

Article

A Dual-Domain Reverse Distillation Algorithm for Unsupervised Industrial Surface Defect Detection: Application to Non-Woven Fabrics

Rong Lin Yan, Wei Wei *  and Zhen Huang 

School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan 430048, China

* Correspondence: wei.vision@whpu.edu.cn

Abstract

Industrial surface defect detection faces challenges of complex textures, diverse defect morphologies, and scarce labeled data, especially for non-woven fabrics. This paper proposes a dual-domain reverse distillation algorithm for unsupervised defect detection (DDRD). The algorithm integrates frequency-domain wavelet enhancement and spatial-domain self-attention to enhance defect features synergistically. A Wavelet High-frequency Deformable Enhancement Module amplifies fine-grained defect details, while a Convolutional Self-Attention Spatial Global Enhancement Module captures long-range spatial dependencies. The enhanced dual-domain features are embedded into a reverse distillation framework for end-to-end training, achieving precise defect localization via feature reconstruction errors. Extensive experiments are conducted on the self-built large-scale non-woven fabric WFB dataset and the public MVTec AD benchmark dataset. The results show that DDRD achieves 98.0% pixel-level AUROC and 93.0% image-level AUROC on the WFB dataset, outperforming the state-of-the-art RD4AD method by 0.5% and 1.5%, respectively. On the MVTec AD dataset, it attains an average of 97.5% pixel-level AUROC and 99.7% image-level AUROC, with perfect 100% image-level detection accuracy on multiple categories. These results validate the efficacy and robustness of the dual-domain enhancement strategy for industrial defect detection tasks.

Keywords: unsupervised defect detection; industrial surface defect detection; dual-domain feature enhancement; wavelet transform



Academic Editor: Atsushi Mase

Received: 31 May 2026

Revised: 8 July 2026

Accepted: 16 July 2026

Published: 23 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Non-woven fabrics are critical industrial materials with excellent mechanical and functional properties, widely used in medical protection, industrial filtration and other fields. During high-speed manufacturing, their surfaces are prone to various defects due to raw material and environmental disturbances [1]. These defects impair product quality and cause economic losses, so automated defect detection is essential for intelligent production line quality control.

Currently, non-woven fabric defect inspection mainly relies on manual observation and traditional machine vision methods. Manual inspection is inefficient and subjective, with high error rates that cannot meet high-speed production requirements [2]. Traditional methods rely on handcrafted shallow features, lacking robustness to complex random textures and uneven illumination, making it difficult to distinguish defects from normal backgrounds.

Deep learning has become the mainstream for industrial defect detection due to its powerful feature learning ability [3]. However, supervised methods require a large number of labeled defect samples, which are scarce and costly to collect for non-woven fabrics. Unsupervised learning, trained only on defect-free data, identifies anomalies via distribution discrepancies, effectively solving the sample scarcity problem and becoming a promising solution for fabric defect detection [4].

To address the limitations of manual inspection and supervised methods, this paper proposes an unsupervised defect detection method based on dual-domain feature enhancement and reverse distillation. The method learns normal texture representations and locates defects via reconstruction errors. Wavelet transform, with low complexity and high real-time performance, is introduced to enhance high-frequency defect features. By integrating frequency and spatial domain information, the method synergistically enhances defect representation, effectively adapting to complex textured backgrounds.

Existing methods mostly focus on spatial domain optimization, ignoring frequency domain characteristics of defects (high-frequency signals) and global structural information (low-frequency signals). This work breaks the single-domain constraint and proposes collaborative dual-domain enhancement. Two core modules are designed: a Wavelet High-frequency Deformable Enhancement Module and a Convolutional Self-Attention Spatial Global Enhancement Module. The proposed method adopts an unsupervised training strategy and embeds wavelet operations into the network for end-to-end optimization. As a result, it achieves high accuracy and robustness in non-woven fabric defect detection.

From a signal processing perspective, the surface texture and defect patterns of non-woven fabrics exhibit markedly different frequency-domain characteristics. The random fiber texture that constitutes the fabric background is spatially disordered yet statistically stationary, and its energy is broadly distributed across a wide frequency band without a dominant component, resulting in a noise-like broadband spectrum. In contrast, genuine defects are typically localized, sharp discontinuities in the fabric structure and are therefore concentrated in the high-frequency sub-bands (e.g., the LH, HL, and HH components under wavelet decomposition), although with distinct characteristic patterns across defect types: minor flaws (e.g., small holes or thin cracks) produce sparse but high-magnitude high-frequency responses due to abrupt fiber breakage, whereas low-contrast defects, despite exhibiting only subtle gray-level differences from the background in the spatial domain, still generate detectable high-frequency edge responses that would otherwise be masked by their low intensity contrast. Meanwhile, the coarse spatial layout of the fabric, such as large-scale structural distortion, is mainly encoded in the low-frequency sub-band. This frequency-domain decoupling implies that operating purely in the spatial domain conflates these markedly different signal components: high-frequency defect cues can be overwhelmed by the broadband spectral energy of random background texture, while low-contrast defects, although weak in spatial intensity, retain a distinguishable high-frequency signature that spatial-domain-only methods tend to overlook. This observation motivates the necessity of the proposed dual-domain fusion strategy: the frequency-domain branch (WHAEM) selectively amplifies the high-frequency components associated with fine-grained and low-contrast defects while suppressing broadband texture noise, whereas the spatial-domain branch (CSGEM) complements this by modeling long-range structural dependencies to further discriminate defects from complex background patterns. Only by jointly exploiting both frequency- and spatial-domain information can discriminative features be enhanced synergistically across these three representative defect scenarios.

The main contributions of this paper are summarized as follows:

- (1) This paper presents a Wavelet High-frequency Deformable Enhancement Module (WHAEM) for industrial defect detection. Conventional methods only optimize fea-

tures in the spatial domain and neglect frequency-domain properties of defects. As defects are high-frequency signals against low-frequency texture backgrounds, the proposed WHAEM adopts wavelet transform to enhance frequency-domain features and highlight subtle defect information. Unlike prior high-frequency enhancement approaches that apply fixed, spatially invariant filters uniformly across wavelet sub-bands, WHAEM introduces a deformable, content-adaptive enhancement mechanism that adjusts sampling positions and enhancement strength according to the local high-frequency response, thereby selectively amplifying genuine, including low-contrast, defect signals while suppressing background texture noise residing in the same frequency band.

- (2) A Convolutional Self-Attention Spatial Global Enhancement Module (CSGEM) is developed to capture long-range spatial dependencies. It improves spatial perception and defect localization. Combined with WHAEM, a dual-domain collaborative enhancement mechanism is formed to boost feature discrimination between defects and normal textures. Different from standard self-attention modules that model global dependencies at high computational cost, or conventional channel/spatial attention mechanisms (e.g., SE, CBAM) restricted to a fixed local receptive field, CSGEM integrates convolutional locality with self-attention's global modeling capacity through an adaptive residual fusion mechanism, capturing long-range structural dependencies while retaining fine-grained local sensitivity—an advantage that purely attention-based or purely convolutional spatial enhancement methods cannot simultaneously achieve.
- (3) Experiments on the self-built WFB dataset and public MVTec AD dataset show that the proposed dual-domain reverse distillation algorithm outperforms state-of-the-art unsupervised methods. The results validate the effectiveness, robustness and generalization of the dual-domain enhancement framework in industrial applications.

2. Related Work

Industrial surface defect detection, particularly for non-woven fabrics, is commonly addressed via supervised and unsupervised learning paradigms [5,6]. Deep learning has become a mainstream solution for fabric defect detection due to its powerful feature representation ability in handling complex textile textures [7,8]. A variety of supervised detection algorithms have been developed. Wang et al. proposed a CNN–Transformer hybrid model to balance local fine-grained feature extraction and global dependency modeling [9]. Zhang et al. constructed a causal inference-guided network to suppress background interference [10]. Guo et al. introduced a multi-scale deformable attention-based model for irregular defect localization [11]. Zhou et al. designed a cross-level feature enhancement framework to mitigate detection errors under complex textures [12]. Xiang proposed a lightweight multi-scale integration method for edge deployment [13]. Novakov et al. built a hybrid network integrating CNN, Transformer and Kolmogorov–Arnold Networks to enhance nonlinear feature fitting [14]. Despite their promising performance, supervised methods rely heavily on large-scale annotated defect samples, which are extremely scarce and costly to collect in industrial non-woven fabric production, severely restricting their practical deployment.

Unsupervised learning avoids dependence on labeled data, making it well-suited for industrial scenarios with limited defect samples. Knowledge distillation (KD) has emerged as a core unsupervised paradigm, which trains a student network to mimic the feature outputs of a teacher network on normal samples and identifies anomalies via inter-network feature discrepancies [15]. Liu et al. proposed the DAERD framework with dual-attention residual modules to improve feature reconstruction accuracy [16]. Pei et al. proposed a reverse knowledge distillation framework with a dual-teacher attentional fusion module

to improve the accuracy of industrial defect detection and localization [17]. Wan et al. designed a high-frequency feature mapping enhanced distillation method for tiny defect detection [6]. Zhou et al. proposed an Attention-based Multi-scale Dual-flow Reverse Distillation framework (AMDRD) to improve the accuracy of industrial anomaly detection and localization [18]. Tran et al. proposed the DM-GRD framework with dual memory banks and a multi-scale feature fusion module to guide reverse distillation for unsupervised anomaly detection [19]. Notably, DM-GRD shares the same reverse distillation baseline as our work but pursues a fundamentally different enhancement route. It equips the teacher–student pipeline with two external, updatable memory banks that explicitly store normal and abnormal feature prototypes retrieved from the teacher network, together with pseudo-anomaly images synthesized via Fourier phase mixing and Perlin noise, and enforces separation between the two memory banks through a dedicated triplet loss. This retrieval-based paradigm operates entirely in the spatial domain. In contrast, our method requires no external memory storage, no retrieval mechanism, and no synthetic anomaly generation. It instead enhances the intrinsic feature representation of the teacher–student pipeline itself by jointly exploiting frequency-domain and spatial-domain cues, motivated by the observation that defect signals in complex-textured surfaces are predominantly high-frequency in nature. This distinction in enhancement paradigm—external memory retrieval versus internal dual-domain representation enhancement—together with the difference in target application scenario (general-purpose industrial objects/textures for DM-GRD versus complex-textured non-woven fabrics in this work), constitutes the primary novelty of the proposed approach relative to [19].

These studies have collectively advanced the performance of unsupervised defect detection from different perspectives, such as attention design, multi-source fusion, and multi-scale modeling. Building on this progress, the present work further explores the frequency domain as a complementary direction, aiming to enhance the extraction of tiny, low-contrast defect features under complex random textures.

Beyond algorithmic performance, it is worth highlighting the industrial implementation potential of knowledge distillation-based anomaly detection frameworks. Such frameworks are particularly attractive for real-world deployment because they eliminate the need for large-scale defect annotation, which is often infeasible on active production lines. A typical industrial pipeline built upon such methods generally consists of three stages: (1) an image acquisition module, usually a high-resolution industrial area-scan camera synchronized with the production line via hardware triggering, capturing surface images of the material under controlled illumination; (2) an inference module, where the pre-trained teacher–student network runs on an embedded or edge-computing platform to generate real-time anomaly score maps; and (3) a decision module that thresholds the anomaly map to trigger alarms or reject defective products. This deployment paradigm is conceptually consistent with the automated optical inspection (AOI) systems widely adopted in textile and film production lines, where continuous, unsupervised quality monitoring is required without interrupting production throughput. Compared with supervised detectors, this pipeline substantially reduces the cost of dataset construction and model maintenance, as the model only needs to be retrained when the normal texture pattern of the product changes, rather than each time a new defect type appears.

Overall, existing deep learning-based fabric defect detection methods, whether supervised or unsupervised, mostly optimize features solely in the spatial domain and neglect the complementary value of frequency-domain information. Defects typically manifest as high-frequency signals relative to normal textures, while global structural information corresponds to low-frequency components. Integrating frequency-domain and spatial-domain information can effectively enhance the discrimination between defects and background. To

address these limitations, this paper proposes a dual-domain reverse distillation algorithm for unsupervised industrial surface defect detection. Two complementary enhancement modules—WHAEM for frequency-domain high-frequency feature enhancement and CSGEM for spatial-domain global feature modeling—are constructed and embedded into the reverse distillation framework. Benefiting from its unsupervised nature, the proposed algorithm inherently alleviates the sample scarcity issue, while WHAEM and CSGEM further boost detection accuracy for tiny defects under complex textured backgrounds.

3. Materials and Methods

3.1. Overall Framework

Figure 1 presents the architecture of the proposed dual-domain reverse distillation algorithm. The framework is developed based on reverse distillation and consists of a frozen teacher encoder E, WHAEM, OCBE, CSGEM, and a student decoder D. The teacher encoder extracts multi-scale features from input samples, and the OCBE module compresses these features into a low-dimensional embedding space that preserves only normal sample patterns. As a result, the student decoder can reconstruct only normal features and fails to restore features in defective regions. To overcome the limitation of relying solely on spatial-domain processing, we combine frequency-domain and spatial-domain information for joint feature enhancement. Specifically, WHAEM leverages wavelet transform to achieve frequency-domain enhancement, while CSGEM performs spatial-domain global feature modeling. The two modules cooperate to strengthen defect feature representation. Deployed on the branch of teacher encoder E, WHAEM amplifies feature differences between teacher and student networks to distinguish defects from normal textures. Embedded after the OCBE module, CSGEM optimizes global semantic features and further boosts the detection performance. At each of the three encoder stages, the WHAEM-enhanced teacher feature E_i is compared against the corresponding student-reconstructed feature D_i via cosine similarity to compute a scale-wise discrepancy term $L_i = 1 - \cos(E_i, D_i)$ ($i = 1, 2, 3$), as illustrated in Figure 1. These three terms are summed to form the total training loss and also serve as the basis for generating the multi-scale anomaly score map at test time.

The specific embedding positions of WHAEM and CSGEM are not arbitrary design choices, but are dictated by the underlying principle of reverse distillation: because the teacher-student pair is trained exclusively on defect-free samples, the student decoder can only faithfully reconstruct feature patterns that are representative of normal texture, whereas at test time it systematically fails to reconstruct patterns associated with unseen defects, and this failure manifests as a large reconstruction error that serves as the anomaly indicator. Both modules are designed to widen this training-inference discrepancy specifically at defective regions, but they act on different stages of the pipeline and therefore play complementary, non-interchangeable roles.

WHAEM is attached to the teacher encoder branch E, prior to the OCBE bottleneck, rather than to the student decoder D. Because the teacher branch defines the distillation target that the student is trained to reproduce, placing WHAEM at this stage embeds the frequency-domain enhancement directly into the target representation: during training, the student decoder learns to reconstruct the WHAEM-enhanced high-frequency patterns characteristic of normal fiber texture. At test time, when a defect is present, its high-frequency response is structurally inconsistent with the normal-texture patterns the student has learned, and this response is further amplified by WHAEM before being passed to the teacher-side target, while the student decoder, having never observed such patterns, cannot reproduce them—thereby widening the feature discrepancy specifically at defect locations. If WHAEM were instead placed on the student branch, the enhancement would act on the student's own reconstructed, already normal-biased features rather than on the

teacher's target, offering no additional discriminative signal and potentially amplifying reconstruction noise rather than genuine defect cues. Moreover, embedding WHAEM before, rather than after, the OCBE compression ensures that fine-grained high-frequency detail is enhanced while the multi-scale spatial resolution of the teacher encoder is still preserved; applying this enhancement after OCBE's bottleneck would operate on an already information-impoverished embedding in which subtle high-frequency defect cues have been attenuated.

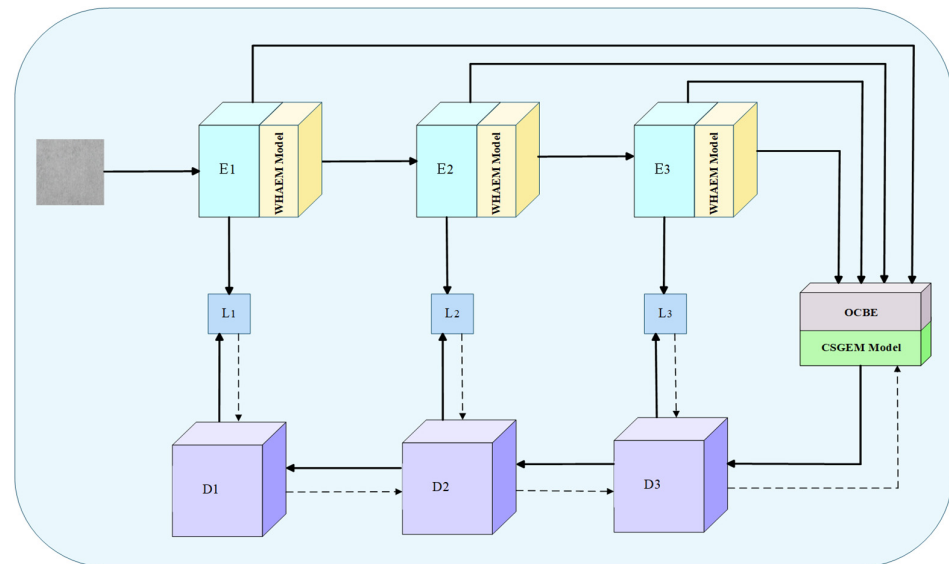


Figure 1. Architecture Diagram of Dual-Domain Reverse Distillation Algorithm for Unsupervised Industrial Surface Defect Detection (E1–E3: multi-scale feature maps extracted by the frozen teacher encoder at three stages; WHAEM: Wavelet High-frequency Deformable Enhancement Module, attached to each teacher stage; L1–L3: the scale-wise cosine similarity discrepancy (1—cosine similarity) computed between the teacher-side feature E_i and the corresponding student-side reconstructed feature D_i at each scale $i = 1, 2, 3$, which are summed to form the total training loss and also used to derive the multi-scale anomaly score map at inference time; OCBE: Omni-dimensional Compression Bottleneck Embedding module; CSGEM: Convolutional Self-Attention Spatial Global Enhancement Module; D1–D3: multi-scale features reconstructed by the student decoder. Solid arrows indicate the forward feature flow; dashed arrows indicate the feature comparison used to compute L1–L3).

CSGEM, in contrast, is embedded between OCBE and the student decoder D, i.e., on the student-side reconstruction path rather than on the teacher branch. This placement is likewise dictated by the reverse distillation mechanism: OCBE compresses the teacher's multi-scale features into a low-dimensional embedding that retains only the statistical regularities of normal samples, and it is the student decoder's task to expand this embedding back into full-resolution features. By inserting CSGEM immediately after OCBE, the student decoder's reconstruction is conditioned on globally aggregated contextual information that reflects the long-range spatial regularity typical of normal fabric texture. For normal regions, this global context is consistent with the local content and therefore facilitates accurate reconstruction, keeping the residual error low. For defective regions, however, the defect locally violates the long-range structural regularity that CSGEM aggregates from the surrounding normal context; as a result, the self-attention-guided reconstruction is biased toward the incorrect, normal-consistent pattern, which further enlarges the discrepancy with the teacher's target feature at the defect location. Critically, embedding CSGEM on the student side rather than the teacher side preserves the asymmetry between the two branches that reverse distillation fundamentally relies on: if CSGEM were instead applied to the teacher branch, the resulting globally regularized target would already

encode long-range consistency that the student could trivially match even in the presence of local anomalies, thereby weakening rather than strengthening the reconstruction-error-based anomaly signal.

In summary, WHAEM operates on the teacher branch prior to bottleneck compression to inject frequency-domain discriminability into the distillation target itself, while CSGEM operates on the student branch after bottleneck compression to regularize the reconstruction process with global spatial context. Together, the two modules widen the teacher-student reconstruction discrepancy at defective regions from complementary frequency and spatial perspectives, while preserving the teacher-student asymmetry that is essential for the reverse distillation framework to remain sensitive to unseen defects.

3.2. Wavelet High-Frequency Deformable Enhancement Module

In defect detection, defect features correspond to high-frequency signals relative to background features. Therefore, we propose to enhance detection performance in the frequency domain by designing a frequency-domain enhancement module named the Wavelet High-frequency Deformable Enhancement Module (WHAEM). Inspired by recent advances in Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT) [20–26], which have become active research topics in deep learning, WHAEM incorporates wavelet-based operations to strengthen high-frequency feature representation. The structure of the module is illustrated in Figure 2.

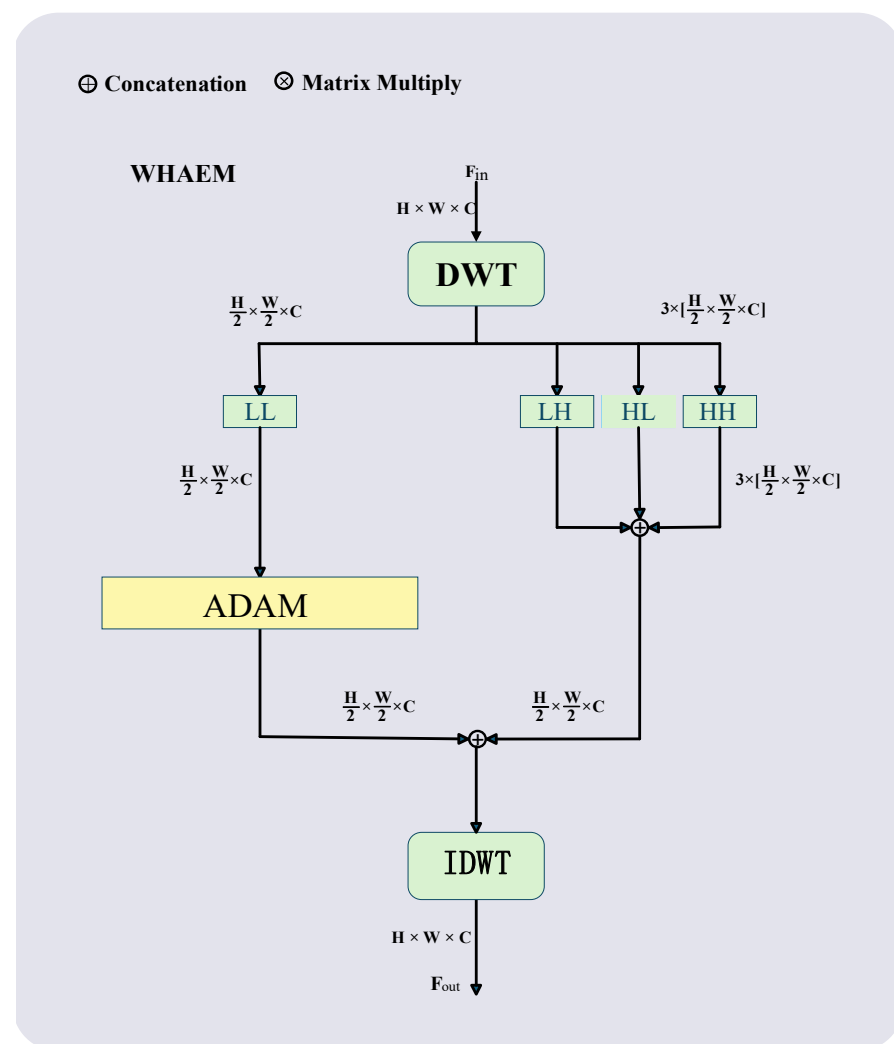


Figure 2. WHAEM Module Structure Diagram.

The overall framework of the Wavelet High-frequency Deformable Enhancement Module (WHAEM) first employs the Haar wavelet transform to decompose the input feature map into four frequency subbands:

$$I_{LL}, \{I_{LH}, I_{HL}, I_{HH}\} = \text{DWT}(I) \quad (1)$$

The low-frequency subband retains the global structure of the input feature map, while the three high-frequency subbands capture texture detail; the biorthogonal property of the wavelet transform ensures no information loss [27]. The high-frequency subbands are processed by the Atrous Deformable Attention Module (ADAM), while the low-frequency subband is left unchanged. Finally, IDWT converts the processed subbands back to the spatial domain to obtain the output feature map:

$$I_{\text{final}} = \text{IDWT}(\hat{I}_{LL} + \hat{I}_{LH} + \hat{I}_{HL} + \hat{I}_{HH}) \quad (2)$$

Formally, the Haar wavelet is an orthogonal wavelet whose scaling function is a piecewise-constant function equal to one over the unit interval from zero to one and equal to zero elsewhere, and whose wavelet function is equal to one over the first half of that unit interval, equal to negative one over the second half, and equal to zero elsewhere. Correspondingly, the associated one-dimensional low-pass filter consists of two equal coefficients, each equal to one over the square root of two, while the associated one-dimensional high-pass filter consists of a first coefficient equal to one over the square root of two and a second coefficient equal to negative one over the square root of two; both filters are used for decomposition and reconstruction in this work. In the proposed WHAEM module, these two one-dimensional filters are combined through outer products to construct four separable two-dimensional kernels: the low-low kernel is obtained from the outer product of the low-pass filter with itself, the low-high kernel from the outer product of the low-pass filter with the high-pass filter, the high-low kernel from the outer product of the high-pass filter with the low-pass filter, and the high-high kernel from the outer product of the high-pass filter with itself. These four kernels correspond, respectively, to the LL (low-low), LH (low-high), HL (high-low), and HH (high-high) sub-bands, and are applied as strided two-dimensional convolutions to implement the discrete wavelet transform described above, with their transposed counterparts used for the inverse transform.

The Haar wavelet is specifically adopted, rather than other wavelet families such as Daubechies or Symlet, for three main reasons. First, the Haar wavelet has the simplest computational form among orthogonal wavelets, involving only addition and subtraction operations without floating-point multiplication overhead, which makes it well suited for real-time industrial inspection and potential deployment on resource-constrained embedded platforms. Second, the Haar wavelet possesses compact support and strict orthogonality, which guarantees precise spatial localization of frequency components without introducing redundant or overlapping information across sub-bands—a property that is particularly desirable for accurately localizing small and sharply bounded surface defects. Third, the multi-scale, multi-directional decomposition provided by the Haar wavelet naturally separates the low-frequency global texture structure from high-frequency directional details (horizontal, vertical, and diagonal), which aligns well with the goal of disentangling the random background texture of non-woven fabrics from the localized, high-frequency signatures of genuine defects.

WHAEM adaptively strengthens high-frequency defect features while suppressing background noise. However, traditional attention mechanisms model features along a single dimension with a fixed receptive field, often losing fine details of tiny defects [28]. To address this, we propose the Atrous Deformable Attention Module (ADAM), which

expands the receptive field via multi-scale atrous convolution, achieves precise focus on high-frequency defects through mean and maximum feature fusion, and adaptively learns key feature weights, thereby enhancing defect representation and suppressing noise. The structure of the module framework is presented in Figure 3.

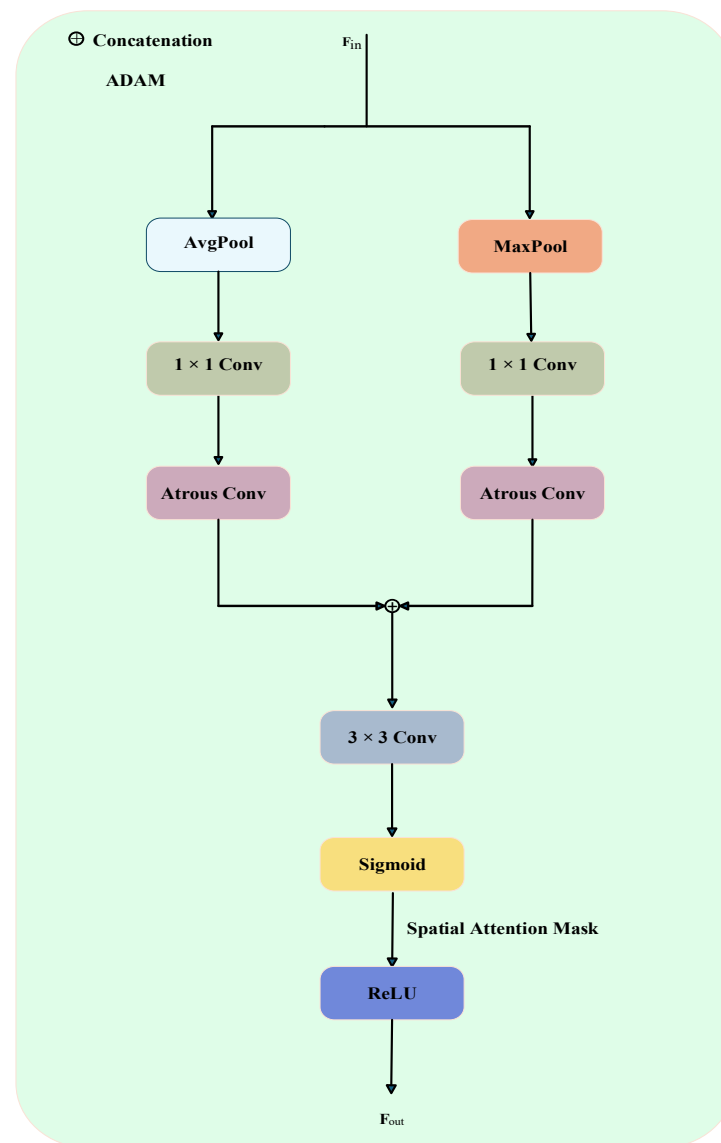


Figure 3. Structural Diagram of the ADAM Module.

ADAM is applied in the comparison stage of the teacher–student model, leveraging local channel information to highlight the defect features extracted by the teacher model [29]. As shown in Figure 3, given an input multi-channel feature map, ADAM sequentially performs six steps: channel statistical information extraction, bottleneck compression, atrous multi-scale feature encoding, deformable gated interaction, attention weight generation, and weighted feature output, ultimately producing an attention-calibrated feature map. For the input convolutional feature map:

$$X \in \mathbb{R}^{C \times H \times W} \quad (3)$$

First, global average pooling and global maximum pooling are applied along the channel dimension to obtain two single-channel feature maps that respectively capture

the overall feature distribution and salient local responses, while preserving the original spatial resolution and avoiding early information loss.

These two maps are then compressed by a 1×1 convolution to reduce redundancy and computational overhead, followed by a 3×3 atrous convolution (dilation rate 2) that expands the receptive field for multi-scale context modeling without downsampling or changing the spatial size. In the deformable gated interaction stage, independent convolutional layers—zero-initialized for training stability—learn modulation weights for the mean and maximum branches; after activation, the two branches undergo cross-gated interaction to achieve complementary feature enhancement and distortion correction, expressed as:

$$F_{\text{cat}}^{\text{d}} = \text{concat}\left(F_{\text{avg}}^{\text{d}} \cdot F_{\text{max}}^{\text{a}}, F_{\text{max}}^{\text{d}} \cdot F_{\text{avg}}^{\text{a}}\right) \quad (4)$$

A backpropagation-scaling mechanism is further applied to the two branches to suppress gradient oscillation and improve convergence stability. In the attention weight generation step, the interacted features are concatenated in the channel dimension and fused via a 3×3 convolution, followed by a Sigmoid activation [30] to generate a full-resolution spatial attention mask, given as:

$$M = \sigma(C_{3 \times 3}(F_{\text{cat}})) \quad (5)$$

This avoids the need for upsampling interpolation. Finally, the mask is multiplied pixel-wise with the input features to enhance key target regions and suppress background information, followed by a ReLU activation [31] to output the calibrated feature map, expressed as:

$$X_{\text{out}} = \text{ReLU}(X \cdot M) \quad (6)$$

By preserving full spatial resolution, expanding the receptive field without added parameters, and adaptively modeling feature deformation, ADAM effectively captures fine-grained defect details in a lightweight manner, while maintaining strong modeling capability for multi-scale and complex distortion scenarios.

Figure 4 illustrates the intermediate feature representations of the WHAEM module at each key processing stage on a representative WFB nonwoven fabric patch containing subtle defects (a low-contrast patch and a small dot defect, visible in the bottom-right corner of the input). As the starting point of WHAEM, the Input fabric patch is the original grayscale image directly fed into the teacher encoder stage E1. After DWT decomposition, the low-frequency subband LL (low-frequency, kept as-is) preserves the rough global texture structure of the fabric. Defect regions with low contrast are nearly indistinguishable at this stage, confirming that the low-frequency components carry minimal defect-discriminative information in this context and require no further enhancement. Meanwhile, the raw high-frequency subbands LH (raw, after DWT split), HL (raw, after DWT split), and HH (raw, after DWT split) before ADAM processing capture horizontal, vertical, and diagonal detail components, respectively. Although defect-related high-frequency responses are already present in these subbands, they are mixed with broadband background texture noise, making it difficult to disentangle the true defect clues from the background. Additionally, the bright edge artifacts visible in the LH and HL subbands originate from the boundary padding applied prior to the DWT transformation.

To address the issue of high-frequency noise mixing, the ADAM-processed high-frequency subbands LH after DeformableAttention, HL after DeformableAttention, and HH after DeformableAttention exhibit a critical feature evolution: the defect-related feature responses are selectively amplified by ADAM, while the spatially uniform background texture noise is significantly suppressed. This enhancement effect is intuitively reflected in the corresponding spatial attention masks, namely the LH attention mask, HL attention

mask, and HH attention mask; these masks assign higher attention weights (bright regions) to subtle defect areas, while assigning lower weights (dark regions) to uniform background areas and boundary artifacts. Finally, the Output after IDWT (reconstructed) integrates the unmodified LL low-frequency structure with the three ADAM-enhanced high-frequency subbands via inverse discrete wavelet transform, generating a spatial-domain feature map with superior defect discriminability. This output serves as the WHAEM-enhanced teacher feature and is subsequently passed to the teacher-student distillation stage.

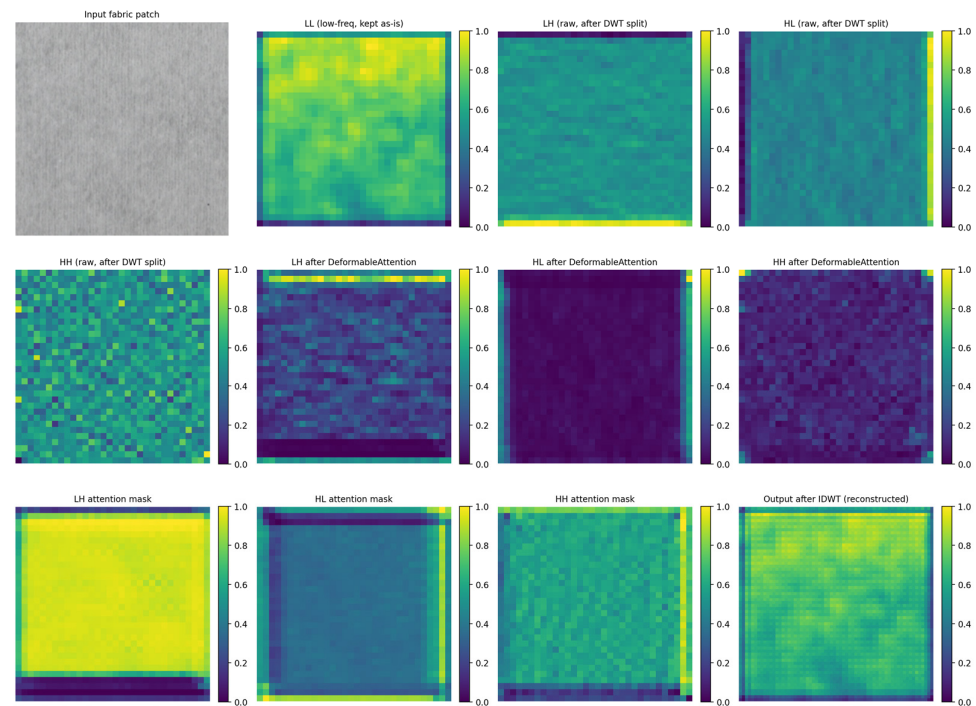


Figure 4. Intermediate feature visualization of the WHAEM module on a WFB fabric patch with subtle defects.

3.3. Convolutional Self-Attention Spatial Global Enhancement Module

This paper proposes the Convolutional Self-Attention Spatial Global Enhancement Module (CSGEM). Based on convolutional self-attention, this module realizes global dependency modeling in the spatial domain. It preserves local detailed features through residual connections and captures long-range spatial correlations via global self-attention, achieving fine-grained enhancement of spatial features and global information fusion. It significantly improves the model's capability to perceive and locate spatial features of defective regions. The overall structure of the module is illustrated in Figure 5. The input to the module is a batch of high-order convolutional feature maps output by the network, where B , C , H , and W denote the batch size, number of channels, height, and width of the feature maps, respectively, which is defined as:

$$X \in \mathbb{R}^{B \times C \times H \times W} \quad (7)$$

where B , C , H , and W denote the batch size, channel dimension, and spatial height and width, respectively. CSGEM sequentially performs three processes: local convolutional feature refinement, multi-head global self-attention modeling, and adaptive residual feature fusion.

First, a 3×3 convolution with symmetric padding extracts local fine-grained features (e.g., textures, edges) while preserving spatial resolution, followed by batch normalization and ReLU activation to stabilize training and introduce nonlinearity.

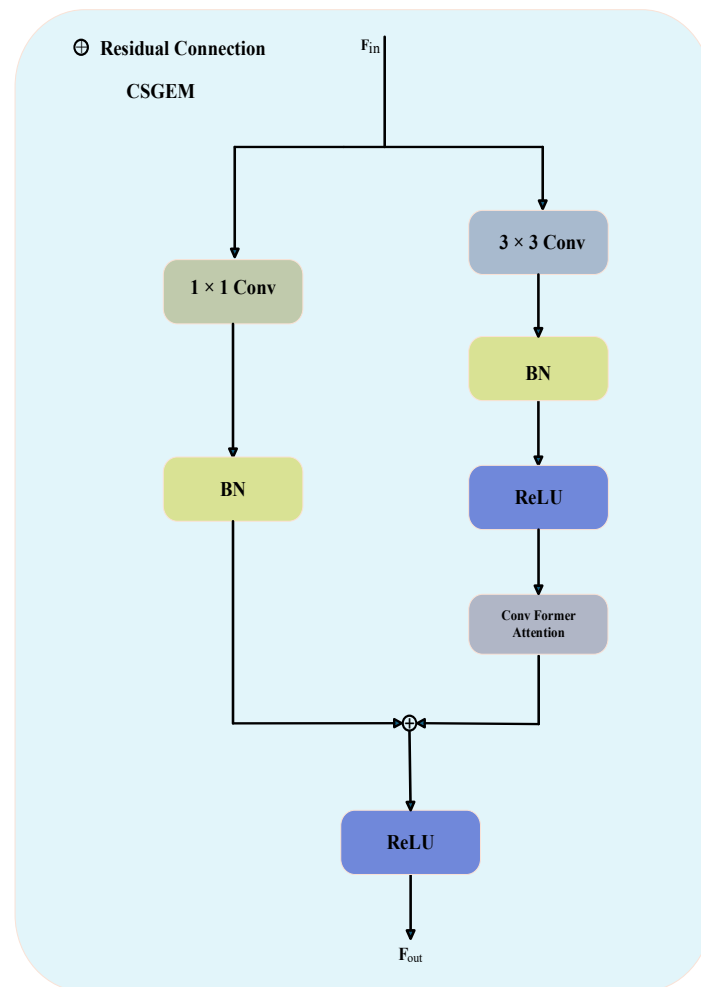


Figure 5. Structural Diagram of the CSGEM Module.

The core of the module is the multi-head global self-attention branch, which overcomes the fixed receptive field of standard convolutions to model long-range spatial dependencies. The refined features are projected via 1×1 convolutions into Query, Key, and Value representations, then flattened into a spatial sequence and processed in multiple heads, with correlation weights computed via scaled dot-product attention:

$$\text{Attn}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

To ease gradient propagation, CSGEM applies an adaptive residual connection: a 1×1 convolution aligns the residual branch when input–output dimensions or strides differ, otherwise identity mapping is used. The residual features are fused element-wise with the attention-enhanced features and calibrated by a ReLU activation:

$$X_{\text{out}} = \text{ReLU}(F_{\text{att}} + F_{\text{res}}) \quad (9)$$

By integrating local convolutional feature refinement with global self-attention modeling and adaptive residual fusion, CSGEM effectively combines fine-grained local details with long-range global context, enhancing feature discriminability while maintaining stable and efficient training.

4. Results

4.1. Datasets and Implementation Details

4.1.1. Datasets

To verify the effectiveness of the proposed method, experiments are conducted on two anomaly detection datasets in this study. One is the self-built non-woven fabric texture dataset (WFB) tailored to the research requirements, which serves as the primary dataset for the anomaly detection task in this work. The other is MVTec AD [32], a widely recognized standard benchmark in the field of anomaly detection. This benchmark comprises 15 categories of real-world scenarios, including 5 texture categories and 10 object categories, providing strong support for verifying the reliability of the experimental results.

For the self-constructed WFB dataset, all images are resized to 256×256 pixels during preprocessing. The dataset is divided into a training set (21,737 images) and a test set (6217 images) according to the standard unsupervised anomaly detection protocol, where the training set contains only defect-free samples. These defect-free training samples cover four different background textures of non-woven fabrics, including spunlace mesh texture, plain spunlace, patterned mesh spunlace, and moiré spunlace, which are used to guide the model in learning the intrinsic feature distribution of normal samples. Correspondingly, the test set consists of both defect-free and anomalous samples. To comprehensively evaluate the model's anomaly recognition capability, the test set contains two types of anomalous defects in different colors: point defects and filamentous defects. The sample composition of the self-built WFB dataset is illustrated in Figure 6.

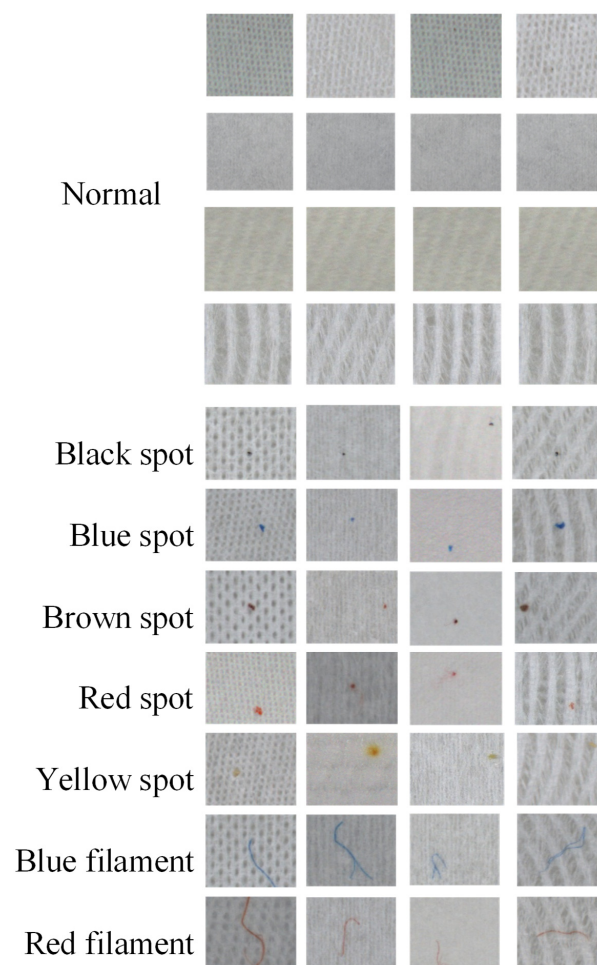


Figure 6. Normal samples and defective samples. The dataset includes normal fabric samples with four different background textures, and defective samples with two types of anomalies (point defects

and filamentous defects) in different colors, including black spot, blue spot, brown spot, red spot, yellow spot, blue filament, and red filament.

For the public benchmark dataset MVTec AD, all images are resized to 256×256 pixels. As a standard dataset for unsupervised anomaly detection, it adopts a similar experimental protocol: the training set contains 3629 defect-free images to learn the inherent feature patterns of standard samples, while the test set contains 1725 images including both normal and defective samples, which can comprehensively measure the overall anomaly detection performance of the model. In addition, each sub-dataset of WFB and MVTec AD covers multiple types of anomalous samples, which can fully verify the generalization and representation ability of the model under different anomaly morphologies. Meanwhile, all anomalous samples in the test sets of the two datasets are equipped with complete pixel-level ground truth annotations, laying a solid data foundation for quantitative comparison and performance evaluation of the anomaly localization task.

4.1.2. Data Acquisition System

The nonwoven fabric image dataset (WFB) was acquired using a custom-built imaging platform, as shown in Figure 7. The system consists of an industrial area-scan camera (model MV-AH7160CU000, iRAYPLE Technology Co., Ltd. [Zhejiang HuaRay Technology Co., Ltd.], Hangzhou, China) equipped with a Sony IMX273 global-shutter CMOS sensor (Sony Corporation, Tokyo, Japan) ($1/2.9$ -inch, 1440×1080 resolution, $3.45 \mu\text{m} \times 3.45 \mu\text{m}$ pixel size), connected via a USB 3.0 interface to an NVIDIA Jetson-based embedded processing unit (NVIDIA Corporation, Santa Clara, CA, USA) for image acquisition and storage.

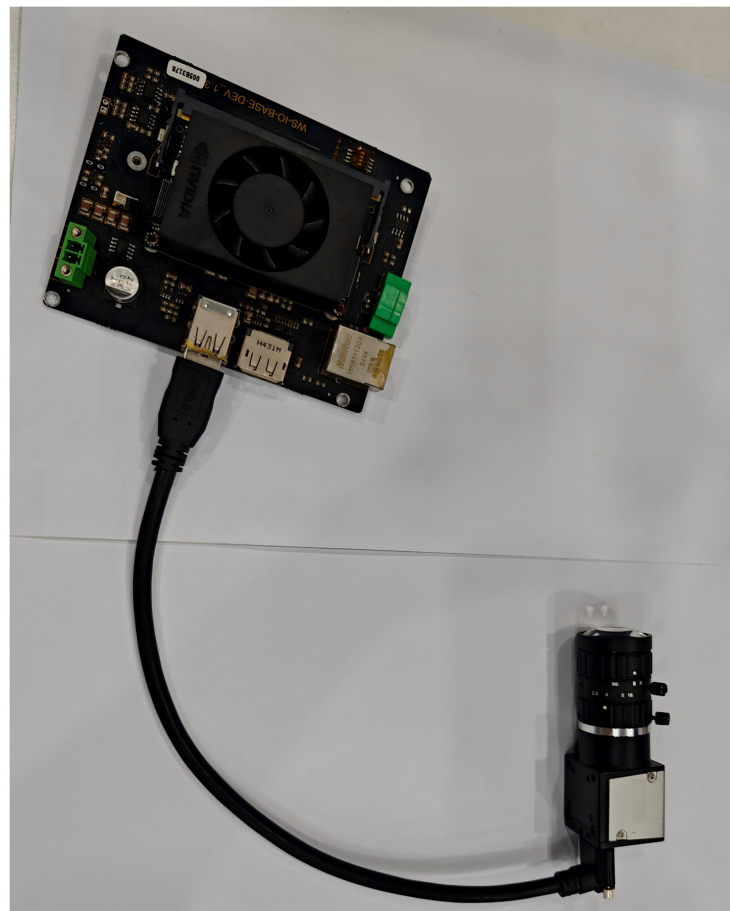


Figure 7. Hardware setup of the WFB acquisition system (including the embedded processing unit and the industrial camera).

The experimental setup for data acquisition is illustrated in Figure 8. The camera was mounted on a vertically adjustable rig positioned directly above the fabric sample, with the lens axis oriented strictly perpendicular to the fabric surface. The working distance between the camera lens and the fabric surface was fixed at 30 cm to minimize geometric distortion and ensure imaging stability. A C-mount lens (Zhejiang HuaRay Technology Co., Ltd., Hangzhou, China) was used to provide sufficient field of view while maintaining image sharpness across the sampling area.



Figure 8. Experimental setup for WFB dataset acquisition.

Illumination was provided by a white panel light source (Zhejiang HuaRay Technology Co., Ltd., Hangzhou, China) positioned above the imaging area, offering uniform, diffuse lighting across the fabric surface and minimizing shadows and specular reflections that could interfere with defect detection.

A total of seven (7) types of nonwoven fabric were used for dataset construction, covering variations in fiber composition, surface texture, and color. For each fabric type, images were repeatedly captured 10 times under identical conditions to account for measurement variability and enhance the robustness of the dataset.

Regarding measurement uncertainty, the fixed 30 cm working distance, strictly perpendicular shooting angle, and stable white panel illumination ensured high repeatability across acquisitions, indicating that hardware-induced measurement error has negligible impact on the final defect detection performance of the proposed algorithm.

4.1.3. Implementation Details

All experimental simulations are performed on a server equipped with an NVIDIA GeForce RTX 4090 GPU (NVIDIA Corporation, Santa Clara, CA, USA), with all algorithmic implementations developed based on the PyTorch (2.7.0) framework. The Adam optimizer is adopted for model training, where the hyperparameters β_1 and β_2 are set to 0.5 and 0.999, respectively. To achieve a trade-off between model convergence efficiency and training stability, the initial learning rate is fixed at 0.005. The model is trained for 200 epochs with a batch size of 16. These training hyperparameters, including the learning rate, batch size, number of epochs, and the Adam optimizer's β coefficients, are kept consistent with the original RD4AD configuration on which our teacher-student backbone is based, ensuring a fair, controlled comparison in which the observed performance improvements can be attributed to the proposed dual-domain enhancement modules rather than to differences in the training protocol. To suppress background noise in anomaly score maps and refine pixel-level defect localization results, a Gaussian filter with a standard deviation of $\sigma = 4$ is employed for post-processing the raw anomaly feature maps. This smoothing parameter is likewise set consistent with the original RD4AD configuration, so that the post-processing protocol applied to the raw anomaly maps does not introduce an additional confounding factor when comparing localization performance against the RD4AD baseline and other compared methods.

Targeting the unsupervised detection of non-woven fabric surface defects, this study constructs a wavelet-enhanced dual-domain reverse distillation network. A ResNet34 backbone [33] is deployed for the dual-branch teacher–student feature extraction structure. Compared with ResNet50 [34] and WideResNet50 [29], ResNet34 achieves a smaller parameter scale and higher computational efficiency, thereby improving inference performance and better meeting the lightweight requirements of industrial defect detection tasks. On this basis, two customized feature enhancement modules, namely the Wavelet High-frequency Deformable Enhancement Module (WHAEM) and the Convolutional Self-Attention Spatial Global Enhancement Module (CSGEM), are embedded into the distillation framework. The core design principle is the synergistic optimization of frequency-domain and spatial-domain features: WHAEM extracts and amplifies fine-grained high-frequency defect features via wavelet transformation, while CSGEM captures long-range spatial dependencies to model holistic structural information. This design enables the complementary fusion of frequency-domain detail features and spatial-domain global features. Combined with the residual learning mechanism of the backbone network and channel dimension optimization, the proposed network can effectively extract multi-scale and multi-level semantic features from fabric images, providing robust and discriminative feature representations for unsupervised anomaly identification and defect classification.

Consistent with the standard evaluation paradigm of unsupervised anomaly detection, all experiments adopt a category-independent training and testing strategy. Each sample category is trained and evaluated separately to eliminate cross-category interference, thereby guaranteeing the accuracy and credibility of experimental results.

4.1.4. Model Complexity and Inference Efficiency

To further substantiate the practical value of DDRD for industrial deployment, we quantitatively evaluate its model complexity and inference efficiency, and compare it

against the RD4AD baseline under identical hardware and input conditions. Both models are tested on the aforementioned NVIDIA GeForce RTX 4090 GPU with an input resolution of 256×256 and a batch size of 1. Parameter count and FLOPs are measured using both the thop and fvcore libraries for cross-validation. Inference latency is averaged over 200 forward passes following a 20-iteration warm-up to eliminate CUDA initialization overhead, and peak GPU memory usage is recorded via `torch.cuda.max_memory_allocated()`. The results are summarized in Table 1.

Table 1. Comparison of model complexity and inference efficiency between DDRD and the RD4AD baseline.

Method	Params (M)	FLOPs (G)	Latency (ms)	FPS	GPU Memory (MB)
RD4AD	161.08	30.53	18.66 ± 0.02	53.60	994.75
DDRD (Ours)	47.91	9.10	16.46 ± 0.04	64.68	298.42

As shown in Table 1, although DDRD introduces two additional feature enhancement modules (WHAEM and CSGEM) into the reverse distillation pipeline, it achieves a substantially lighter and faster model than the RD4AD baseline: parameter count, FLOPs, and peak GPU memory are all reduced by approximately 70%, while inference speed improves by roughly 21% (from 53.60 to 64.68 FPS). This efficiency gain is primarily attributable to our adoption of a lighter ResNet34 backbone in place of the WideResNet50-2 used in RD4AD, which more than offsets the modest additional computation introduced by the two enhancement modules. Moreover, both WHAEM and CSGEM are purely feed-forward modules that operate directly on intermediate feature maps, without relying on any external memory, retrieval, or exemplar-matching mechanism, thereby incurring no additional storage overhead or retrieval latency during inference. With less than 300 MB of peak GPU memory and a throughput of nearly 65 FPS, DDRD comfortably satisfies the computational and memory budget of typical industrial inline inspection systems, confirming its feasibility for real-time deployment on resource-constrained industrial hardware.

4.2. Evaluation Metrics

The performance of the fabric defect detection system is evaluated using key object detection metrics described in the equations. The formulas of the evaluation metrics are given as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (11)$$

$$\text{Image-level AUROC} = \int_0^1 \text{TPR}(t) d\text{FPR}(t) \quad (12)$$

$$\text{Pixel-level AUROC} = \int_0^1 \text{TPR}_{\text{pixel}}(t) d\text{FPR}_{\text{pixel}}(t) \quad (13)$$

Here, TP (True Positive) represents the number of correctly predicted defects, and FP (False Positive) represents the number of false predictions when no defect exists. Image-level AUROC is a metric that evaluates the classification performance of the model at the image level. It calculates the area under the curve between the True Positive Rate (TPR) and False Positive Rate (FPR) under different thresholds. A higher AUROC value indicates better classification performance of the model at the image level. Pixel-level AUROC is a metric that evaluates the localization performance of the model at the pixel level. It calculates the area under the curve between the pixel-level TPR and FPR under different thresholds. A higher pixel-level AUROC value indicates more accurate defect localization.

4.3. Comparative Experiments

In this experiment, we compare the proposed DDRD with the following state-of-the-art methods on the public MVTec AD dataset, including RD4AD [29], DBFAD [35], SupersimpleNet [36], MKD [37], SPADE [38], and SimpleNet [39].

RD4AD proposes a reverse distillation paradigm and adopts a heterogeneous architecture consisting of a teacher encoder and a student decoder. Following the knowledge distillation order of high-level abstract features, the OCBE module preserves key information of normal patterns, and realizes anomaly detection and localization based on differences in feature similarity. Based on reverse distillation technology, DBFAD is optimized for fabric defect detection. It introduces residual connections and customized attention layers to retain shallow high-dimensional features, and designs a texture-specific bottleneck embedding module. Layer weights are assigned via an adaptive loss function, and domain generalization can be achieved by removing residual connections. SuperSimpleNet integrates unsupervised and supervised learning. It obtains target domain features through a pre-trained feature extractor and a feature adapter, and generates synthetic anomalies using binarized Perlin noise masks in the feature space. Equipped with a segmentation head and a classification head, it realizes anomaly localization and global semantic judgment respectively, balancing detection performance and inference speed. MKD adopts a knowledge distillation framework, which distills multi-layer intermediate features from the pre-trained expert network to a compact clone network. Euclidean distance loss and directional similarity loss are utilized to ensure the integrity of feature transfer, and a gradient interpretability algorithm is applied to achieve pixel-level anomaly localization. SimpleNet consists of a feature extractor, a feature adapter, an anomaly feature generator and a discriminator. The adapter alleviates the domain deviation of pre-trained features, and Gaussian noise is added in the feature space to generate anomalous samples. A lightweight MLP discriminator is trained to distinguish normal and abnormal features, and the maximum value of the anomaly map is taken as the detection score during inference. Without additional training, SPADE leverages the pre-trained ResNet [40] to extract a multi-scale feature pyramid. It first retrieves the most similar normal images to the test image via K-nearest neighbor search, and then calculates distances through pixel-level feature correspondence. Regions without matching features are judged as anomalies, thereby achieving both image-level and pixel-level anomaly detection.

The quantitative experimental results are presented in Table 2, which specifically focuses on the pixel-level AUROC metric. DDRD (marked as “Ours” in the table) achieves the best performance across all six categories on the MVTec AD dataset and attains the highest average pixel-level AUROC of 97.5%.

Table 2. Anomaly detection and localization results in terms of pixel-level AUROC on the MVTec AD dataset.

	Category	RD4AD	SSN	SPADE	SimpleNet	Ours
MVTec	Carpet	99.0	91.9	98.9	97.6	99.2
	Leather	98.8	97.8	99.0	96.7	99.3
	Transistor	92.1	94.1	88.6	93.9	94.1
	Wood	95.0	87.9	94.5	93.8	95.2
	Cable	97.2	96.9	92.4	97.4	98.3
	Bottle	98.1	96.2	97.0	98.0	98.8
	Average	96.7	94.1	95.0	96.2	97.5

Compared with existing methods, DDRD demonstrates comprehensive and significant advantages. Notably, it performs particularly well in the Leather and Bottle categories,

reaching 99.3% and 98.8%, respectively. In terms of average pixel-level performance, our method achieves a substantial improvement of +0.8% over the second-best method, RD4AD (96.7% average), proving the robustness and effectiveness of DDRD in pixel-level anomaly localization tasks.

The quantitative experimental results are presented in Table 3, which specifically displays the image-level AUROC metric evaluation on the MVTec AD dataset.

Table 3. Anomaly detection and localization results in terms of image-level AUROC on the MVTec AD dataset.

	Category	RD4AD	DBFAD	SSN	MKD	SPADE	SimpleNet	Ours
MVTec	Carpet	99.1	99.1	88.6	64.5	98.9/92.8	99.4	100
	Leather	100	100	100	63.9	99.0/95.4	100	100
	Transistor	95.8	96.8	99.1	68.9	88.6/90.3	96.8	99.2
	Wood	98.6	99.2	98.2	92.1	94.5/95.8	99.2	99.6
	Cable	97.8	98.6	98.5	82.7	92.4/84.8	99.3	99.3
	Bottle	99.3	99.7	99.6	92.2	97.0/97.2	99.6	100
	Average	98.4	98.9	97.3	77.4	95.0/92.7	99.1	99.7

Our proposed DDRD (marked as “Ours”) demonstrates exceptionally outstanding performance in image-level anomaly detection classification. Among the six selected categories, DDRD achieves the highest or tied-highest scores on five categories. Notably, it attains a perfect 100% image-level detection accuracy on both the ‘Leather’ and ‘Bottle’ categories.

In terms of overall average performance, our method achieves an image-level AUROC of 99.7%, significantly outperforming the second-best method, SimpleNet (average 99.1%), and DBFAD (average 98.9%). Compared to the existing suboptimal methods, our approach achieves a significant improvement of +0.6%, fully proving the powerful global discriminative capability of DDRD in image-level classification tasks.

It should be noted that the aforementioned experiments were performed on six selected categories (e.g., leather, transistor, and bottle) from the MVTec AD dataset. The remaining categories (e.g., pill, zipper, and screw) were excluded from our comparative scope. For these excluded categories, the irregular defect shapes and fine-grained anomaly patterns inherent to these objects present significant challenges to the current architecture of DDRD, consequently leading to suboptimal performance where our method fails to achieve the best results. We plan to address these shortcomings and improve the performance on these challenging categories in our future work.

Table 4 presents the quantitative experimental results for anomaly detection and localization specifically for the pixel-level AUROC metric on the self-built WFB dataset. DDRD (denoted as “Ours” in the table) continues to exhibit excellent performance on this pixel-level evaluation metric, achieving the best overall results and significantly outperforming all compared methods. Compared with the second-best RD4AD method, our method achieves a +0.5% improvement (rising from 97.5% to 98.0%); while compared with the SimpleNet method, the improvement reaches +2.4%. This result regarding the single metric further demonstrates the superior advantage of the proposed model in pixel-level localization accuracy, enabling it to precisely identify and localize anomalous regions.

Table 4. Anomaly detection and localization results in terms of pixel-level AUROC on the self-built WFB dataset.

Self-Check Dataset	SimpleNet	RD4AD	Ours
wfb	95.6	97.5	98.0

Table 5 presents the quantitative experimental results for anomaly detection specifically for the image-level AUROC metric on the self-built WFB dataset. DDRD (denoted as “Ours” in the table) still achieves the optimal performance on this image-level evaluation metric, with highly significant advantages. Specifically, compared with the second-best RD4AD method, our method achieves a stable improvement of +1.5% (increasing from 91.5% to 93.0%); while compared with the MKD method, which performs poorly (only scoring 36.4%), our method demonstrates an overwhelming lead. This image-level result further substantiates the exceptional capability of the proposed model in discriminating global image features, enabling it to perform anomaly detection and classification tasks at the image level with extremely high accuracy.

Table 5. Anomaly detection and localization results in terms of image-level AUROC on the self-built WFB dataset.

Self-Check Dataset	MKD	RD4AD	Ours
wfb	36.4	91.5	93.0

As shown in Figures 9 and 10, DDRD produces sharper anomaly maps with more accurate defect boundaries compared to baseline methods, and significantly reduces false positives caused by complex random textures in non-woven fabrics. The results demonstrate that our proposed DDRD has an excellent ability to distinguish between normal and anomalous samples.

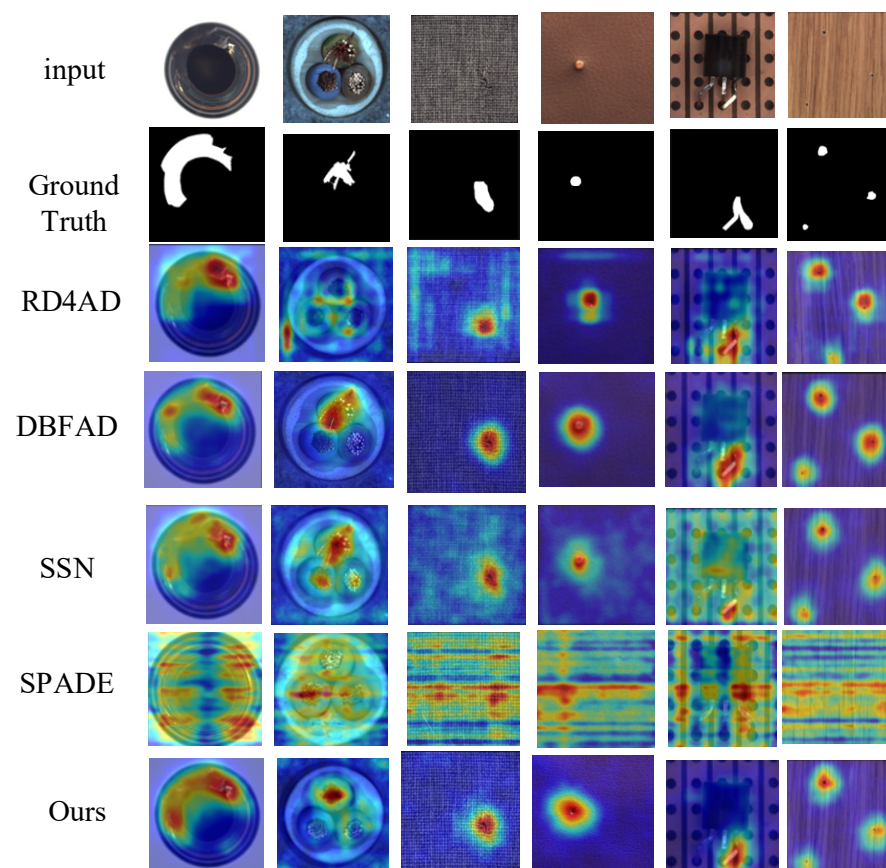


Figure 9. Qualitative results of DDRD and different SOTA methods on the MVTec AD dataset.

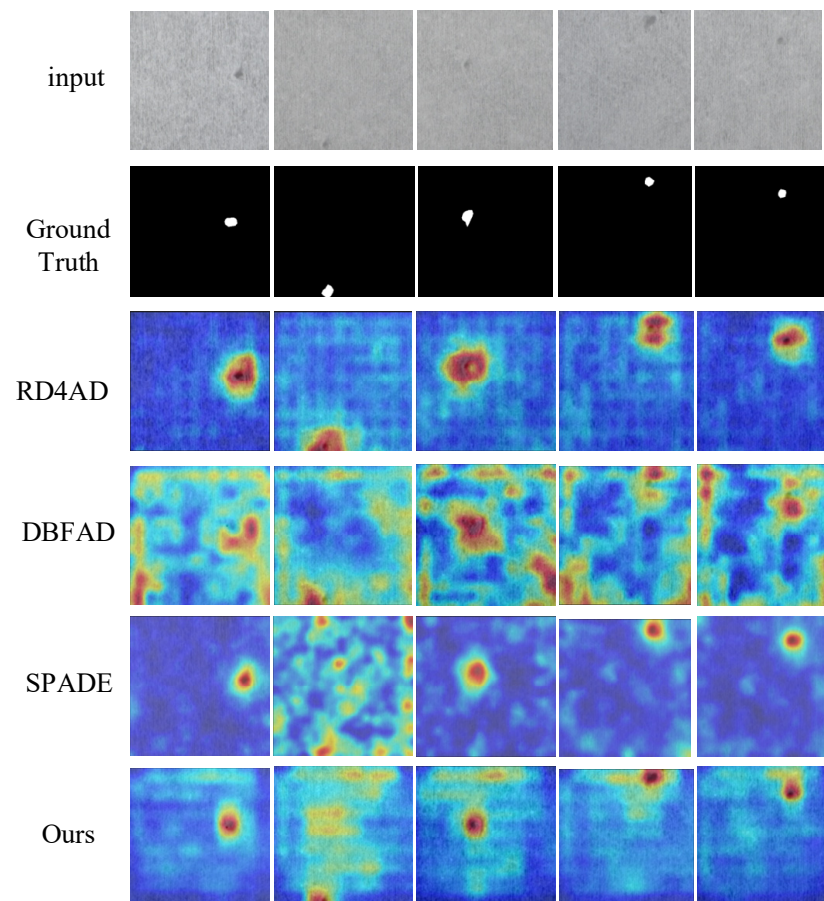


Figure 10. Qualitative results of DDRD and various SOTA methods on the self-built WFB dataset.

4.4. Ablation Experiments

We investigate the individual effects of the WHAEM and CSGEM modules on the TS model, as well as their combined effect. A total of four experiments are designed, and the quantitative results are presented in Table 6. All defect detection experiments are conducted on the self-built WFB dataset. We take the pre-trained residual block as the baseline. The embeddings of pre-trained residual blocks may contain anomalous features, which reduce the representation discrepancy of the T-S model. The proposed WHAEM module is attached to the side of the teacher encoder to enhance high-frequency information and fine-grained details. The CSGEM module is embedded between the OCBE module and the student decoder to strengthen global dependency modeling. This design realizes dual-domain collaborative enhancement in the frequency and spatial domains. WHAEM extracts high-frequency detailed features in the wavelet frequency domain, while CSGEM captures long-range global dependencies in the spatial dimension, achieving complementary fusion of frequency-domain detail representation and spatial global correlation.

- (1) TS: Baseline Model T-S Model
- (2) TS + WHAEM
- (3) TS + CSGEM
- (4) TS + WHAEM + CSGEM

This ablation study (Table 6) validates the effectiveness of each module on the complete WFB dataset, with 95% confidence intervals obtained via bootstrap resampling reported alongside each metric. Taking the baseline model TS (based on RD4AD) as the reference, its performance reaches 97.5 [97.0, 97.9] for AUROC_P, 91.5 [91.0, 92.1] for AUROC_S, and 88.7 [88.0, 89.4] for AUPRO_P.

Table 6. Ablation experiments of TS, WHAEM, and CSGEM.

Method	TS	TS + WHAEM	TS + CSGEM	TS + WHAEM + CSGEM
AUROC _P	97.5	97.6	97.5	97.8
AUROC _S	91.5	92.2	92.5	93.0
AUPRO _P	88.7	89.0	88.9	90.0

When the frequency-domain detection module WHAEM and the spatial-domain detection module CSGEM are introduced separately, both lead to performance improvements. Specifically, TS + WHAEM raises AUROC_S and AUPRO_P to 92.2 [91.5, 92.9] and 89.0 [88.1, 89.9] respectively, indicating that the frequency-domain module effectively enhances the model's discriminative ability for defects. TS + CSGEM shows a more notable improvement in AUROC_S, reaching 92.5 [91.9, 92.7], while its AUPRO_P reaches 88.9 [88.0, 90.0], suggesting that the spatial-domain module contributes more substantially to sample-level anomaly detection.

When the frequency-domain module WHAEM and the spatial-domain module CSGEM are jointly employed (TS + WHAEM + CSGEM), all three metrics reach their best values: AUROC_P of 97.8 [96.9, 98.5], AUROC_S of 93.0 [92.0, 94.1], and AUPRO_P of 90.0 [89.5, 90.5], representing improvements of 0.3, 1.5, and 1.3 percentage points over the baseline TS, respectively. Notably, the confidence intervals of the joint configuration do not overlap with those of the baseline TS for AUROC_S and AUPRO_P, indicating that these improvements are statistically meaningful rather than due to random variation. This demonstrates the complementary nature of frequency-domain and spatial-domain information—their combination captures defect characteristics more comprehensively from different perspectives, significantly boosting the model's overall performance in pixel-level localization accuracy, sample-level discrimination, and region overlap, thereby confirming the effectiveness and necessity of the proposed WHAEM and CSGEM module designs.

To further validate the specific contributions of the WHAEM and CSGEM modules across different defect scenarios, we divide the WFB dataset into four sub-categories according to defect characteristics: minor defects, low-contrast defects, complex texture interference, and edge defects. The corresponding ablation results are presented in Tables 7–10, respectively, with 95% confidence intervals obtained via bootstrap resampling reported alongside each metric. The results show that the improvement effects of each module vary significantly across scenarios, addressing the details previously obscured by reporting only the overall AUROC.

Table 7. Ablation experiments on minor defects.

Method	TS	TS + WHAEM	TS + CSGEM	TS + WHAEM + CSGEM
AUROC _P	99.1	100	100	100
AUROC _S	98.6	99.3	99.5	100
AUPRO _P	99.2	99.4	99.7	99.9

As shown in Table 7, the baseline model TS already achieves relatively high performance (AUROC_P 99.1 [98.1, 99.9], AUROC_S 98.6 [97.6, 99.6], AUPRO_P 99.2 [98.7, 99.6]), indicating that this type of defect is inherently distinctive and relatively easy to detect. The addition of WHAEM and CSGEM further pushes all metrics toward saturation, with the joint configuration reaching 100 [99.6, 100] for both AUROC_P and 100 [99.5, 100] for AUROC_S, and AUPRO_P rising to 99.9 [99.8, 100]. This demonstrates that both modules still provide stable, albeit limited, gains in simple scenarios—which also explains why the overall pixel-level AUROC on the complete WFB dataset improved by only 0.3%: since minor defects constitute a large proportion of the samples and the baseline is already

close to the detection ceiling, there is little room for further improvement, resulting in a “diminishing marginal effect” on the overall metric.

Table 8. Ablation experiments on low-contrast defects.

Method	TS	TS + WHAEM	TS + CSGEM	TS + WHAEM + CSGEM
AUROC _p	96.6	97.5	97.7	98.2
AUROC _s	98.0	98.4	98.4	98.5
AUPRO _p	83.3	84.6	83.8	86.4

As shown in Table 8, the contribution of the modules becomes considerably more pronounced. The TS baseline achieves only 96.6 [95.6, 97.6] for AUROC_P and 83.3 [82.3, 84.3] for AUPRO_P. With WHAEM added, AUROC_P rises to 97.5 [96.5, 98.5] and AUPRO_P to 84.6 [83.6, 85.6], reflecting that the frequency-domain module effectively alleviates the detection difficulty caused by the low contrast between defects and background by enhancing high-frequency detail responses. With CSGEM added, AUROC_P further increases to 97.7 [96.7, 98.7], while AUPRO_P fluctuates slightly (83.8 [82.8, 84.8]), suggesting that spatial-domain context modeling contributes more moderately in this scenario. When WHAEM and CSGEM are jointly applied, AUROC_P reaches 98.2 [97.2, 99.2] and AUPRO_P rises markedly to 86.4 [85.4, 87.4], a 3.1-percentage-point improvement over the baseline, clearly demonstrating that the complementarity between frequency-domain and spatial-domain information is most evident in this challenging low-contrast scenario. Notably, the confidence interval of the joint configuration’s AUPRO_P does not overlap with that of the baseline, confirming that this improvement is statistically meaningful.

Table 9. Ablation experiments on complex texture interference.

Method	TS	TS + WHAEM	TS + CSGEM	TS + WHAEM + CSGEM
AUROC _p	97.4	97.7	97.5	97.9
AUROC _s	97.7	98.1	98.8	99.2
AUPRO _p	81.9	82.7	83.8	85.5

As shown in Table 9, where defect features are easily submerged by background texture, the baseline AUPRO_P is only 81.9 [80.9, 82.9], the lowest among the four scenarios. The introduction of WHAEM raises AUPRO_P to 82.7 [81.7, 83.7], while CSGEM yields a more substantial improvement, reaching 83.8 [82.8, 84.8], indicating that the spatial-domain module more effectively separates genuine defect regions from complex textures by capturing local structural consistency. When the two modules are combined, AUPRO_P increases to 85.5 [84.5, 86.5], a 3.6-percentage-point improvement over the baseline, while AUROC_S rises from 97.7 [96.7, 98.7] to 99.2 [98.8, 99.5]—demonstrating that the synergy between the frequency-domain and spatial-domain modules is critical for suppressing texture interference and improving region-overlap accuracy. Here again, the non-overlapping confidence intervals between the baseline and the joint configuration support the statistical significance of the observed gains.

As shown in Table 10, the baseline model already performs well in terms of AUROC_P (99.3 [98.9, 99.6]) and AUPRO_P (98.5 [97.5, 99.5]), but AUROC_S is comparatively lower (97.5 [96.5, 98.5]). With WHAEM added, AUROC_S rises to 98.7 [97.7, 99.7]; with CSGEM, it further increases to 98.9 [97.9, 99.9]; and with the joint module, it reaches 99.1 [98.3, 99.8]. This indicates that both the frequency-domain and spatial-domain modules make positive contributions to the sample-level discriminative ability for edge defects, more fully exploiting the high-frequency characteristics and local spatial continuity present at defect edges.

Table 10. Ablation experiments on edge defects.

Method	TS	TS + WHAEM	TS + CSGEM	TS + WHAEM + CSGEM
AUROC _p	99.3	99.6	99.7	99.9
AUROC _s	97.5	98.7	98.9	99.1
AUPRO _p	98.5	98.9	99.0	99.2

5. Discussion

The proposed DDRD framework achieves competitive performance on both the MVTec AD dataset and the self-built WFB dataset, with an average Pixel-level AUROC of 97.5% and Image-level AUROC of 99.7%, surpassing all compared baseline methods including RD4AD, DBFAD, SuperSimpleNet, MKD, SPADE, and SimpleNet.

The performance improvements are primarily attributed to the dual-domain collaborative enhancement strategy. WHAEM explicitly extracts defect-related high-frequency components via the Haar wavelet transform and adaptively captures irregular defect morphologies through deformable convolution, compensating for the insufficient response of existing spatial-domain methods to subtle defects. CSGEM enhances low-frequency global structural representations by integrating local convolutional features with global self-attention context. The two modules work collaboratively within the reverse distillation framework, effectively amplifying the reconstruction discrepancy between normal and anomalous regions to achieve precise anomaly localization.

Compared with existing methods, reverse distillation-based approaches such as RD4AD primarily perform feature modeling in the spatial domain, while feature embedding methods such as SPADE and SimpleNet conduct anomaly detection based on pre-trained feature extractors combined with distance-based scoring mechanisms. Building upon the reverse distillation framework, the proposed method directly embeds wavelet operations into the backbone feature extraction process, enabling joint modeling of frequency-domain and spatial-domain information, and achieves an average Pixel-level AUROC of 97.5%, representing an improvement over the aforementioned methods.

Nevertheless, certain limitations of the proposed method should be acknowledged. First, the introduction of WHAEM and CSGEM increases computational overhead, which imposes constraints on deployment to edge devices; this will be addressed in future work through model compression techniques such as knowledge distillation and channel pruning. Second, the current method shows relatively limited sensitivity to defect types characterized predominantly by low-frequency features, such as large-area color deviations; adaptively balancing high- and low-frequency enhancement remains a direction for future investigation. Furthermore, the self-built WFB dataset is limited in scale and defect diversity, and expanding the dataset to validate the generalization of the proposed method across a broader range of fabric types will be an important focus of future research.

6. Conclusions

This work presents two novel functional modules for unsupervised defect detection of non-woven fabrics, namely the Wavelet High-frequency Deformable Enhancement Module (WHAEM) and the Convolutional Self-Attention Spatial Global Enhancement Module (CSGEM).

Most existing methods for non-woven fabric defect detection predominantly perform feature analysis only in the spatial domain and generally neglect the complementary information hidden in the frequency domain. From the perspective of signal characteristics, surface defects correspond to high-frequency components, while normal texture backgrounds belong to low-frequency components. Targeting this limitation, this study

innovatively combines frequency-domain and spatial-domain enhancement strategies to strengthen defect feature representation and overall detection performance.

Specifically, WHAEM is constructed based on the Haar wavelet transform and its inverse operation. It applies deformable convolution to high-frequency features sensitive to defects, so as to accurately highlight local anomaly-related texture details. In contrast, CSGEM focuses on spatial feature modeling, which fully integrates global structural context and local spatial features to effectively enhance low-frequency global representations. The two modules work in a collaborative manner to achieve targeted enhancement of defect features, establishing a complementary dual-domain feature representation paradigm.

Embedded within the reverse distillation framework, the proposed approach eliminates the dependence on large-scale labeled defective samples. This advantage makes the method well applicable to practical industrial scenarios, where defective samples are scarce and manual data annotation incurs high costs. Sufficient experiments on two standard anomaly detection datasets verify the effectiveness and practicability of our algorithm, which achieves competitive performance in non-woven fabric defect detection. The proposed scheme can be deployed to online quality inspection systems for high-speed non-woven fabric production lines.

Despite these promising results, the proposed method still has certain limitations that warrant further investigation. First, the introduction of the WHAEM and CSGEM modules, while improving detection accuracy, inevitably increases the computational complexity and parameter count of the overall network compared with the baseline reverse distillation framework; this additional overhead may constrain real-time inference speed and pose challenges for deployment on resource-constrained edge devices commonly used in industrial production lines. Future work will explore lightweight model compression techniques, such as knowledge distillation, channel pruning, or neural architecture search, to reduce computational cost while preserving detection accuracy. Second, since WHAEM is specifically designed to enhance high-frequency components associated with local texture anomalies, the current method exhibits relatively limited sensitivity to defect types that are predominantly characterized by low-frequency variations, such as large-area color deviations, uneven dyeing, or gradual thickness inconsistencies; how to adaptively balance the enhancement of high- and low-frequency information according to different defect characteristics remains an open direction for future research. Third, although the self-built WFB dataset provides a valuable benchmark for non-woven fabric defect detection, its scale and defect diversity are still relatively limited compared with large-scale industrial datasets, which may restrict a comprehensive evaluation of the method's generalization capability across more complex fabric types and defect patterns; expanding the dataset with a broader range of fabric materials, textures, and defect categories will be an important focus of future research to further validate the robustness and applicability of the proposed framework.

In addition, this research provides a solid technical reference for intelligent quality control and process optimization across the textile manufacturing industry. For future research, we will further optimize the model architecture to adapt to edge device deployment, and extend the proposed framework to various fabric defect detection tasks, aiming to fully validate its generalization ability.

Author Contributions: Conceptualization, R.L.Y.; methodology, R.L.Y. and W.W.; software, R.L.Y.; validation, R.L.Y.; formal analysis, R.L.Y.; investigation, R.L.Y.; resources, W.W.; data curation, R.L.Y.; writing original draft preparation, R.L.Y.; writing—review and editing, W.W. and Z.H.; visualization, Z.H.; supervision, W.W.; funding acquisition, W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Hubei Provincial Department of Education, grant number B2023051.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to commercial confidentiality restrictions, as the industrial defect images were collected from Jiumu (Wuhan) Technology Co., Ltd.'s production lines, and public release requires authorization from the industrial partner to protect proprietary process information and intellectual property.

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

References

- Li, C.; Wei, K.; Zhao, Y.; Tian, X.; Qian, Y.; Zhang, L.; Wang, R. Improvement of High-Speed Detection Algorithm for Nonwoven Material Defects Based on Machine Vision. *J. Donghua Univ.* **2024**, *41*, 416–427.
- He, F.; Xin, B.; Zhan, Z.; Yang, W.; Liu, J. A review of fabric defect detection in textile manufacturing. *J. Text. Inst.* **2026**, *117*, 327–345.
- Cui, Y.; Liu, Z.; Lian, S. A Survey on Unsupervised Industrial Anomaly Detection Algorithms. *arXiv* **2022**, arXiv:2204.11161.
- Carrilho, R.; Yaghoubi, E.; Lindo, J.; Hambarde, K.; Proença, H. Toward Automated Fabric Defect Detection: A Survey of Recent Computer Vision Approaches. *Electronics* **2024**, *13*, 3728. [[CrossRef](#)]
- Kahraman, Y.; Durmusoglu, A. Deep learning-based fabric defect detection: A review. *Text. Res. J.* **2023**, *93*, 1485–1503.
- Wan, D.; Gao, C.; Zhou, J.; Shen, X.; Shen, L. Unsupervised fabric defect detection with high-frequency feature mapping. *Multimed. Tools Appl.* **2024**, *83*, 21615–21632.
- Haobo, Y. A Survey of Industrial Surface Defect Detection Based on Deep Learning. In *Proceedings of the 2024 International Conference on Cyber-Physical Social Intelligence (ICCSI), Doha, Qatar, 8–12 November 2024*; IEEE: New York, NY, USA, 2024; pp. 1–6.
- Arshad, S.R.; Shahzad, M.K. Deep Learning Based Fabric Defect Detection. *Res. Rep. Comput. Sci.* **2024**, *3*, 1–11. [[CrossRef](#)]
- Ramachandran, V.; Madhavan, R.; Anand, P.; Vishanth, A.; Pradeep, K.V. Improved Fabric Defect Detection Using a Vision Transformer and ResNet Hybrid Model. *Int. J. Adv. Res. Sci. Commun. Technol.* **2023**, *3*, 734–743. [[CrossRef](#)]
- Liang, T.; Liu, T.; Wang, J.; Zhang, J. Causal inference-guided deep learning method for vision-based defect detection of complex patterned fabrics. *Sci. Sin. Technol.* **2023**, *53*, 1138–1149. [[CrossRef](#)]
- Guo, Y.; Kang, X.; Li, J.; Yang, Y. Automatic Fabric Defect Detection Method Using AC-YOLOv5. *Electronics* **2023**, *12*, 2950. [[CrossRef](#)]
- Zhou, L.; Ma, B.; Dong, Y.; Yin, Z.; Lu, F. DCFE-YOLO: A novel fabric defect detection method. *PLoS ONE* **2025**, *20*, e0314525. [[CrossRef](#)] [[PubMed](#)]
- Xiang, Z. Lightweight fabric defect detection method based on high precision multi-scale integration. *J. Zhejiang Univ. (Eng. Sci.)* **2025**, *59*, 2556–2565.
- Li, M.; Ye, P.; Cui, S.; Zhu, P.; Liu, J. HKAN: A Hybrid Kolmogorov–Arnold Network for Robust Fabric Defect Segmentation. *Sensors* **2024**, *24*, 8181. [[CrossRef](#)] [[PubMed](#)]
- Hinton, G.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network. *Comput. Sci.* **2015**, *14*, 38–39.
- Zhang, H.; Liu, S.; Lu, S.; Yao, L.; Li, P. Knowledge distillation for unsupervised defect detection of yarn-dyed fabric using the system DAERD: Dual attention embedded reconstruction distillation. *Color. Technol.* **2024**, *140*, 125–143.
- Pei, M.; Liu, N.; Gao, P.; Sun, H. Reverse Knowledge Distillation with Two Teachers for Industrial Defect Detection. *Appl. Sci.* **2023**, *13*, 3838. [[CrossRef](#)]
- Zhou, J.; Wong, W.; Liao, F. Attention-based Multi-scale Dual Flow Reverse Distillation for Unsupervised Anomaly Detection. *Expert Syst. With Appl.* **2026**, *327*, 132791. [[CrossRef](#)]
- Tran, C.D.; Pham, L.H.; Tran, D.N.N.; Ho, Q.P.N.; Jeon, J.W. Dual Memory Networks Guided Reverse Distillation for Unsupervised Anomaly Detection. In *Proceedings of the 17th Asian Conference on Computer Vision (ACCV 2024), Hanoi, Vietnam, 8–12 December 2024*; pp. 2650–2666.
- Gao, N.; Jiang, X.; Zhang, X.; Deng, Y. Efficient Frequency-Domain Image Deraining with Contrastive Regularization. In *Proceedings of the 18th European Conference on Computer Vision (ECCV 2024), Milan, Italy, 29 September–4 October 2024*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; pp. 240–257.
- Yang, Y.; Yuan, G.; Li, J. SFFNet: A Wavelet-Based Spatial and Frequency Domain Fusion Network for Remote Sensing Segmentation. In *Proceedings of the IEEE Transactions on Geoscience and Remote Sensing*; IEEE: Piscataway, NJ, USA, 2024.
- Tan, J.; Pei, S.; Qin, W.; Fu, B.; Li, X.; Huang, L. Wavelet-based Mamba with Fourier Adjustment for Low-light Image Enhancement. In *Proceedings of the 17th Asian Conference on Computer Vision (ACCV 2024), Hanoi, Vietnam, 8–12 December 2024*.

23. Hwang, S.; Han, D.; Jung, C.; Jeon, M. WaveDH: Wavelet Sub-bands Guided ConvNet for Efficient Image Dehazing. *arXiv* **2024**, arXiv:2404.01604.
24. Finder, S.E.; Amoyal, R.; Treister, E.; Freifeld, O. Wavelet Convolutions for Large Receptive Fields. In *Proceedings of the 18th European Conference on Computer Vision (ECCV 2024), Milan, Italy, 29 September–4 October 2024*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; Volume 15112, pp. 363–380.
25. Zhao, C.; Cai, W.; Dong, C.; Hu, C. Wavelet-based Fourier Information Interaction with Frequency Diffusion Adjustment for Underwater Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 8281–8291.
26. Yan, F.; Jiang, X.; Lu, Y.; Cao, J.; Chen, D.; Xu, M. Wavelet and Prototype Augmented Query-based Transformer for Pixel-level Surface Defect Detection. In *Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 23860–23869.
27. Le, A.; Nguyen, H.; Seo, S.; Bae, Y.S.; Nguyen, T. Biorthogonal Tunable Wavelet Unit with Lifting Scheme in Convolutional Neural Network. In *Proceedings of the 2025 33rd European Signal Processing Conference (EUSIPCO), Palermo, Italy, 8–12 September 2025*; IEEE: Piscataway, NJ, USA, 2025.
28. Li, Y.; Tian, S.; Sun, W.; Chen, L.; Li, J.; Hu, J.; Meng, N. FMW-YOLO: A Frequency-Enhanced and Multi-Scale Context-Aware Framework for PCB Defect Detection. *Micromachines* **2026**, *14*, 531. [\[CrossRef\]](#)
29. 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*.
30. Wei, Y.; Wang, R.; Zhang, M.; Wang, Y.; Zhou, F.; Bian, X. Ade-yolo: Real-time steel surface flaw recognition through enhanced adaptive attention and dilated convolution fusion. *Signal Image Video Process.* **2025**, *19*, 457. [\[CrossRef\]](#)
31. Chakraborti, A.; Chaudhuri, B.B. DSReLU: A Novel Dynamic Slope Function for Superior Model Training. *arXiv* **2024**, arXiv:2408.09156.
32. Heckler-Kram, L.; Neudeck, J.H.; Scheler, U.; König, R.; Steger, C. The MVTec AD 2 Dataset: Advanced Scenarios for Unsupervised Anomaly Detection. *Int. J. Comput. Vis.* **2026**, *134*, 175. [\[CrossRef\]](#)
33. Bhavanasi, G.; Neven, D.; Arteaga, M.; Ditzel, S.; Dehaeck, S.; Bey-Temsamani, A. Enhanced Vision-Based Quality Inspection: A Multiview Artificial Intelligence Framework for Defect Detection. *Sensors* **2025**, *25*, 1703. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Leng, Y.; Liu, J. Improved faster R-CNN for steel surface defect detection in industrial quality control. *Sci. Rep.* **2025**, *15*, 30093. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Thomine, S.; Snoussi, H. Distillation-based fabric anomaly detection. *Text. Res. J.* **2024**, *94*, 14.
36. Rolih, B.; Fučka, M.; Skočaj, D. SuperSimpleNet: Unifying Unsupervised and Supervised Learning for Fast and Reliable Surface Defect Detection. In *Proceedings of the 27th International Conference, ICPR 2024, Kolkata, India, 1–5 December 2024*; Lecture Notes in Computer Science; Springer Nature: Cham, Switzerland, 2025; pp. 47–65.
37. Salehi, M.; Sadjadi, N.; Baselizadeh, S.; Rohban, M.H.; Rabiee, H.R. Multiresolution Knowledge Distillation for Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021*; IEEE: Piscataway, NJ, USA, 2021. [\[CrossRef\]](#)
38. Cohen, N.; Hoshen, Y. Sub-Image Anomaly Detection with Deep Pyramid Correspondences. *arXiv* **2020**, arXiv:2005.02357.
39. Liu, Z.; Zhou, Y.; Xu, Y.; Wang, Z. SimpleNet: A Simple Network for Image Anomaly Detection and Localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023*; pp. 20402–20411.
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.

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.