds ensuring reproducibility. Training and inference leverage single-GPU acceleration for optimal throughput.

4.1.3. Evaluation Indicators

We evaluate fabric defect detection using three key metrics: Image-level AUROC assesses normal and anomaly image classification, with values near 1 indicating optimal performance; Pixel-level AUROC extends this to localized anomaly identification in complex textures; and Pixel-level AUPRO quantifies spatial overlap between predictions and ground truth while equally weighting differently sized regions. Crucially, AUPRO calculates prediction reality overlap at 30% false positive rate, providing a robust measure for subtle defects under class imbalance where higher values denote superior fine-grained detection capability.

4.2. Quantitative Results Analysis

To evaluate the performance of the proposed method on fabric defect detection, we compare it with several state-of-the-art unsupervised anomaly detection approaches, including DBFAD [24], RDMS [18], Anomaly Clip [25], DSR [26], EfficientAD [27], and RD4AD [28]. Experiments are conducted on the ZJU-Leaper dataset, which covers 15 representative fabric texture categories. We adopt image-level and pixel-level evaluation metrics, including AUCROC, PixROC, and PixPRO.

As shown in Table 1, the proposed method achieves superior image-level performance. The large model variant obtains the highest average AUCROC, slightly outperforming RD4AD and EfficientAD, and matching the performance of DSR. In contrast, methods like DBFAD and Anomaly Clip struggle with fine-grained defects under complex textures. Notably, even the lightweight variant shows strong competitiveness, surpassing most compared methods in both accuracy and inference speed.

Table 1. Comparison of AUCROC performance of each method on 15 types of fabric textures. The best results in each row are highlighted in bold.

Category/Method	DBFAD	RDMS	Anomaly Clip	DSR	EfficientAD	RD4AD	Small	Large
Thick Stripe	98.36	97.97	99.34	**99.96**	99.73	99.25	99.68	99.71
Thin Stripe	99.38	99.52	98.93	**99.81**	99.12	99.48	98.57	99.76
White Plain	99.17	99.41	96.05	99.60	99.29	99.49	99.54	**99.67**
Dot Pattern	99.69	99.71	99.06	99.70	99.78	99.3	**99.94**	99.79
Gingham	98.29	98.54	94.78	98.82	99.2	99.12	97.3	**99.68**
Houndstooth	97.71	98.59	95.47	98.82	97.7	**99.23**	98.86	**99.23**
Knot Pattern	96.57	99.39	98.57	99.41	**99.83**	99.3	99.56	99.59
Blue Plaid	97.02	98.03	95.63	98.77	97.57	**98.92**	97.28	98.33
Brown Plaid	96.31	98.51	94.82	98.65	98.39	**99.13**	97.8	98.87
Gray Plaid	96.00	98.22	92.24	93.54	95.36	**98.76**	97.52	98.47
Red Plaid	96.38	97.62	90.78	96.83	98.63	**99.31**	98.36	98.81
Twill Plaid	94.84	97.80	86.72	98.67	95.97	**98.91**	98.11	98.50
Floral Print1	93.28	91.58	97.44	**99.49**	98.53	98.43	93.53	97.69
Floral Print2	89.78	82.93	92.58	**99.44**	96.43	97.98	93.81	97.58
Floral Print3	84.03	82.73	91.75	**98.79**	96.96	95.47	90.87	96.80
Total Average	95.79	96.03	94.94	98.69	98.17	98.81	97.38	**98.83**

FAD-RNet demonstrates robustness across challenging categories such as Gingham, Gray Plaid, and Twill Plaid, highlighting its effectiveness in complex pattern scenarios and validating its classification capability.

Pixel-level results in Table 2 further support the effectiveness of our method. The large model leads in PixPRO and is slightly behind RD4AD in PixROC. Compared with DBFAD, it achieves gains of +6.18% and +11.83% in PixROC and PixPRO, respectively. Although the lightweight variant shows slightly lower scores, it still remains highly competitive and comparable to EfficientAD.

Table 2. A comparison of the detection performance of each method under PixROC and PixPRO on 15 datasets.The best results in each row are highlighted in bold.

Category/Method	DBFAD	RDMS	Anomaly Clip	DSR	EfficientAD	RD4AD	Small	Large
Thick Stripe	98.58/80.51	**99.04**/88.67	98.6/**93.3**	91.09/81.36	98.06/83.82	98.51/86.68	98.28/80.65	98.55/83.55
Thin Stripe	90.28/74.68	98.72/85.15	99.3/**95.5**	89.3/86.11	**99.49**/83.05	98.9/85.82	99.23/90.97	99.38/92.45
White Plain	97.24/86.27	**99.39**/92.3	98.8/**92.9**	94.27/86.36	98.26/72.92	98.86/92.41	99.00/86.70	99.28/88.92

Table 2. *Cont.*

Category/Method	DBFAD	RDMS	Anomaly Clip	DSR	EfficientAD	RD4AD	Small	Large
Dot Pattern	95.51/73.34	98.05/86.02	96.0/**86.9**	90.24/79.91	96.13/86.79	**98.98**/85.00	97.88/81.32	98.19/85.44
Gingham	90.99/80.01	98.12/87.78	95.9/86.7	92.5/77.13	97.99/78.98	**98.61**/**90.14**	97.85/84.82	98.52/87.77
Houndstooth	94.87/79.03	98.34/87.07	96.6/84.6	89.56/75.32	**98.65**/70.78	98.56/86.30	97.47/85.80	98.08/**88.30**
Knot Pattern	94.95/68.40	95.77/78.12	87.3/71.9	89.58/75.80	94.02/78.54	**98.78**/**82.41**	93.84/68.25	98.69/81.51
Blue Plaid	97.95/82.86	98.68/89.15	93.47/79.91	94.86/88.03	98.01/87.95	**98.80**/**90.18**	97.38/79.36	98.35/86.76
Brown Plaid	91.41/83.62	98.58/89.14	93.26/78.12	91.4/84.63	97.46/87.61	97.69/82.78	98.66/86.65	**98.83**/**89.52**
Gray Plaid	92.64/73.09	97.97/88.65	90.18/67.81	86.37/67.23	96.82/83.31	**98.22**/84.78	97.30/86.84	97.23/**88.89**
Red Plaid	89.29/73.96	97.4/84.77	90.92/75.22	91.75/78.86	96.96/84.92	97.41/79.87	97.84/83.23	**98.28**/**88.56**
Twill Plaid	92.21/70.33	97.76/86.59	89.33/71.02	94.22/87.21	96.59/78.17	**99.14**/**89.52**	96.90/81.27	97.13/85.76
Floral Print1	80.73/69.15	92.88/79.08	96.1/86.00	95.09/**86.72**	97.56/82.01	**95.67**/80.20	90.98/71.73	93.81/79.73
Floral Print2	87.48/61.98	97.12/83.04	96.0/85.00	96.68/**87.56**	97.18/76.99	**98.68**/87.46	95.29/73.74	97.91/85.36
Floral Print3	78.89/46.36	95.13/76.09	90.7/75.10	92.21/68.93	92.97/56.41	**96.81**/**76.99**	91.54/60.16	96.44/73.52
Total Average	91.53/73.57	97.53/85.51	94.16/82.00	91.94/80.71	97.08/79.48	**98.24**/85.37	96.93/80.10	97.71/**85.74**

Figure 8 illustrates the anomaly score distribution. Normal (blue) and anomalous (red) samples form distinct, non-overlapping regions, confirming the discriminative capability of our reverse distillation strategy. Particularly, high localization accuracy is achieved on categories like Gray Plaid, Houndstooth, and Dot Pattern, benefiting from our frequency-aware multi-scale fusion.

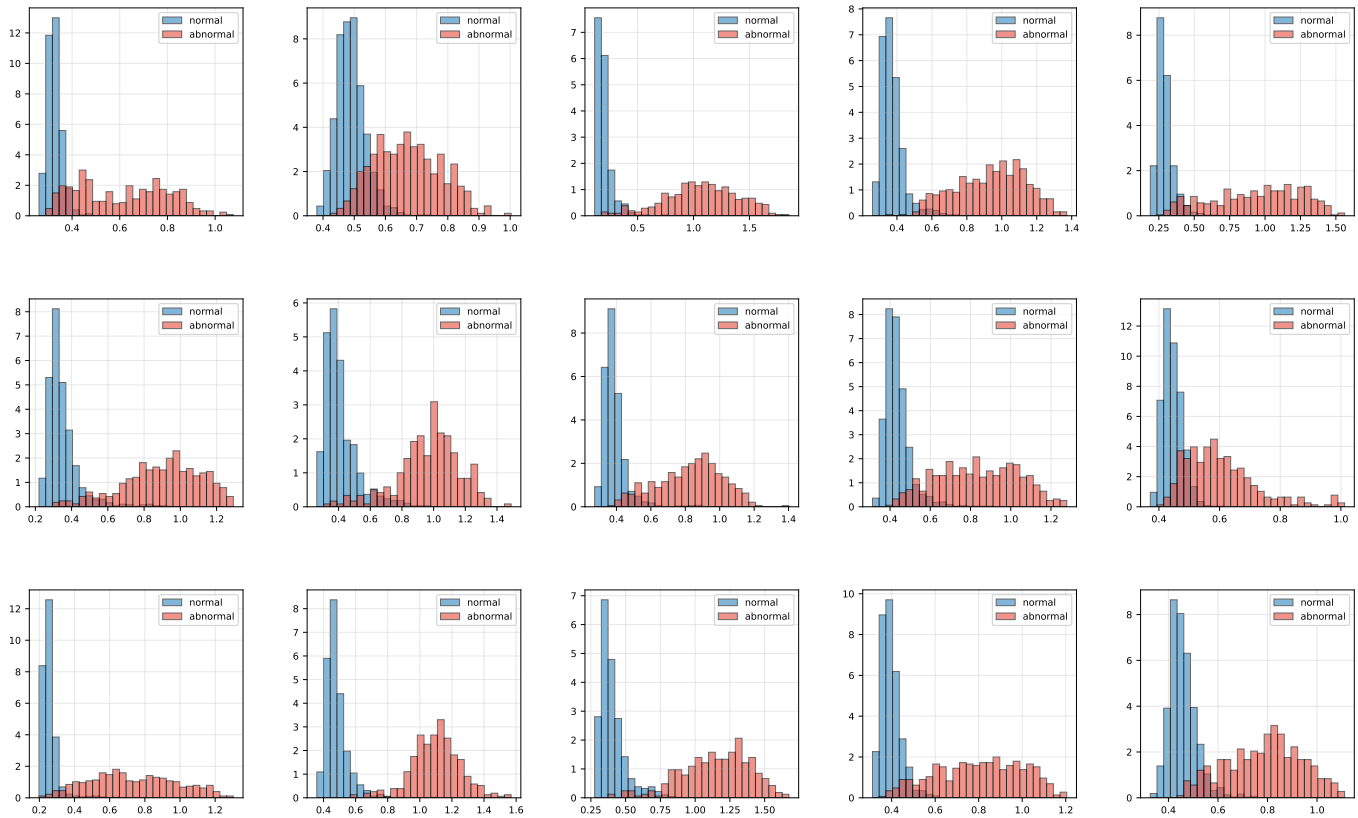


Figure 8. ZJU-Leaper anomaly score histograms for all categories.

The visual comparison in Figure 9 shows that our method precisely localizes anomalies of varying size, shape, and distribution. This robustness across different anomaly patterns reflects the method's generalization capability in real-world industrial applications.

By providing accurate and fine-grained localization-even for subtle and scattered defects-the proposed method proves to be both practical and versatile for deployment in industrial anomaly detection systems.

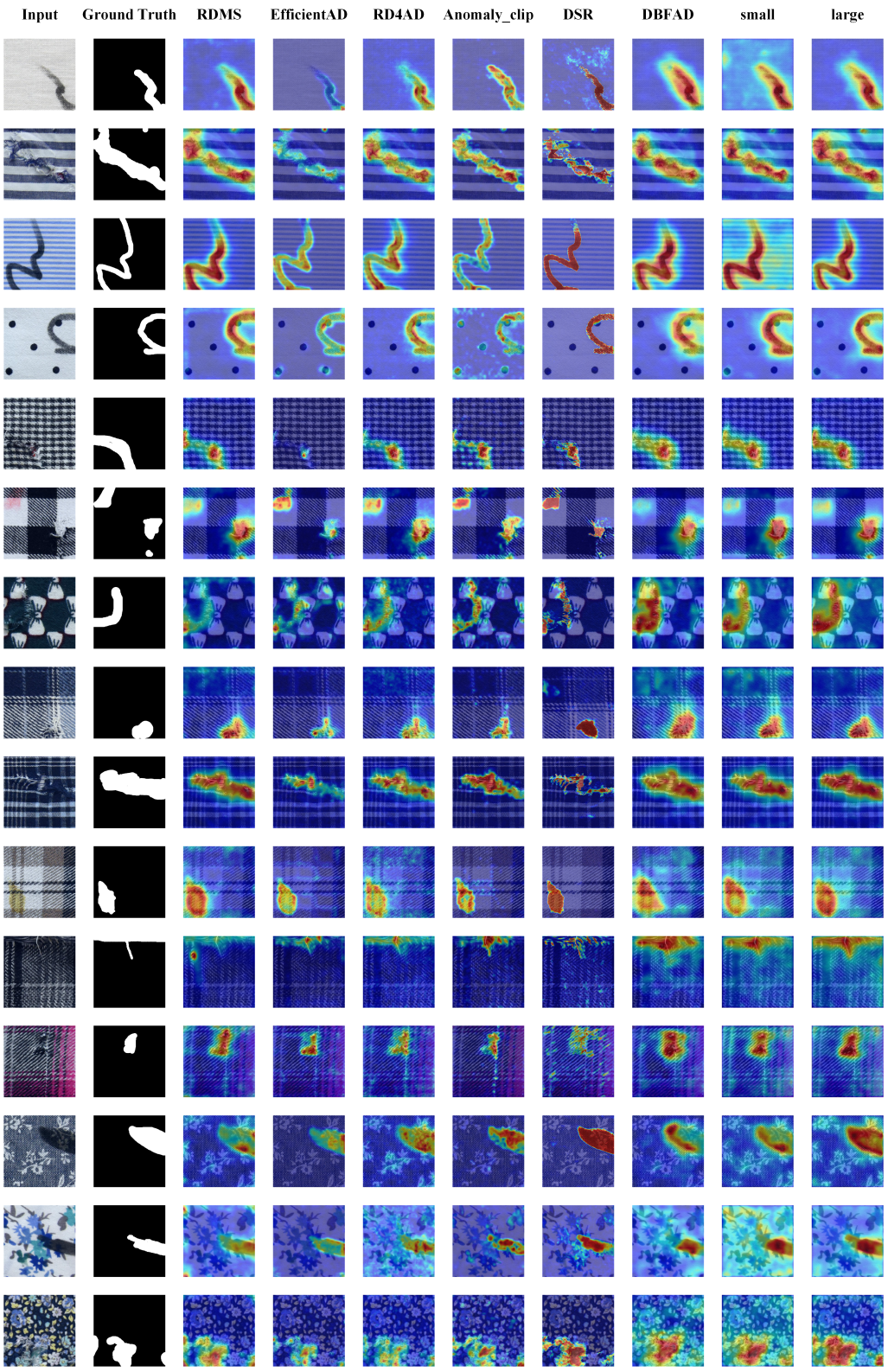


Figure 9. Visual comparison of the detection results of 15 categories in the ZJU-leaper dataset. The heatmaps represent the predicted anomaly scores, where warmer colors (e.g., red) indicate higher anomaly responses and cooler colors (e.g., blue) indicate lower anomaly responses.

4.3. Ablation Study

To further analyze the contribution of each module, we perform ablation studies by incrementally introducing the OCBE module, FDFM without frequency modeling, FDFM

with ALPF or AHPF only, and the full FDFM. Table 3 summarizes the results, where the ✓ and × indicate that the corresponding module is enabled or disabled, respectively.

The baseline model lacks both OCBE and FDFM modules. Adding OCBE alone yields slight improvements, confirming its role in enhancing spatial consistency. Introducing FDFM, even without frequency modeling, brings significant performance gains by improving texture representation and multi-scale fusion.

The full FDFM, with both ALPF and AHPF, further boosts performance. The lightweight variant incorporating complete FDFM achieves the best scores across all metrics, emphasizing the importance of frequency-aware design in structural reconstruction and boundary perception.

Table 3. The ablation experimental results of OCBE and FDFM modules and their frequency domain mechanisms on a small model.The best results in each row are highlighted in bold.

Model	w/RL	w/FD	Params	AUCROC	PixROC	PixPRO
baseline	×	×	4.0	95.00	95.22	73.50
+OCBE	×	×	10.1	95.71	96.10	73.77
+OCBE	✓	×	10.1	95.90	96.42	74.54
+FDFM	✓	×	10.3	95.98	95.41	76.43
+FDFM	✓	ALPF	10.4	96.72	96.15	78.62
+FDFM	✓	AHPF	10.4	96.85	96.38	79.07
+FDFM	✓	✓	10.5	**97.38**	**96.93**	**80.10**

We also investigate the effect of the self-supervised reconstruction loss by varying the loss weight λ (Table 4). A small λ (e.g., 0.01) leads to insufficient structural learning, while an overly large value (e.g., 1.0 or 10.0) dominates optimization and reduces anomaly sensitivity, especially in AUCROC. Optimal performance is observed at $\lambda = 0.1$, which balances structural fidelity and anomaly discrimination.

Table 4. Analysis of the Impact of the weight λ of the reconstruction loss term on Model Performance on Small models. The best results in each row are highlighted in bold.

λ	AUCROC	PixROC	PixPRO
1	96.84	96.70	78.47
0.01	96.20	94.94	74.18
10	96.85	96.17	78.64
0.1	**97.38**	**96.93**	**80.10**

4.4. Parameter and Time Cost Analysis

To assess the deployability of the proposed method in real-world industrial scenarios, we compare its computational efficiency with representative unsupervised anomaly detection approaches in terms of model size and inference speed, as shown in Table 5. All methods are evaluated on 100 randomly selected samples from the Thick Stripe subset of the ZJU-Leaper dataset, each tested five times to report average results.

DBFAD, though compact in trainable parameters, suffers from low inference speed (19.55 FPS), limiting its applicability in real-time settings. RD4AD, despite strong accuracy, has a large model size and moderate speed (38.62 FPS). In contrast, our small variant achieves both a lightweight model and the highest inference speed (41.14 FPS), surpassing most existing methods. EfficientAD provides a good balance with decent speed and size but compromises on accuracy due to its simple architecture. Our large variant, while heavier, maintains a solid 28.25 FPS thanks to the FDFM module's efficient frequency decoupling.

Overall, the proposed method balances accuracy, model size, and speed effectively. The lightweight variant, in particular, demonstrates strong potential for deployment in resource-limited industrial environments.

Table 5. A comparative analysis of the number of parameters and average reasoning speed of each method. The best results in each row are highlighted in bold.

Method	Tr.Params (M)	NonTr.Params (M)	Avg.FPS
EfficientAD	5.4	2.7	28.31
S-T	6.8	0	0.39
RDMS	96.5	24.9	34.35
Patchcore	0	68.9	6.74
DSR	36.2	4.0	11.03
RD4AD	92.2	68.9	38.62
DBFAD	10.8	8.2	19.55
small	10.5	8.2	**41.14**
large	187.1	24.9	28.25

5. Conclusions

We propose an efficient unsupervised fabric defect detection method using reverse distillation with feature reconstruction. The architecture integrates a teacher encoder, student decoder, bottleneck embedding module, and frequency-decoupled fusion module. This design enhances local structure recovery in complex textures through structure-aware reconstruction, while high-low frequency separation improves sensitivity to subtle anomalies.

On ZJU-Leaper, the large model excels in AUCROC and PixPRO, while the lightweight version maintains competitive accuracy for resource-constrained deployment. Ablations validate component synergy.

However, this study has several limitations. First, the method is evaluated on a single dataset, and its generalization ability to unseen fabric types or other industrial domains remains unverified. Second, the model currently relies on a static encoder without adaptive fine-tuning, which may limit its flexibility under domain shifts. Third, although the frequency-decoupled fusion improves anomaly localization, its computational design may require further optimization for real-time high-resolution deployment.Additionally, while our evaluation relies on pixel-level annotations from domain experts, direct physical measurement of detected defects—such as actual dimensions or severity grading—remains unexplored in this study. Incorporating such physical validation would be an important step toward practical industrial deployment.

Future work may explore the following directions: Incorporating cross-image memory mechanisms to improve global contextual contrast, particularly for hard-to-detect anomalies; integrating multi-modal information (e.g., infrared or reflectance maps) to enhance robustness and extend applicability in more complex industrial inspection settings.

Author Contributions: S.L. performed the data analysis; J.L. (Jun Liu) performed the formal analysis; J.L. (Jiuzhen Liang) performed the validation; H.L. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no competing interests.

References

1. Hütten, N.; Alves Gomes, M.; Hölken, F.; Andricevic, K.; Meyes, R.; Meisen, T. Deep learning for automated visual inspection in manufacturing and maintenance: A survey of open-access papers. *Appl. Syst. Innov.* **2024**, *7*, 11. [CrossRef]
2. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. MVTec AD–A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9592–9600.
3. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [CrossRef]
4. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Langs, G.; Schmidt-Erfurth, U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Med. Image Anal.* **2019**, *54*, 30–44. [CrossRef] [PubMed]
5. Gong, D.; Liu, L.; Le, V.; Saha, B.; Mansour, M.R.; Venkatesh, S.; Hengel, A.v.d. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1705–1714.
6. Mishra, P.; Verk, R.; Fornasier, D.; Piciarelli, C.; Foresti, G.L. VT-ADL: A vision transformer network for image anomaly detection and localization. In *Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 01–06.
7. Zavrtanik, V.; Kristan, M.; Skočaj, D. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 8330–8339.
8. Li, C.L.; Sohn, K.; Yoon, J.; Pfister, T. Cutpaste: Self-supervised learning for anomaly detection and localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9664–9674.
9. Schlüter, H.M.; Tan, J.; Hou, B.; Kainz, B. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 474–489.
10. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4183–4192.
11. Wang, G.; Han, S.; Ding, E.; Huang, D. Student-teacher feature pyramid matching for anomaly detection. *arXiv* **2021**, arXiv:2103.04257. [CrossRef]
12. Fakoor, R.; Mueller, J.W.; Erickson, N.; Chaudhari, P.; Smola, A.J. Fast, accurate, and simple models for tabular data via augmented distillation. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 8671–8681.
13. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. Carafe: Content-aware reassembly of features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3007–3016.
14. Bergmann, P.; Löwe, S.; Fauser, M.; Sattlegger, D.; Steger, C. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv* **2018**, arXiv:1807.02011.
15. Defard, T.; Setkov, A.; Loesch, A.; Audigier, R. Padim: A patch distribution modeling framework for anomaly detection and localization. In *Proceedings of the International Conference on Pattern Recognition, Virtual Event, 10–15 January 2021*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 475–489.
16. Cohen, N.; Hoshen, Y. Sub-image anomaly detection with deep pyramid correspondences. *arXiv* **2020**, arXiv:2005.02357.
17. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009*; IEEE: Piscataway, NJ, USA, 2009; pp. 248–255.
18. Chen, Z.; Lyu, C.; Zhang, L.; Li, S.; Xia, B. RDMS: Reverse distillation with multiple students of different scales for anomaly detection. *IET Image Process.* **2024**, *18*, 3815–3826. [CrossRef]
19. Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; Gehler, P. Towards total recall in industrial anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 14318–14328.
20. Lee, S.; Lee, S.; Song, B.C. Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access* **2022**, *10*, 78446–78454. [CrossRef]
21. Chen, L.; Fu, Y.; Gu, L.; Yan, C.; Harada, T.; Huang, G. Frequency-aware feature fusion for dense image prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10763–10780. [CrossRef] [PubMed]
22. Zagoruyko, S.; Komodakis, N. Wide residual networks. *arXiv* **2016**, arXiv:1605.07146. [CrossRef]
23. Zhang, C.; Feng, S.; Wang, X.; Wang, Y. Zju-leaper: A benchmark dataset for fabric defect detection and a comparative study. *IEEE Trans. Artif. Intell.* **2021**, *1*, 219–232. [CrossRef]
24. Thomine, S.; Snoussi, H. Distillation-based fabric anomaly detection. *Text. Res. J.* **2024**, *94*, 552–565. [CrossRef]

25. Zhou, Q.; Pang, G.; Tian, Y.; He, S.; Chen, J. Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. *arXiv* **2023**, arXiv:2310.18961.
26. Zavrtanik, V.; Kristan, M.; Skočaj, D. Dsr–a dual subspace re-projection network for surface anomaly detection. In *Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 539–554.
27. Batzner, K.; Heckler, L.; König, R. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2024; pp. 128–138.
28. Deng, H.; Li, X. Anomaly detection via reverse distillation from one-class embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9737–9746.

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.