processes

MDPI

Article

Industrial Product Surface Defect Detection via Memory-Bank Feature Matching and Lightweight Diffusion Reconstruction

Xiaojun Tang [1,2], Jun He [1,2,*], Yixin Tang [1,2], Jihong Sun [1,2] and Xiang Zhao [1,2]

[1] School of Information Engineering, Kunming University, Kunming 650214, China; txjswee@163.com (X.T.); tyxsmt@163.com (Y.T.); sjh918a@163.com (J.S.); zhaoxiang@kmu.edu.cn (X.Z.)
[2] Yunnan Key Laboratory of Intelligent Logistics Equipment and Systems, Kunming 650214, China
* Correspondence: 2508540014@stu.kmu.edu.cn; Tel.: +86-183-8431-3128

Abstract

Surface defect detection of industrial products is a critical component of intelligent manufacturing quality control. However, existing unsupervised anomaly detection methods still struggle to achieve a favorable balance among image-level detection accuracy, pixel-level localization performance, and inference efficiency. To address this issue, this paper proposes a Fusion PatchCore with Lightweight Diffusion Reconstruction (FPC-LDR) framework for industrial surface defect detection, exploiting the complementary properties of discriminative feature matching and generative reconstruction. The proposed method first constructs a memory-bank module based on deep semantic features to identify anomalies via feature matching. Then, a lightweight latent diffusion reconstruction module is developed, in which single-step reconstruction replaces conventional multi-step diffusion processes, reducing computational overhead while maintaining reconstruction quality. Furthermore, an adaptive fusion strategy is introduced to integrate feature distance and reconstruction residuals, improving detection robustness and localization accuracy. Extensive experiments on the MVTec AD dataset demonstrate that FPC-LDR achieves an image-level AUROC of 94.65%, a pixel-level AUROC of 94.54%, and a PRO score of 95.28%. Cross-dataset evaluations on the VisA dataset further verify its generalization capability. The proposed method achieves an effective trade-off among accuracy, localization performance, and inference efficiency, demonstrating its potential for real-time industrial defect inspection.

Keywords: industrial product surface defect detection; unsupervised anomaly detection; memory bank; diffusion reconstruction; feature matching; MVTec AD; VisA

Check for updates

Academic Editor: José de Jesús Rangel Magdaleno

Received: 2 September 2026
Revised: 12 October 2026
Accepted: 17 October 2026
Published: 4 November 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

Surface defect detection in industrial products [1–3] is essential for quality control in intelligent manufacturing, directly affecting product reliability and production efficiency [4]. With advances in computer vision, deep-learning-based automated inspection methods [5–8] have increasingly replaced manual inspection on industrial production lines. However, defective samples are scarce and highly diverse in real-world scenarios [9], making large-scale annotated datasets difficult to obtain. Consequently, normal-only unsupervised anomaly detection (UAD), in which model training uses only defect-free normal samples without defect labels [10], has attracted considerable attention. Although UAD methods have achieved substantial progress on industrial benchmarks such as MVTec AD [11], practical deployment still faces several challenges [3,12,13]. First, defective samples are scarce and highly diverse, making comprehensive defect collection and annotation difficult [9].

Second, industrial surfaces exhibit substantial variations in texture, illumination, product structure, and defect scale, requiring both robust image-level anomaly discrimination and precise pixel-level localization. Third, online inspection imposes strict computational constraints, resulting in an inherent trade-off among detection accuracy, localization precision, and inference efficiency. These practical challenges, summarized in Figure 1, motivate the development of an anomaly-detection framework that can learn from normal samples while maintaining robust localization and efficient inference.

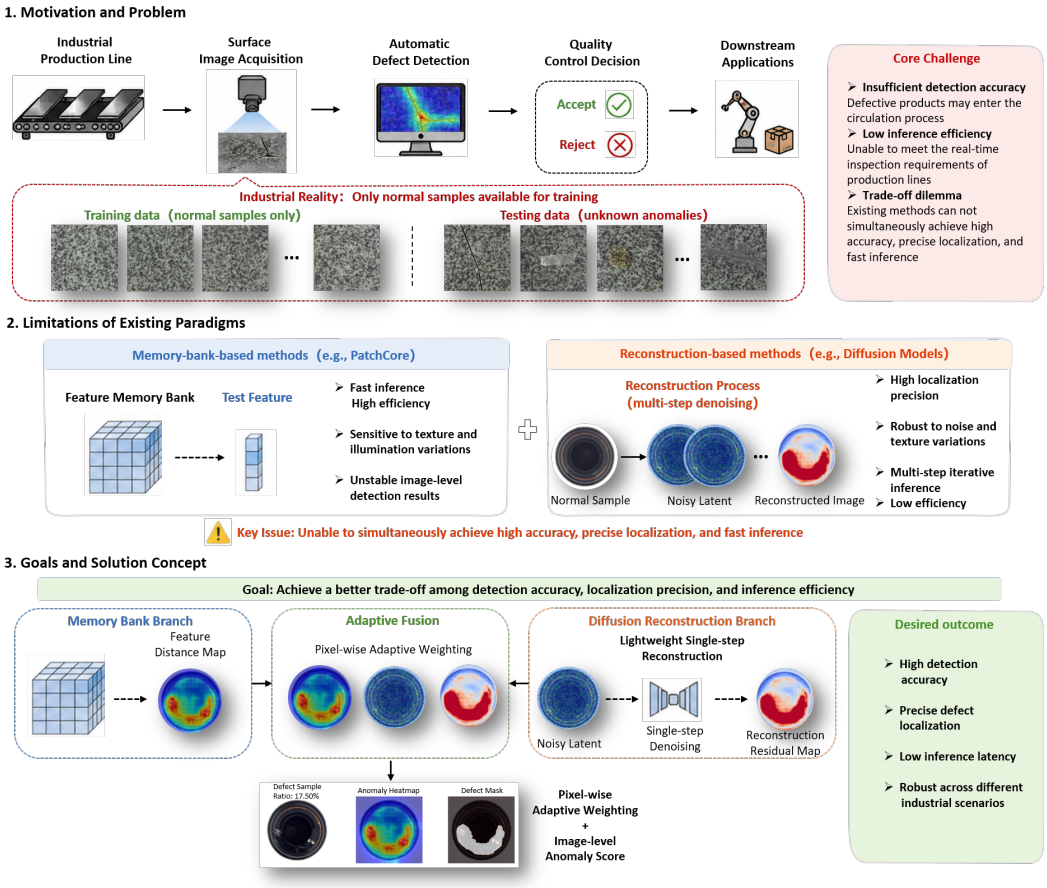


Figure 1. Motivation of integrating memory bank and diffusion reconstruction for industrial surface defect detection. Highlighted regions indicate defect locations, the middle panel illustrates the diffusion-based noise-injection process, and red regions represent defects.

Recent advances in contrastive learning [14], zero-shot anomaly detection [15,16], diffusion models [17], and multi-class anomaly detection [18] have further promoted the transition from supervised learning toward zero- and few-shot paradigms [19]. Memory-bank-based methods, represented by PatchCore [20–22], detect anomalies by constructing a repository of normal features [22,23] and measuring feature distances. These methods offer simple training and efficient inference, and their performance has been further improved through multi-scale feature fusion [24,25], self-attention mechanisms [21], and lightweight designs [26,27]. Nevertheless, their anomaly scores rely heavily on local feature matching and are therefore sensitive to texture variations, noise, and illumination changes, which can lead to unstable image-level predictions.

More recent studies have further expanded this landscape through diffusion-based industrial anomaly modeling and defect synthesis [28–30], reconstruction–matching integration [31], and cross-modal zero-shot anomaly detection [32,33]. These studies demonstrate the rapid development of industrial anomaly detection; however, most primarily improve

a specific paradigm, such as reconstruction quality, synthetic defect generation, representation alignment, or zero-shot transfer. The joint optimization of image-level anomaly discrimination, pixel-level localization, and inference efficiency under normal-only training remains insufficiently explored.

Distribution-modeling methods constitute another important branch of discriminative anomaly detection. PaDiM [34] relies on a Gaussian modeling assumption, which may limit its ability to represent complex feature distributions. In contrast, SPADE [35] relies on deep feature correspondences and may be sensitive to variations in feature matching across complex surface patterns. Generative reconstruction methods [36,37] learn the distribution of normal samples and localize defects using reconstruction errors. Diffusion models have recently attracted considerable interest because of their stable training and high-quality reconstruction capabilities [38,39]. DiffusionAD [38] demonstrates their advantages in pixel-level defect localization, while latent diffusion models [39,40] reduce computational costs. EfficientAD [41] adopts a lightweight student–teacher architecture [42] to achieve millisecond-level inference while maintaining competitive detection accuracy. Nevertheless, conventional diffusion models generally require tens or hundreds of iterative denoising steps during inference [43,44], limiting their applicability to real-time industrial inspection. Existing one-step diffusion methods [38,39] improve inference efficiency, although accelerated reconstruction may compromise reconstruction quality or fine-grained localization.

Overall, memory-bank methods provide efficient inference but limited detection stability [20,27,45], whereas diffusion-based reconstruction methods offer accurate localization at a high computational cost [38,43,44]. Their complementary characteristics motivate the development of a unified framework that simultaneously improves detection accuracy, localization precision, and inference efficiency.

To this end, this study proposes Fusion PatchCore with Lightweight Diffusion Reconstruction (FPC-LDR) for industrial surface defect detection. FPC-LDR first constructs a memory-bank branch that extracts multi-scale semantic features from normal samples, thereby establishing a reference feature distribution. A lightweight latent diffusion reconstruction branch is then introduced, replacing conventional multi-step sampling with single-step reconstruction to accelerate inference while preserving reconstruction quality. Subsequently, an adaptive weighting mechanism dynamically combines the memory-bank and diffusion anomaly maps using learned branch-level weights. Finally, the image-level anomaly scoring strategy is optimized to improve detection stability and robustness. Figure 1 illustrates the motivation and objective of the proposed method.

The novelty of FPC-LDR lies not in simply combining two existing detectors, but in coordinating them at different functional levels: lightweight single-step latent diffusion supplies region-sensitive reconstruction evidence; memory-bank matching provides efficient discriminative feature-distance evidence; adaptive branch-level weighting dynamically reconciles their heterogeneous anomaly responses; and image-level score optimization further improves decision stability. Consequently, detection accuracy, localization completeness, and inference efficiency are jointly addressed within a unified normal-only framework.

The main contributions of this study are summarized as follows:

- A lightweight single-step latent diffusion reconstruction branch is developed as a region-sensitive complementary pathway. By replacing iterative reverse denoising with one denoising-network evaluation, the reconstruction branch reduces diffusion inference overhead while retaining useful reconstruction-residual cues for defect localization.
- A complementary dual-branch architecture is proposed to jointly exploit the efficient local discrimination of memory-bank feature matching and the spatially continuous anomaly evidence provided by diffusion reconstruction. This design targets the differ-

ent failure modes of the two paradigms rather than merely increasing the capacity of a single detector.

- An adaptive branch-level fusion mechanism and an optimized image-level anomaly-scoring strategy are introduced to dynamically reconcile heterogeneous feature-distance and reconstruction-residual distributions, thereby improving image-level decision stability and suppressing isolated local noise.
- Extensive experiments on the MVTec AD and VisA datasets demonstrate the effectiveness of FPC-LDR in detection accuracy, localization performance, and inference efficiency, as well as its generalization across industrial scenarios.

2. Related Work

Among existing unsupervised industrial anomaly detection approaches [46], two paradigms particularly relevant to this work are memory-based feature matching and generative reconstruction. Recent advances in multimodal learning and attention mechanisms [47–49] have also encouraged the integration of multi-source information [50] and attention-enhanced representations, offering new solutions and benchmarks [51] for complex industrial anomaly detection.

2.1. Memory-Bank-Based Methods

Memory-bank-based approaches [20–23] detect anomalies by measuring the distance between test features and stored representations of normal samples. PatchCore [20–22] employs coreset sampling to reduce the memory bank size while maintaining strong detection performance, making it a widely adopted baseline. Subsequent studies have improved this paradigm through multi-scale feature fusion [24], self-attention modeling [21], lightweight deployment [26] and simplified discriminative architectures [52]. Despite their simple training procedure and efficient inference, memory-bank methods fundamentally depend on local feature-distance matching [23]. This makes them sensitive to texture variations, illumination changes, and noise; in complex scenes, even normal regions may produce large feature distances [20,27], leading to unstable image-level predictions. Therefore, their main limitation lies in robust global anomaly discrimination rather than computational efficiency.

2.2. Generative Reconstruction Methods

Generative reconstruction methods [36,37] learn representations of normal samples and detect anomalies through reconstruction discrepancies. Diffusion models have also been explored for realistic defect synthesis and anomaly-data generation [53,54]. Diffusion models have recently attracted considerable attention for their stable training, high-quality generation, and strong distributional modeling capabilities. DiffusionAD [38] introduces a norm-guided mechanism to improve reconstruction quality and defect localization, while latent diffusion models [39,40] reduce computational complexity. To address inefficient inference [43,44], several acceleration strategies have been investigated, including multi-scale guided denoising [44] and one-step diffusion [38,39]. Nevertheless, conventional diffusion models still require iterative denoising, whereas accelerated variants may compromise reconstruction quality or fine-grained localization. Achieving efficient inference without sacrificing reconstruction capability therefore remains a major challenge. Unlike diffusion-centered methods such as DiffusionAD [38], FPC-LDR does not rely on reconstruction evidence alone; instead, single-step latent diffusion is used as complementary generative evidence and is adaptively fused with non-parametric memory-bank feature matching to jointly improve anomaly discrimination and localization.

2.3. Motivation for Dual-Branch Anomaly Detection

Existing methods exhibit a clear trade-off between inference efficiency and localization accuracy. Memory-bank approaches are computationally efficient but may produce unstable image-level predictions, whereas diffusion-based reconstruction achieves accurate pixel-level localization at a relatively high computational cost. Although multimodal information [47,50], attention mechanisms [21,47], and feature-fusion strategies [48,49] have been explored, most existing methods rely on fixed fusion weights or a single detection mechanism. Consequently, the key research gap lies in how to jointly exploit the efficient discrimination of memory-bank matching and the region-sensitive reconstruction capability of diffusion models, particularly while maintaining stable image-level scoring and efficient inference.

To address these limitations, the proposed FPC-LDR combines lightweight single-step diffusion reconstruction for accurate pixel-level localization with a memory bank for efficient anomaly perception. An adaptive weighted fusion mechanism dynamically integrates the two anomaly scores, achieving a better balance among detection accuracy, localization precision, and inference efficiency.

3. Methods

Existing industrial surface defect detection methods still struggle to simultaneously achieve high image-level detection accuracy, precise pixel-level localization, and efficient inference. To address this challenge, this study proposes FPC-LDR, which exploits the complementary characteristics of memory-bank-based feature matching and diffusion-based reconstruction. Memory-bank methods offer simple training and efficient inference but are sensitive to texture variations and environmental noise, resulting in limited detection stability. In contrast, diffusion-based reconstruction methods accurately localize pixel-level defects by learning the distribution of normal samples; however, their iterative denoising process incurs substantial inference latency, limiting their applicability to real-time production-line inspection. These two paradigms are inherently complementary, and their effective integration can overcome the performance limitations of a single detection mechanism.

Accordingly, FPC-LDR adopts a modular dual-branch architecture. The memory-bank branch uses multi-scale feature enhancement to obtain more discriminative representations of normal patterns. The lightweight single-step latent diffusion branch reduces inference cost while preserving reconstruction and localization capability. Finally, adaptive weighted fusion integrates the anomaly maps generated by the two branches using learned branch-level weights. The following sections describe these components in detail.

3.1. Overall Architecture of FPC-LDR

The overall architecture of the proposed method is illustrated in Figure 2, while its detailed implementation and data flow are presented in Figure 3. FPC-LDR adopts a modular dual-branch architecture comprising a configurable memory-bank branch and a lightweight single-step latent diffusion branch (Figure 4a). Their complementary strengths are integrated through an adaptive weighted fusion mechanism.

During training, only normal samples are used. The memory-bank branch extracts multi-scale deep semantic features from normal images and constructs a representative memory bank through coreset sampling. In parallel, the lightweight diffusion branch learns the distribution of normal samples in the latent space and trains a single-step denoising network. The two branches are trained independently to prevent mutual interference.

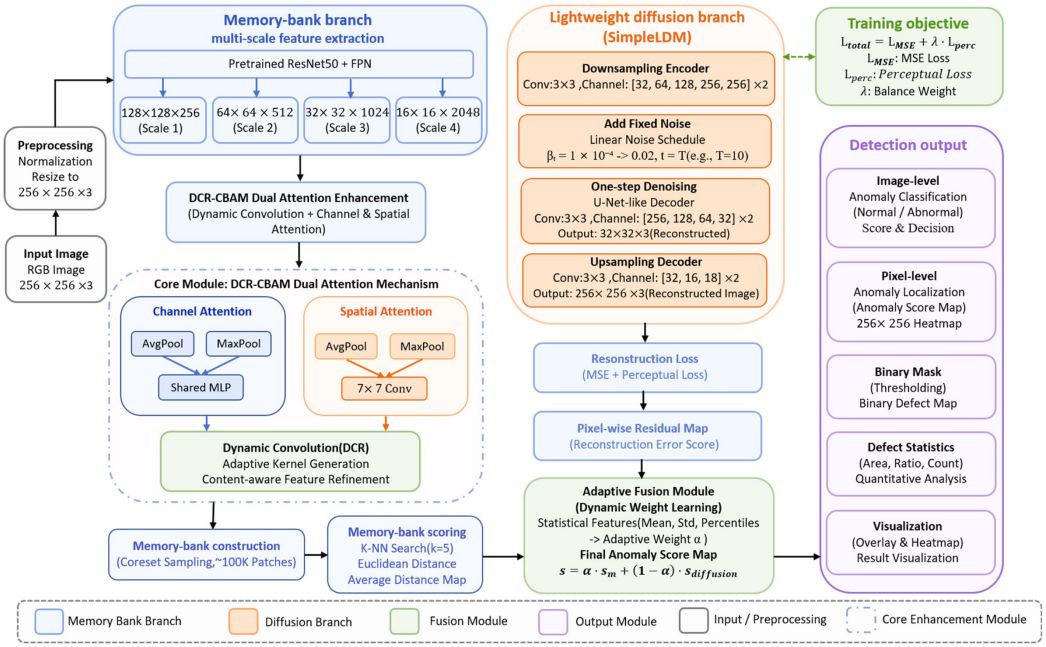


Figure 2. Overall end-to-end architecture of FPC-LDR.

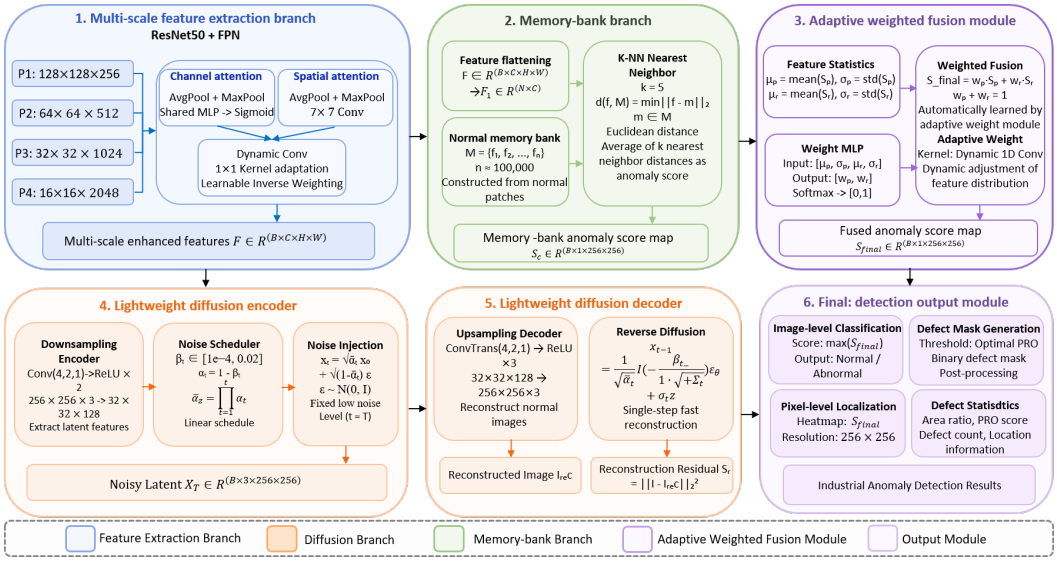


Figure 3. Detailed implementation and data flow of the core FPC-LDR modules.

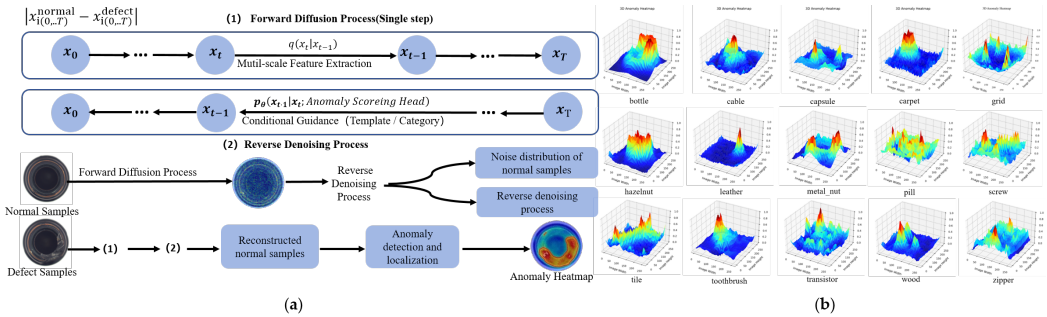


Figure 4. Lightweight latent diffusion reconstruction and 3D anomaly heatmaps. (**a**) Forward noise injection, single-step denoising, and defect localization. (**b**) 3D anomaly heatmaps for the 15 MVTec AD categories. Highlighted regions indicate defect locations.

During inference, each test image is processed by both branches simultaneously. The memory-bank branch generates a pixel-level anomaly map based on feature distances, whereas the diffusion branch performs rapid single-step reconstruction to obtain a pixel-level residual map. These two maps are adaptively weighted and fused to produce the final pixel-level anomaly map. An optimized image-level scoring strategy is subsequently applied to predict the image-level anomalies.

The main advantage of this architecture lies in its modular and decoupled design, which allows the two branches to be optimized independently. The single-step diffusion reconstruction strategy improves inference efficiency while preserving pixel-level localization accuracy. Moreover, adaptive fusion and image-level score optimization address inconsistencies in score distributions and erroneous score orientations commonly encountered in conventional methods. The following sections describe each component of FPC-LDR in detail.

The two anomaly maps are dynamically combined using adaptively learned branch-level weights. Finally, the framework produces image-level classification results and pixel-level defect localization maps, effectively combining the complementary strengths of discriminative and generative anomaly detection paradigms. Figure 4a illustrates the principle of diffusion-based reconstruction, while Figure 4b presents the corresponding 3D defect heatmaps.

3.2. Lightweight Single-Step Latent Diffusion Reconstruction Module

The proposed diffusion branch is designed specifically for anomaly reconstruction rather than general-purpose sample generation. During training, a normal image is first encoded into the latent space, where Gaussian noise is injected once at a predefined noise level. The denoising network is trained to predict the injected noise from this single corrupted latent state. During inference, the same fixed noise setting is used, and only one denoising-network evaluation is performed to estimate the clean latent representation through the analytical reverse relation. Therefore, no iterative timestep-by-timestep denoising is required.

This differs from a conventional denoising auto-encoder, which directly learns corrupted-to-clean mapping; the proposed module explicitly predicts Gaussian noise under the diffusion formulation and reconstructs the latent representation accordingly. One reverse step is sufficient because the objective is to recover the normal counterpart of an observed industrial sample, rather than generate diverse samples from random noise. The influence of diffusion noise design is further evaluated in Section 4.5.2.

The fundamental principle of diffusion models is illustrated in Figure 4a. Gaussian noise is progressively added to the data during the forward diffusion process, while a reverse denoising process is learned to recover the underlying data distribution. Given clean data x_0, the forward process generates a sequence of noisy samples $(x_1, x_2, \ldots, x_T)$:

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon_t, \epsilon_t \sim N(0, I) \quad (1)$$

where $\alpha_t = 1 - \beta_t$ denotes the noise schedule. Defining $\overline{\alpha_t} = \prod_{s=1}^{t} \alpha_s$, a sample at an arbitrary timestep can be obtained directly as

$$z_t = \sqrt{\overline{\alpha}_t} z_0 + \sqrt{1-\overline{\alpha}_t}\epsilon_t, \epsilon_t \sim N(0, I) \quad (2)$$

To reduce computational complexity, the diffusion process is transferred from the high-dimensional image space to a compact latent space using a latent diffusion model (LDM). An autoencoder {E,D} is first employed, where the encoder E maps an image x to a latent representation z, and the decoder D reconstructs the image: $z = E(x), x = D(z)$.

The autoencoder is optimized using

$$L_{AE} = \| x - D(E(x)) \|_2^2 + \lambda_{KL} D_{KL}\left(N\left(\mu_z, \sigma_z^2\right) \| N(0, I)\right) \tag{3}$$

where λ_{KL} controls the regularization imposed on the latent distribution. In the latent space, the forward diffusion process and the reverse mean estimator are respectively expressed as

$$\mu_\theta(z_t, t) = \frac{1}{\sqrt{\alpha_t}}(z_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon_\theta(z_t, t)) \tag{4}$$

Because the latent representation has substantially lower dimensionality than the original image, latent diffusion considerably reduces computational cost.

Industrial anomaly detection aims to reconstruct the normal counterpart of a potentially anomalous input rather than generate diverse samples. Based on this characteristic, a lightweight single-step latent diffusion strategy is developed to replace iterative denoising with a single network evaluation. The model considers a clean normal latent variable z_0 and its noisy counterpart z_1. The single-step noise schedule is defined as $\overline{\alpha_1} == \alpha = \tau, \beta_1 = 1-\tau$, where τ controls the signal retained after noise injection. The same single diffusion state and fixed noise parameter are used during both training and inference. The forward process is given by

$$z_1 = \sqrt{\tau} z_0 + \sqrt{1-\tau}\epsilon, \epsilon \sim N(0, I) \tag{5}$$

The clean latent representation is estimated in a single reverse step:

$$\hat{z}_0 = \frac{1}{\sqrt{\tau}}\left(z_1 - \sqrt{1-\tau}\epsilon_\theta(z_1, 1)\right) \tag{6}$$

The denoising network is trained exclusively on normal samples by minimizing the noise-prediction error:

$$L_{diffusion} = E_{z_0 \sim p(z_0), \epsilon \sim N(0,I)}\left[\| \epsilon - \epsilon_\theta\left(\sqrt{\tau} z_0 + \sqrt{1-\tau}\epsilon, 1\right) \|_2^2\right] \tag{7}$$

Compared with a conventional T-step reverse diffusion process requiring T repeated denoising-network evaluations, the proposed strategy performs only one such evaluation, reducing reverse-process evaluations by $(T-1)/T$. For example, a 50-step setting corresponds to 98% fewer denoising evaluations. This computational simplification is consistent with previous one-step diffusion studies such as DiffusionAD [38], which reported substantial acceleration over iterative denoising while preserving reconstruction quality.

During inference, an input image x is first encoded as $z_0 = E(x)$. A noisy latent representation z_1 is generated according to Equation (5), after which the denoising network estimates z_0 using Equation (6). The reconstructed normal image is then obtained as $\hat{x} = D(\hat{z}_0)$. The pixel-level anomaly map of the diffusion branch is calculated from the absolute reconstruction residual:

$$S_{diffusion}(h, w) = |x(h, w) - \hat{x}(h, w)| \tag{8}$$

where (h,w) denotes the pixel coordinates.

3.3. Modular and Configurable Memory-Bank Module

The memory-bank branch adopts a modular design that supports multiple backbone networks, including ResNet18, ResNet50, and MobileNetV2. Based on comparative experiments, ResNet50 was selected as the optimal backbone. Given an input image $x \in R^{3\times H\times W}$, an ImageNet-pretrained convolutional neural network ϕ is employed to extract feature maps from the third and fourth stages:

$$f_l = \phi_l(x), l \in \{3,4\} \tag{9}$$

where $f_l \in R^{C_l \times H_l \times W_l}$ denotes the feature map extracted from stage l, and C_l, H_l, and W_l represent its number of channels, height, and width, respectively. After spatial alignment, the multi-scale feature maps are concatenated along the channel dimension and flattened into patch-level feature vectors:

$$F = Flatten(Concat(f_3, f_4)) = \{f_i\}_{i=1}^{N}, f_i \in R^{C_3+C_4} \tag{10}$$

where $Up(\cdot)$ denotes spatial upsampling and $N = H_3 \times W_3 + H_4 \times W_4$ is the total number of patch features after spatial alignment. To reduce the memory-bank size while retaining representative normal features, coreset sampling is applied to the patch features extracted from all training images. A representative subset M is selected as follows:

$$M = Coreset(F_{train}, k) \tag{11}$$

where F_{train} denotes the complete set of training patch features and k is the memory-bank size. In this study, k is set to 10% of $|F_{train}|$. Coreset sampling employs a greedy selection strategy to identify a compact subset that provides broad coverage of the normal feature space, thereby reducing memory consumption while preserving detection performance. For each patch feature f_i extracted from a test image, its Euclidean distances to all features in the memory bank are calculated. The K nearest features are then selected, and their mean Euclidean distance is used as the patch-level anomaly score. In this study, $K = 5$:

$$s_{patchcore}(f_i) = \frac{1}{K}\sum_{j=1}^{K} \| f_i - m_j \|_2 \tag{12}$$

Finally, the anomaly scores of all patches are spatially rearranged and upsampled to the original image resolution, thereby producing the pixel-level anomaly map of the memory-bank branch: $s_{patchcore} \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}}$.

3.4. Adaptive Weighted Fusion

The memory-bank and diffusion branches generate complementary anomaly maps, denoted as $\mathrm{S}_{\mathrm{mem}}$ and $\mathrm{S}_{\mathrm{rec}}$, respectively. Before fusion, both maps are normalized to a common scale. Their global statistical characteristics, including the mean and standard deviation, are then extracted and concatenated as

$$\begin{gathered} \mathrm{z} = \left[\mu_{\mathrm{mem}}, \sigma_{\mathrm{mem}}, \mu_{\mathrm{rec}}, \sigma_{\mathrm{rec}}\right] \\ \mu_{\mathrm{mem}} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{\omega=1}^{W} \mathrm{S}_{\mathrm{mem}}(h,w), \sigma_{\mathrm{mem}} = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{\omega=1}^{W}\left(\mathrm{S}_{\mathrm{mem}}(h,w) - \mu_{\mathrm{mem}}\right)^2} \end{gathered} \tag{13}$$

A two-layer multilayer perceptron (MLP) is employed to estimate the relative contributions of the two branches: $\mathrm{a} = \mathrm{MLP}(\mathrm{z})$, where $\mathrm{a} = [\mathrm{a}_{\mathrm{mem}}, \mathrm{a}_{\mathrm{rec}}]$. The fusion weights are normalized using a Softmax function: $[\omega_{mem}, \omega_{rec}] = \mathrm{Softmax}(\mathrm{a})$ with $\omega_{mem} + \omega_{rec} = 1$. The final anomaly map is obtained as

$$S_{fusion} = \omega_{mem} S_{mem} + \omega_{rec} S_{rec} \tag{14}$$

This adaptive strategy dynamically balances discriminative feature-matching responses and generative reconstruction residuals, improving robustness across different industrial surface textures and defect patterns.

3.5. Image-Level Anomaly Scoring and Automatic Direction Calibration

To improve the stability of image-level predictions and address potential anomaly-score inversion, an image-level scoring strategy is developed that jointly considers the spatial extent and peak intensity of anomalous regions. First, the defect-area ratio R_{defect} is

calculated as the proportion of pixels with fused anomaly scores that exceed a predefined threshold t:

$$R_{\text{defect}} = \frac{\sum_{h,w} \mathbb{I}(S_{\text{fusion}}(h,w) > t)}{H \times W} \tag{15}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. In this study, the detection threshold is the candidate value for continuous anomaly scores traversed on the validation set. The defect-area ratio is then combined with the maximum anomaly score to obtain the preliminary image-level anomaly score:

$$s_{img_raw} = \alpha \cdot R_{defect} + (1 - \alpha) \cdot max(S_{fusion}) \tag{16}$$

where α is a weighting coefficient set to 0.6. This formulation incorporates both the spatial extent of a potential defect and its strongest local response, thereby reducing the sensitivity of image-level prediction to isolated noisy pixels. The score direction is subsequently determined using the validation set. The validation labels are used only for post-training threshold and score-direction calibration and do not participate in model-parameter optimization. Specifically, the mean preliminary scores of normal and anomalous validation samples are calculated as

$$\mu_{\text{normal}} = \mathrm{E}[s_{\text{img_raw}} | y = 0], \quad \mu_{\text{defect}} = \mathrm{E}[s_{\text{img_raw}} | y = 1] \tag{17}$$

If the mean score of normal samples exceeds that of anomalous samples, the score direction is regarded as inverted and is corrected as follows:

$$s_{img} = \begin{cases} s_{img_raw} & if\ \mu_{defect} > \mu_{normal} \\ 1 - s_{img_raw} & otherwise \end{cases} \tag{18}$$

This strategy improves the robustness of image-level scoring for categories with complex textures. Experimental results show that it enables FPC-LDR to achieve an image-level AUROC exceeding 94%.

Taking the grid category as an example (Table 1), its dense periodic textures make anomaly-score discrimination difficult. Before calibration, normal samples have higher mean scores than defective samples, indicating an inverted score direction and yielding an image-level AUROC of 11.36%. Score-direction calibration identifies this inversion using the class-wise mean scores and reverses the score orientation according to Equation (18). After calibration, defective samples obtain higher anomaly scores than normal samples, and the image-level AUROC increases to 88.64%, demonstrating that the calibration effectively corrects inverted anomaly-score orientation under complex textures.

Table 1. Effect of score-direction calibration on the grid category.

Metric	Before Calibration	After Calibration
Mean score of normal samples	2.5497	−1.5497
Mean score of defective samples	2.4570	−1.4570
Image-AUROC (%)	11.36	88.64

4. Experiments

This section evaluates FPC-LDR for industrial defect detection. All models are trained exclusively on defect-free normal samples, while evaluation is performed on test sets containing both normal and anomalous samples. The experiments cover dataset settings, implementation details, comparative evaluation, ablation studies, and cross-dataset validation on VisA.

The overall experimental pipeline is illustrated in Figure 5 and comprises training and inference stages. The training stage consists of memory-bank construction, diffusion-model optimization, and fusion-module training. In the memory-bank branch, multi-scale features are extracted and enhanced using an attention mechanism, after which coreset sampling is applied to construct a compact repository of normal features. The diffusion branch learns the distribution of normal samples by training a lightweight single-step reconstruction model. Finally, an adaptive weighting network is optimized based on the statistical characteristics of the anomaly scores produced by the two branches.

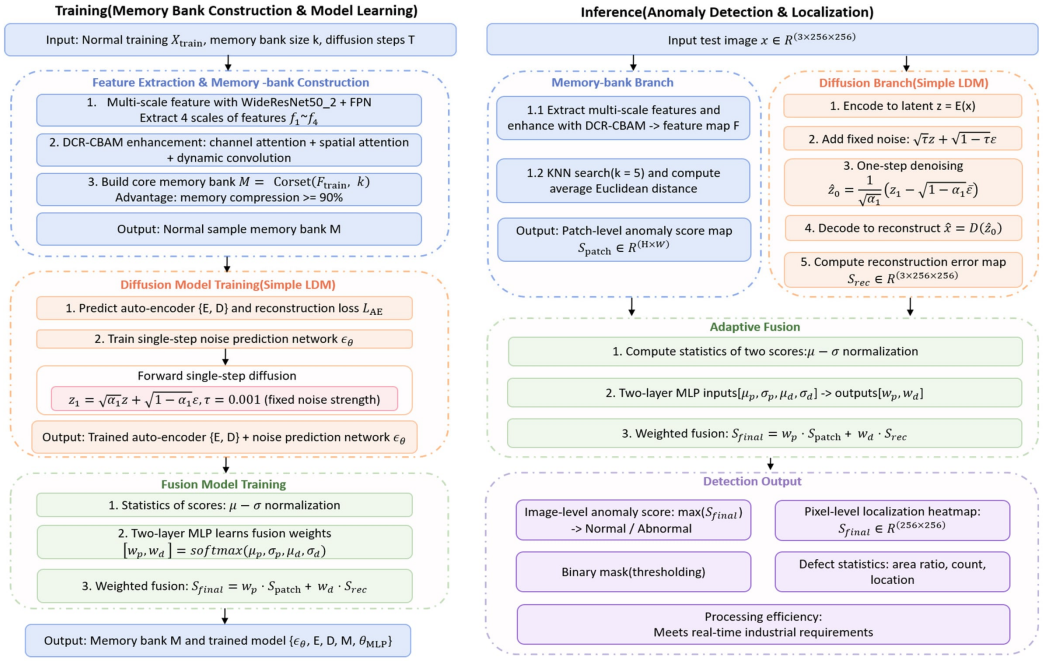


Figure 5. Workflow of the FPC-LDR algorithm.

During inference, each test image is processed in parallel by both branches. The memory-bank branch generates a patch-level anomaly map through k-nearest-neighbor feature matching, whereas the diffusion branch produces a pixel-level reconstruction-error map through single-step denoising and reconstruction. After normalization, the two anomaly maps are dynamically integrated by the adaptive fusion module. The framework ultimately outputs an image-level prediction, a pixel-level localization heatmap, and quantitative defect statistics. Its low per-image inference latency indicates its potential for real-time industrial inspection.

4.1. Datasets and Data Preprocessing

The primary experiments were conducted on MVTec AD [11], a widely used benchmark for industrial anomaly detection. The dataset comprises 15 categories, including five texture categories (carpet, grid, leather, tile, and wood) and ten object categories (e.g., bottle, cable, and capsule). It contains 3629 defect-free training images and 1725 test images covering 73 types of real-world industrial defects, such as scratches, dents, contamination, and structural deformation. Pixel-level ground-truth annotations are provided for all defective regions, supporting both image-level anomaly classification and pixel-level defect localization.

The VisA dataset [55] was further employed to evaluate cross-dataset generalization and robustness. VisA contains 12 industrial object categories and features high-resolution images, complex background textures, diverse defect patterns, and substantial intra-class

variation, making it a challenging benchmark for generalization evaluation. All model architectures and training hyperparameters established on MVTec AD were directly transferred to VisA without dataset-specific model tuning. The operating threshold was calibrated separately using the designated validation data.

All input images were resized to 256 × 256 pixels to ensure compatibility with the backbone networks. Pixel values were normalized using the ImageNet mean and standard deviation, consistent with the preprocessing protocol for the pretrained backbones. No additional data augmentation was applied during training to preserve the original distribution of normal samples and maintain the rigor of the unsupervised training protocol.

4.2. Experimental Settings and Evaluation Metrics

4.2.1. Implementation Details and Parameter Settings

All experiments were implemented using the PyTorch (2.4.1+cu121) framework and conducted on a server equipped with an NVIDIA V100 GPU with 32 GB of memory. The principal configurations of FPC-LDR are described below.

Feature extraction branch: An ImageNet-pretrained ResNet50 was employed as the backbone. A feature pyramid network (FPN) generated feature maps at four scales: 128 × 128 × 256, 64 × 64 × 512, 32 × 32 × 1024, and 16 × 16 × 2048. These features were enhanced using the DCR-CBAM dual-attention module before being stored in the memory bank.

Memory-bank branch: Coreset sampling was applied to compress the normal features and construct a memory bank containing approximately 100,000 feature vectors. A k-nearest-neighbor search with k = 5 was then performed, and the mean Euclidean distance to the five nearest neighbors was used to generate the patch-level anomaly map.

SimpleLDM diffusion branch: SimpleLDM consisted of a three-layer downsampling encoder and a three-layer upsampling decoder. A linear noise schedule with a fixed parameter τ was adopted. Normal-image reconstruction was performed using a single-step reconstruction mechanism, and the pixel-wise reconstruction error was calculated to produce the diffusion-based anomaly map.

Dynamic weighted fusion module: A two-layer multilayer perceptron (MLP) adaptively learned the fusion weights from the means and standard deviations of the anomaly maps generated by the two branches. The predicted weights were normalized with a Softmax function and then used to combine the two anomaly maps, producing the final anomaly score map.

4.2.2. Evaluation Metrics

(1) Image-level AUROC. Image-level AUROC evaluates the ability of a model to distinguish anomalous images from normal ones. Given the image-level ground-truth labels and predicted anomaly scores, true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) are calculated at different decision thresholds. The true-positive rate (TPR) and false-positive rate (FPR) are defined as

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \tag{19}$$

By varying the threshold across all possible anomaly scores, the receiver-operating characteristic (ROC) curve is constructed from the resulting (FPR,TPR) pairs. Image-level $AUROC$ is the area under this curve:

$$AUROC_{image} = \int_0^1 TPRd(FPR) \tag{20}$$

The value ranges from 0 to 1, with a higher value indicating stronger image-level anomaly classification performance.

(2) Pixel-level *AUROC*. Pixel-level AUROC assesses pixel-wise defect localization performance. Each pixel is treated as an independent binary sample, with the ground-truth mask providing its label and the predicted anomaly map providing its anomaly score. Pixel-level *TP*, *FP*, *TN*, and *FN* values are calculated across all thresholds to construct the pixel-level *ROC* curve. The corresponding area under the curve is reported as pixel-level *AUROC*:

$$AUROC_{pixel} = \int_0^1 TPR_{pixel} d\left(FPR_{pixel}\right) \tag{21}$$

A higher value indicates a stronger overall ability to discriminate defective pixels from normal pixels.

(3) PRO score by defect type. The per-region overlap (*PRO*) score evaluates localization performance at the connected-component level. Unlike pixel-level AUROC, which can be dominated by large anomalous regions, PRO assigns equal importance to each defect region and is therefore better suited to evaluating defects of varying sizes.

First, connected-component analysis is applied to the pixel-level ground-truth masks, producing N independent defect regions G_1,G_2,...,G_N. Given a binary prediction mask P obtained at a specific threshold, the overlap ratio for the i-th ground-truth region is calculated as

$$PRO_i = \frac{|G_i \cap P|}{|G_i|} \tag{22}$$

where $|G_i \cap P|$ denotes the number of correctly detected pixels within G_i, and $|G_i|$ is the total number of pixels in that region. The mean regional overlap at the given threshold is

$$PRO = \frac{1}{N}\sum_{i=1}^{N} PRO_i \tag{23}$$

The threshold is varied to construct the PRO–FPR curve. The normalized area under this curve within the commonly adopted range FPR $\in [0, 0.3]$ is reported as the *PRO* score:

$$PRO - score = \frac{1}{0.3}\int_0^{0.3} PRO d(FPR) d(FPR) \tag{24}$$

This region-based metric is particularly sensitive to small defects and provides a more balanced evaluation across defect types and sizes.

(4) Decision threshold. The decision threshold converts a continuous anomaly map into a binary defect mask, making it essential for practical defect identification and quantitative analysis. In this study, the operating threshold is selected on an independent validation set by optimizing a predefined segmentation criterion subject to the required false-positive rate constraint. The selected threshold is then fixed and applied to the test set, thereby balancing defect recall and false detections without introducing test-set information.

(5) Inference time. Inference time measures the computational efficiency of the proposed method and is reported in milliseconds per image. Under a fixed hardware configuration, all test images are processed using the same inference settings. The end-to-end model execution time—from model input to generation of the complete anomaly map—is recorded, excluding data loading and preprocessing overhead. After an initial warm-up, GPU synchronization is performed before and after timing, and the average over repeated runs is reported. A lower inference time indicates greater suitability for real-time industrial inspection.

4.3. Analysis of Experimental Results

Table 2 reports the category-wise performance of FPC-LDR on MVTec AD in terms of image-level AUROC, pixel-level AUROC, PRO score, optimal detection threshold, and average inference time per image.

Table 2. Category-wise performance of FPC-LDR on the MVTec AD dataset.

Class	Image-AUROC (%)	Pixel-AUROC (%)	PRO (%)	Threshold	Inference Time (ms)
bottle	99.76	97.43	94.40	1.3531	27.5479
cable	98.93	97.15	92.50	2.0060	24.8237
capsule	88.55	96.67	90.45	1.0469	23.4968
carpet	96.59	97.83	98.17	1.4382	29.5039
grid	88.64	98.28	83.68	1.5235	29.3211
hazelnut	93.04	96.86	96.75	1.9833	33.0660
leather	90.52	99.41	97.72	1.4224	24.1424
metal_nut	98.48	89.87	99.21	1.6887	27.6758
pill	92.39	85.46	91.31	1.4285	27.3118
screw	99.88	98.39	99.50	1.2538	28.7417
tile	96.07	94.27	97.40	1.8840	29.2100
toothbrush	90.28	94.96	99.78	1.3342	9.7509
transistor	93.79	81.00	94.47	1.7745	21.5545
wood	99.56	93.94	94.25	1.6708	30.0200
zipper	93.30	96.53	99.56	1.1709	25.9205

This table reports the defect-detection performance of FPC-LDR across the 15 industrial categories in the MVTec AD dataset, including image-level AUROC, pixel-level AUROC, PRO score, optimal threshold, and inference time.

Averaged over all 15 categories, FPC-LDR achieves an image-level AUROC of 94.65%, a pixel-level AUROC of 94.54%, and a PRO score of 95.28%, with an average inference time of 26.14 ms per image. These results demonstrate stable detection performance across industrial objects with diverse textures and structures. The inference efficiency also indicates the method's potential for real-time industrial inspection. Figure 4b shows the 3D anomaly heatmaps for all 15 MVTec AD categories. Defective regions exhibit distinct high-scoring peaks, illustrating the model's localization responses to different defect types.

Figure 6 presents the binary confusion matrices for the 15 MVTec AD categories. In each matrix, the rows represent the ground-truth labels and the columns denote the predicted labels. The results show that FPC-LDR correctly identifies most normal and anomalous samples across most categories, with relatively few false negatives and false positives. This demonstrates stable image-level classification performance for diverse industrial defects.

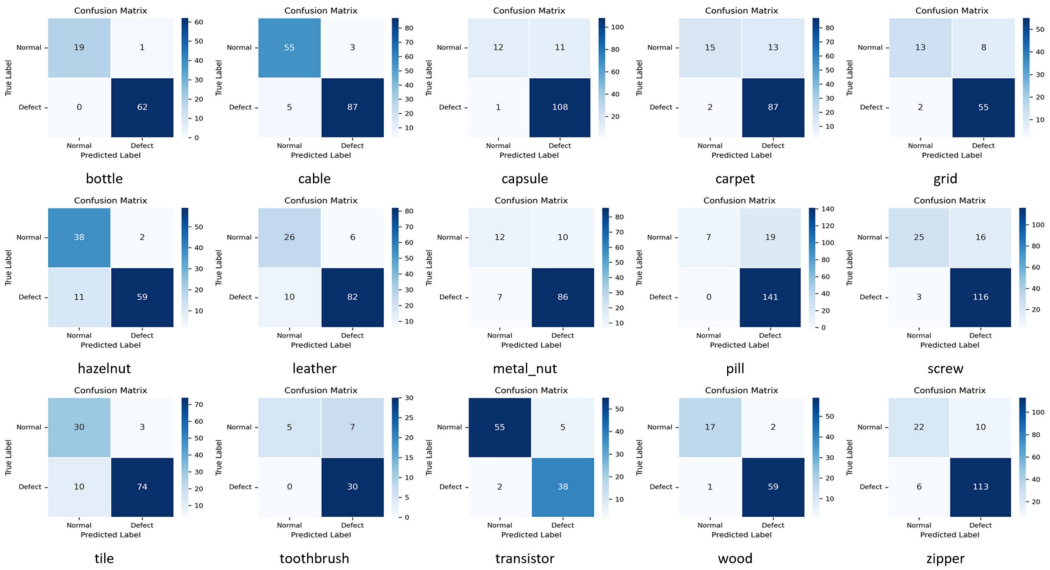


Figure 6. Confusion matrices for different product defect types.

Figure 7 presents the pixel-level localization results of FPC-LDR across the 15 industrial categories in MVTec AD. From left to right, each group shows the defective input image, the anomaly heatmap, the mask overlay, and the binary ground-truth mask. The results cover both object and texture categories and include various defect patterns, such as scratches, missing parts, contamination, dents, and structural deformation. The high-response regions in most anomaly maps are well aligned with the ground-truth defects, demonstrating effective localization across diverse defect patterns. However, as shown by the representative cases below the red dashed line in Figure 7, extremely small or low-contrast defects may still exhibit partial boundary omission or incomplete region coverage.

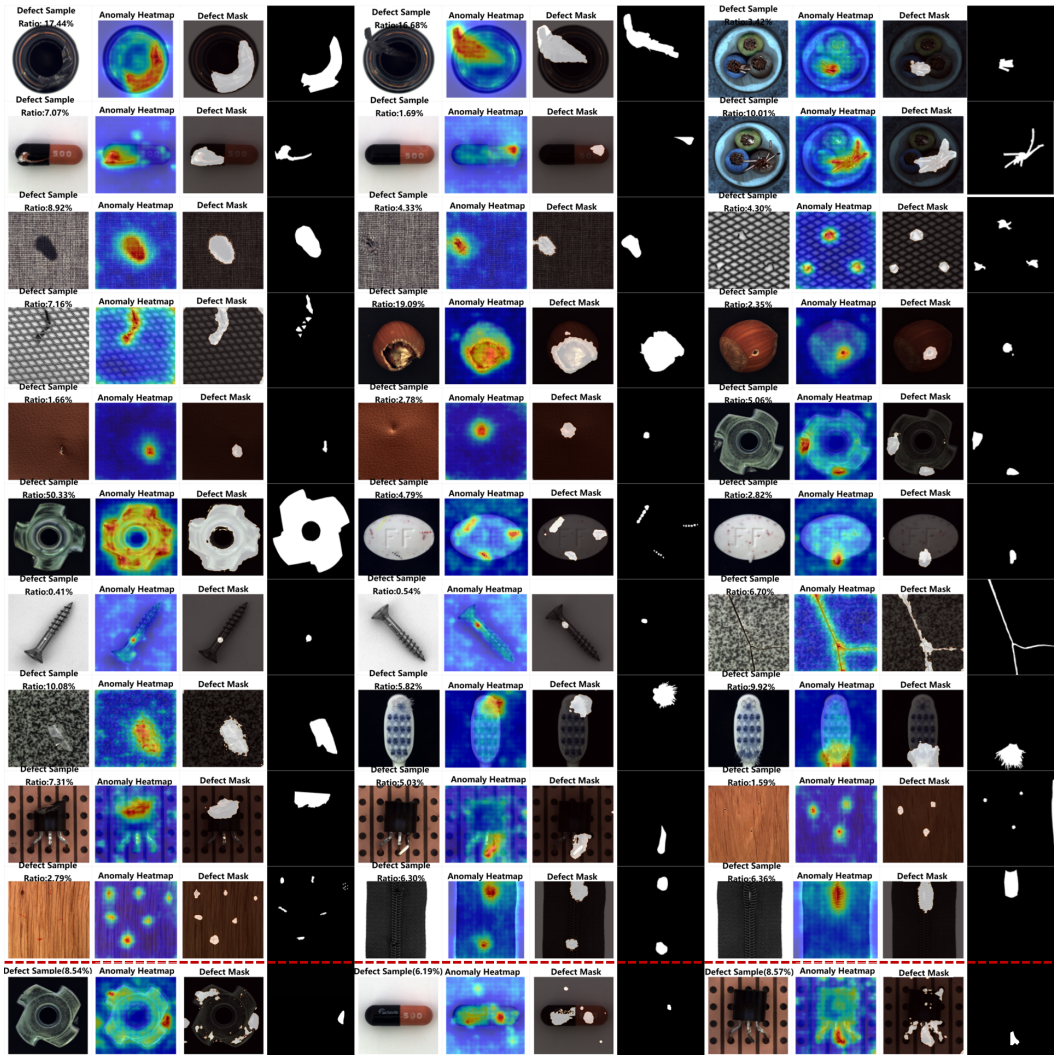


Figure 7. Qualitative localization results of FPC-LDR on MVTec AD. Each example shows the defective image, anomaly heatmap, predicted defect mask, and ground-truth mask. Most defects are accurately localized, while the cases below the red dashed line illustrate typical failures for small or low-contrast defects, including partial boundary omission and incomplete defect-region coverage.

Figure 8 illustrates the diffusion-based reconstruction process for normal samples. The model learns the distributional characteristics of normal samples through noise injection and denoising, and subsequently localizes defective regions in anomalous samples during inference.

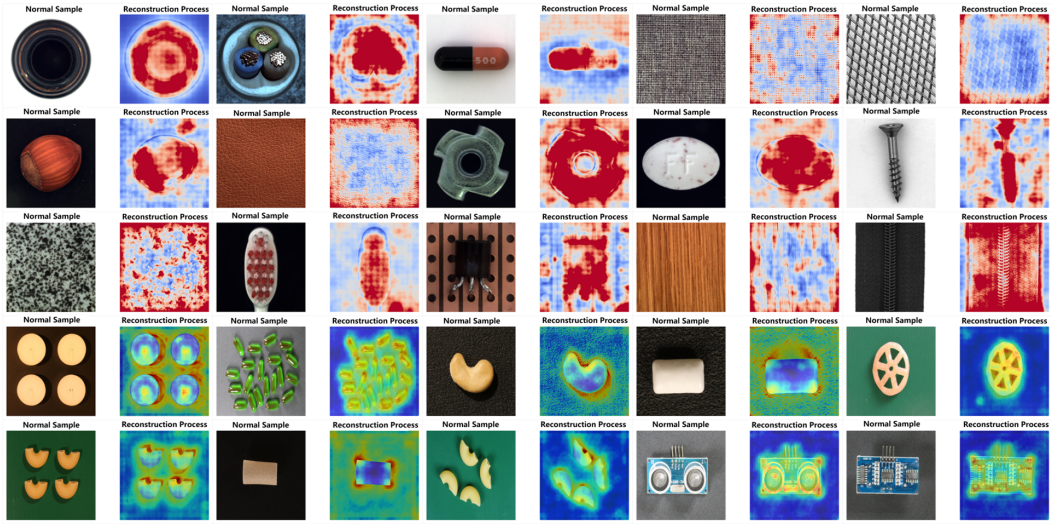


Figure 8. Diffusion-based reconstruction of normal samples. Red indicates reconstruction results on the MVTec AD dataset, while green indicates reconstruction results on the VisA dataset.

Figure 9 illustrates the anomaly-score distributions of FPC-LDR across the 15 MVTec AD categories. For each category, the upper panel shows the probability density distributions of image-level anomaly scores, where blue and orange represent normal and anomalous samples, respectively, while the lower panel presents the histogram of pixel-level anomaly scores. Normal samples are predominantly concentrated in the low-score range, whereas anomalous samples generally shift toward higher scores. The separation between the two distributions provides further evidence of the model's ability to discriminate between normal and defective samples. Figure 4b presents 3D anomaly heatmaps for the 15 categories in the MVTec AD dataset. Defective regions exhibit prominent high-score peaks, clearly illustrating the model's localization responses to different defect types.

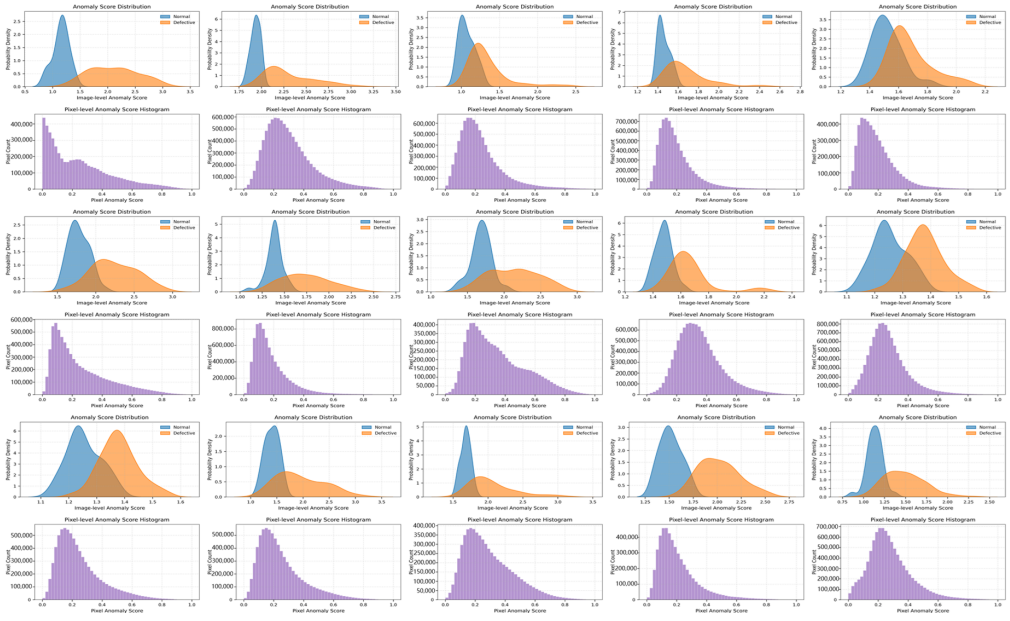


Figure 9. Visualization of anomaly score distributions produced by FPC-LDR.

4.4. Comparative Experiments

4.4.1. Comparative Experiments on the MVTec AD Dataset

To evaluate the overall detection performance of FPC-LDR and determine whether it effectively balances detection accuracy, localization precision, and inference efficiency, five

representative baselines were selected. PatchCore represents memory-bank-based feature matching; PaDiM and SPADE represent feature-distribution and correspondence modeling; AutoEncoder represents reconstruction-based anomaly detection; and Vi-Transformer provides a recent transformer-based comparison. All methods were evaluated on MVTec AD under the same unsupervised training and evaluation protocol. The results are reported in Table 3 and Figure 10.

Table 3. Performance comparison of different models for product defect detection on the MVTec AD dataset.

Model	Image-AUROC (%)	Pixel-AUROC (%)	PRO (%)	Threshold	Inference Time (ms)
PatchCore	90.41	93.50	92.56	4.9734	20.19
AutoEncoder	84.27	86.23	87.64	0.4221	7.83
PaDiM	85.17	88.91	90.91	0.9494	8.31
SPADE	85.07	89.02	90.02	0.9449	8.20
Vi-Transformer	92.54	94.03	94.93	0.3751	93.65
FPC -LDR (ours)	94.65	94.54	95.28	1.5319	26.14

This table compares the comprehensive detection performance of different unsupervised anomaly detection methods on the MVTec AD dataset.

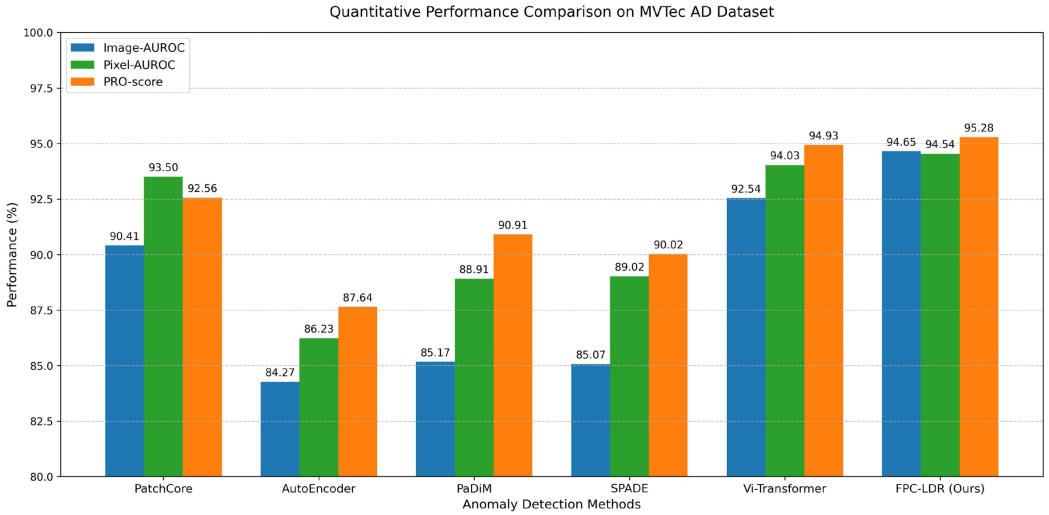


Figure 10. Comparison of the core performance metrics of different anomaly detection methods on the MVTec AD dataset.

Three principal conclusions can be drawn from the experimental results.

First, FPC-LDR achieved the best overall detection performance among the evaluated methods, with Image-AUROC, Pixel-AUROC, and PRO scores of 94.65%, 94.54%, and 95.28%, respectively. Compared with the transformer-based Vi-Transformer, FPC-LDR improves these three metrics by 2.11, 0.51, and 0.35 percentage points, respectively. This result shows that the proposed dual-branch framework remains competitive not only against conventional memory-, distribution-, and reconstruction-based methods, but also against a more recent transformer-based baseline.

Second, the comparison reveals different accuracy–efficiency characteristics across technical paradigms. AutoEncoder shows limited image-level and localization accuracy, while PaDiM and SPADE exhibit imbalanced performance across the three metrics. PatchCore provides efficient inference but remains weaker in overall detection and region-level localization. Vi-Transformer achieves strong localization performance, reaching 94.03% Pixel-AUROC and 94.93% PRO, but requires 93.65 ms per image. In contrast, FPC-LDR

achieves higher values on all three accuracy metrics while reducing inference latency to 26.14 ms. This indicates that the complementary integration of feature matching and lightweight diffusion reconstruction provides a more favorable balance between detection performance and computational efficiency.

Third, FPC-LDR maintained practical inference efficiency. Benefiting from the lightweight single-step diffusion reconstruction strategy, FPC-LDR achieved an average inference latency of 26.14 ms per image, corresponding to approximately 38.3 frames per second (FPS) under the experimental hardware configuration. This substantially alleviates the inference overhead associated with conventional multi-step diffusion models. Although FPC-LDR introduces a moderate increase in latency compared with lightweight baselines, the additional computational cost is accompanied by substantial improvements in detection and localization performance. Notably, FPC-LDR is also substantially faster than Vi-Transformer (26.14 ms vs. 93.65 ms per image), further demonstrating that the improved detection performance is not achieved at the expense of excessive computational cost. These results demonstrate a favorable trade-off between detection performance and computational efficiency and indicate the potential of FPC-LDR for real-time industrial inspection.

Overall, the comparative results on MVTec AD demonstrate that FPC-LDR effectively balances detection accuracy, localization performance, and inference efficiency. Figure 10 provides a visual comparison between FPC-LDR and the baseline methods.

4.4.2. Comparative Experiments on the VisA Dataset

In practical manufacturing environments, inspection systems must handle numerous product categories with substantially different data distributions. Cross-dataset generalization is therefore critical to the practical value of industrial surface defect detection methods. However, it remains a common limitation of single-paradigm approaches. Memory-bank methods rely on local feature matching and may exhibit substantial variations in accuracy under changes in texture and illumination. Generative reconstruction methods may also lose anomaly discriminability when their reconstruction models overfit the training distribution. Consequently, neither paradigm alone can adequately satisfy the multi-category inspection requirements of flexible production lines.

To evaluate the cross-dataset generalization and robustness of FPC-LDR, comparative experiments were conducted on VisA, which contains more complex backgrounds and a wider range of defect patterns. All methods retained the model architectures and training hyperparameters established on MVTec AD, without VisA-specific model retraining or hyperparameter tuning. The operating thresholds were calibrated separately using the designated validation data. This protocol enables a fair evaluation of their inherent cross-dataset transferability. The results are reported in Table 4. Figure 11 presents the defect localization results on the VisA dataset. FPC-LDR accurately localizes complex and fine-grained surface defects, demonstrating strong generalization and cross-dataset transferability.

Under this fixed-model transfer setting, FPC-LDR achieved an image-level AUROC of 95.16%, ranking first among all compared methods and outperforming the second-best method, SPADE, by 2.12 percentage points. This result indicates that the anomaly-discrimination mechanism of FPC-LDR generalizes effectively beyond the specific distribution of the MVTec AD dataset. Within the dual-branch architecture, the multi-scale semantic features stored in the memory bank serve as a transferable reference for anomaly discrimination, while the reconstruction residuals generated by the lightweight diffusion branch provide complementary distributional evidence. Their combination mitigates the performance degradation commonly observed in single-paradigm methods under domain shifts.

Table 4. Performance comparison of different models for product defect detection on the VisA dataset.

Model	Image-AUROC (%)	Pixel-AUROC (%)	PRO (%)	Threshold	Inference Time (ms)
PatchCore	90.18	96.64	89.30	5.0788	43.63
AutoEncoder	86.40	88.61	92.11	0.4185	13.68
PaDiM	91.07	93.31	85.41	4.7300	64.80
SPADE	93.04	89.30	94.50	0.1143	16.55
Vi-Transformer	93.25	92.49	93.45	0.3398	89.66
FPC -LDR (ours)	95.16	93.99	91.57	0.9487	69.00

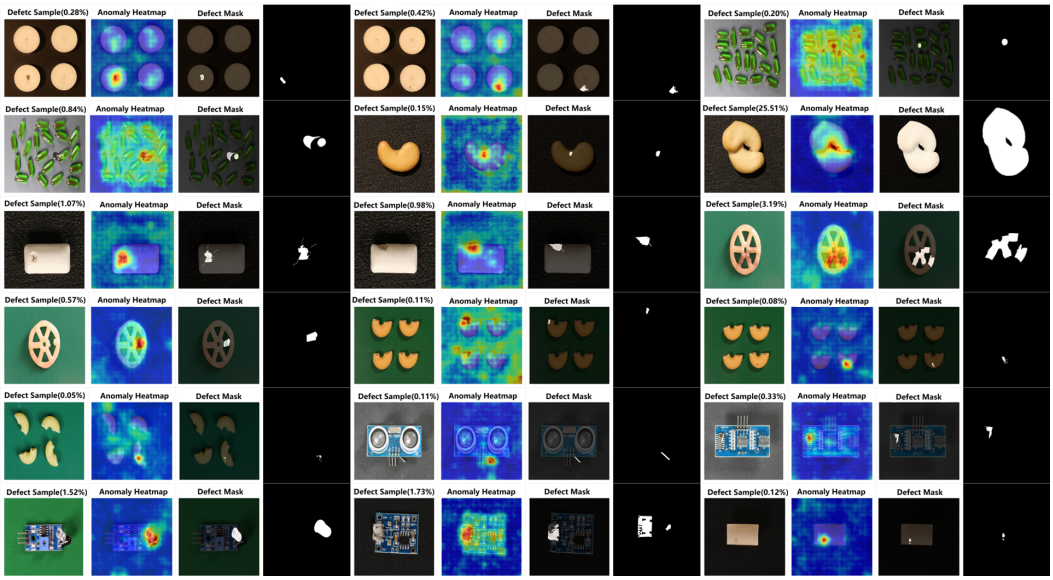


Figure 11. Defect localization results of FPC-LDR on the VisA dataset. The highlighted regions indicate defect locations, and the values in parentheses represent the defect ratios.

For defect localization, FPC-LDR achieved a pixel-level AUROC of 93.99% and a PRO score of 91.57%, placing it among the leading methods for both metrics and demonstrating balanced localization performance. By comparison, PatchCore obtained a high pixel-level AUROC of 96.64% but a lower PRO score of 89.30%, indicating accurate pixel discrimination but incomplete coverage of entire defect regions. Conversely, SPADE achieved a PRO score of 94.50% but only 89.30% pixel-level AUROC, suggesting strong region-level coverage but relatively weak boundary localization. Compared with Vi-Transformer, FPC-LDR improves Image-AUROC and Pixel-AUROC by 1.91 and 1.50 percentage points, respectively, while reducing inference latency from 89.66 ms to 69.00 ms. Although its PRO score is 1.88 percentage points lower, the results indicate a more favorable overall balance between image-level discrimination, pixel-level localization, and computational efficiency under cross-dataset transfer. These results reveal the difficulty faced by single-paradigm methods in simultaneously achieving precise boundaries and complete regional coverage. The performance metrics of the proposed method on the VisA dataset are presented in Table 5.

The category-wise results further explain the gap between image-level detection and pixel-level localization. Lower Pixel-AUROC is mainly observed for fryum and pcb4, while relatively low PRO scores occur for pcb3, macaroni2, pipe_fryum, and pcb2. These categories contain complex structures or irregular defect regions, making precise boundary discrimination and complete defect-region coverage more difficult. In contrast, their generally high Image-AUROC values indicate that FPC-LDR can still reliably identify anomalous images. Therefore, the performance gap mainly arises from fine-grained localization rather than image-level anomaly discrimination.

Table 5. Category-wise performance of FPC-LDR on the VisA dataset.

Class	Image-AUROC (%)	Pixel-AUROC (%)	PRO (%)	Threshold	Inference Time (ms)
candle	96.58	93.85	90.99	0.8416	104.65
capsules	95.64	91.79	97.67	1.0636	62.94
cashew	94.38	91.78	92.69	0.8964	68.36
chewinggum	95.50	95.26	98.53	0.8960	46.66
fryum	92.36	86.36	94.07	0.9916	46.07
macaroni1	95.42	95.53	89.12	1.0130	86.91
macaroni2	95.44	97.55	87.21	0.9477	85.43
pcb1	92.21	97.74	90.86	0.9720	86.76
pcb2	95.53	96.09	88.34	0.9079	96.98
pcb3	96.95	96.42	83.93	0.9950	81.90
pcb4	95.24	91.20	97.10	1.0046	19.62
pipe_fryum	96.67	94.36	88.28	0.8551	41.68
Average	95.16	93.99	91.57	0.9487	69.00

By adaptively integrating boundary-sensitive feature-matching scores with region-sensitive reconstruction residuals, the proposed dual-branch framework maintains a consistently balanced performance across two substantially different industrial datasets. These findings demonstrate the robustness and cross-dataset transferability of FPC-LDR and indicate its potential for rapid deployment in previously unseen industrial inspection scenarios.

4.4.3. Sensitivity Analysis of Memory-Bank Size

Table 6 evaluates the influence of memory-bank size on detection performance and computational cost.

Table 6. Sensitivity analysis of memory-bank size in FPC-LDR.

Memory Bank Size	Configuration Label	Image-Level AUROC (%)	Pixel-Level AUROC (%)	PRO (%)	Total Inference Time (ms)	GPU Memory Usage (MB)	Memory Increment (MB)
20,000	20 K	94.43	93.35	93.49	737.75	2700.21	235.68
50,000	50 K	94.57	95.98	93.74	1432.69	2967.02	366.35
100,000	100 K (Ours)	94.65	94.54	95.28	2664.09	2967.02	335.48
200,000	200 K	94.57	94.51	94.54	4484.52	3233.07	1184.54
300,000	300 K	96.98	96.68	93.03	7028.25	3233.07	1577.73

Increasing the bank size from 20 K to 100 K generally improves the overall performance, with the 100 K setting achieving 94.65% Image-AUROC, 94.54% Pixel-AUROC, and the highest PRO score of 95.28%. Further increasing the bank size to 200 K and 300 K substantially increases total inference time and memory consumption, while the localization performance does not improve consistently. Therefore, 100 K feature vectors were selected as a practical trade-off between detection accuracy, region-level localization, inference efficiency, and memory usage.

4.4.4. Comparison and Validation of Fusion Strategies

To verify the effectiveness of the proposed statistic-driven adaptive fusion method, two representative schemes—fixed-weight fusion and pixel-wise adaptive fusion—are selected for comparison (Table 7). Fixed-weight fusion uses uniform global weights with low computational cost, but it cannot adapt to category-dependent differences in anomaly-score distributions, leading to inferior overall detection performance. Pixel-wise adaptive

fusion dynamically adjusts weights at each pixel and can improve localization performance, but it is sensitive to local texture noise and introduces additional inference latency due to pixel-wise computation.

Table 7. Performance comparison of different fusion strategies.

Fusion Scheme	Image-AUROC (%)	Pixel-AUROC (%)	PRO (%)	Inference Time (ms)
Fixed Weight	92.29	93.89	92.18	32.41
Pixel-wise Adaptive	94.21	94.15	93.65	33.04
Statistical Adaptive (Ours)	94.65	94.54	95.28	26.14

The proposed statistic-driven adaptive fusion uses only the mean and standard deviation of the two anomaly maps as inputs and learns global branch-level fusion weights through a lightweight multilayer perceptron. Experimental results show that it achieves the best overall performance among the compared fusion schemes in Image-AUROC, Pixel-AUROC, and PRO. Meanwhile, its inference time is approximately 20% lower than that of the pixel-wise adaptive scheme, providing a better balance between detection accuracy and computational efficiency.

4.5. Ablation Studies

4.5.1. Ablation Study of the Core Functional Modules

To further evaluate the effectiveness of the dual-branch architecture, adaptive fusion mechanism, and image-level score optimization strategy, a progressive component-wise ablation study was performed, with the results summarized in Table 8.

Table 8. Ablation results of the core functional modules of FPC-LDR on the independent validation set.

Config	Memory	Diffusion	Adaptive Fusion	Score Optimization	Image-AUROC (%)	Pixel-AUROC (%)	PRO (%)	Inference-Time (ms)
A	✓	—	—	—	94.49 ($\Delta + 3.46$)	95.52 ($\Delta + 0.03$)	93.33 ($\Delta + 2.35$)	37.68 ($\Delta - 2.75$)
B	—	✓	—	—	96.08 ($\Delta + 1.87$)	95.40 ($\Delta + 0.15$)	95.06 ($\Delta + 0.62$)	75.78 ($\Delta - 40.85$)
C	✓	✓	Fixed	—	97.00 ($\Delta + 0.95$)	94.98 ($\Delta + 0.57$)	94.28 ($\Delta + 1.40$)	36.34 ($\Delta - 1.41$)
D	✓	✓	Adaptive	—	97.76 ($\Delta + 0.19$)	95.55 ($\Delta = 0$)	95.68 ($\Delta = 0$)	34.93 ($\Delta = 0$)
E	✓	✓	Adaptive	✓	97.95	95.55	95.68	34.93

These results are used for relative component comparison and are not directly comparable with the test-set results in Tables 2 and 3. Δ is calculated as the reference value minus current value relative to the best-performing configuration for each metric. AUROC/PRO differences are expressed in percentage points, and inference-time differences in milliseconds.

The component-wise ablation experiments were conducted using an independent validation protocol and were intended to quantify the relative contribution of individual modules. Therefore, the numerical results in Table 8 are not directly comparable with the final test-set results reported in Tables 2 and 3. Table 8 should be interpreted in terms of relative trends among configurations rather than absolute performance differences from Tables 2 and 3.

Configuration A, which employs only the memory-bank branch, achieves an Image-AUROC of 94.49% and a Pixel-AUROC of 95.52%, demonstrating effective local anomaly discrimination based on the learned normal feature distribution. However, its PRO score is limited to 93.33%, indicating insufficient coverage of spatially continuous or irregular defect regions. In contrast, Configuration B, which relies solely on the diffusion branch, improves the Image-AUROC and PRO score to 96.08% and 95.06%, respectively, confirming the stronger region-level modeling capability of generative reconstruction. Nevertheless, its

inference latency increases to 75.78 ms, indicating that the diffusion branch alone introduces a higher computational cost and is less favorable for efficient online industrial inspection.

In Configuration C, integrating the memory-bank and diffusion branches with fixed-weight fusion increases the Image-AUROC to 97.00%, confirming the complementary nature of discriminative feature matching and generative reconstruction. However, fixed-weight fusion cannot fully adapt to variations in product texture, defect scale, and local structural characteristics. Replacing fixed-weight fusion with adaptive fusion in Configuration D further improves the Image-AUROC, Pixel-AUROC, and PRO to 97.76%, 95.55%, and 95.68%, respectively, while reducing the inference latency from 36.34 to 34.93 ms. These results indicate that adaptive fusion enables more effective coordination of the complementary anomaly responses from the two branches without introducing measurable additional computational overhead.

Configuration E further incorporates the score-optimization strategy, increasing Image-AUROC from 97.76% to 97.95%, with an absolute gain of 0.19 percentage points. Pixel-AUROC and PRO remain unchanged at 95.55% and 95.68%, respectively, with the same inference latency of 34.93 ms. This result shows that score optimization mainly provides a modest refinement of image-level anomaly discrimination and does not affect pixel-level localization. Because repeated-seed significance testing was not performed, the gain is interpreted cautiously as an incremental improvement rather than a statistically significant effect. Overall, the ablation results mainly support the contributions of memory-based feature matching, diffusion reconstruction, and adaptive fusion, while score optimization serves as a lightweight complementary refinement.

The small inference-time fluctuations among configurations with unchanged major network components are attributed to normal runtime variations in GPU execution, including kernel scheduling, memory access, and synchronization, rather than structural differences in the model.

4.5.2. Ablation Study of Key Model Configurations

To quantify the contributions and validate the design of each core component in FPC-LDR, controlled ablation experiments were conducted on the MVTec AD dataset. Four aspects were investigated: the feature-extraction backbone, the diffusion noise schedule, reconstruction strategy, and the loss function. All experimental groups followed the same unsupervised training protocol, evaluation criteria, and hardware configuration. Inference latency was also reported to assess the trade-off between accuracy and efficiency. The complete results are presented in Table 9.

(1) Feature extraction backbone. ResNet-50 achieved the best balance between accuracy and efficiency, obtaining image-level and pixel-level AUROC values of 94.65% and 94.54%, respectively, with an inference time of 26.14 ms per image. Its deep semantic features reduced memory-bank redundancy after coreset sampling, while the resulting decrease in nearest-neighbor search overhead compensated for the additional backbone computation. Lightweight backbones such as MobileNetV2 exhibited noticeably weaker feature discrimination, whereas the custom ConvNet trained without pretraining showed limited generalization and relatively high inference latency.

(2) Diffusion noise schedule. The linear β-schedule was most compatible with the single-step diffusion framework, achieving the best image- and pixel-level AUROC. Its gradually varying noise intensity facilitated stable modeling of the normal-data distribution. Although the cosine and constant schedules produced slightly higher PRO scores, they substantially reduced image-level accuracy by increasing reconstruction errors in normal regions and, consequently, the false-positive rate. The quadratic schedule provided only

moderate accuracy while nearly doubling the inference latency, making it less favorable for real-time inspection applications.

(3) Reconstruction strategy. Diffusion-based reconstruction considerably outperformed direct reconstruction, improving image-level AUROC by 8.6 percentage points and pixel-level AUROC and PRO scores by 2.8 and 4.9 percentage points, respectively. Direct reconstruction primarily learns point-to-point appearance mapping and may generalize excessively to anomalous regions, thereby reducing anomaly separability. In contrast, the noise–denoising paradigm enables the diffusion model to capture the underlying distribution of normal samples and amplify reconstruction discrepancies in defective regions.

(4) Loss function. The combination of mean squared error and perceptual loss achieved the best overall performance. Individual losses exhibited imbalanced behavior: although MSE, L1, and perceptual loss yielded PRO scores of approximately 96.6%, their image-level AUROC values remained around 87.4%, indicating elevated reconstruction errors in normal regions and increased false positives. The combined objective preserved pixel-level details while maintaining semantic structural consistency, reducing the normal-region error baseline and improving image-level AUROC by 7.2 percentage points. This balance is better suited to comprehensive industrial quality inspection.

Table 9. Ablation results of key model configurations in FPC-LDR.

Component Type	Configuration	Image-AUROC (%)	Pixel-AUROC (%)	PRO (%)	Threshold	Inference Time (ms)
FeatureExtractor	ResNet18	87.47 (Δ + 7.18)	90.14 (Δ + 4.40)	94.29 (Δ + 0.99)	0.9851	68.59 (Δ − 42.45)
	ResNet34	88.44 (Δ + 6.21)	92.33 (Δ + 2.21)	97.65 (Δ − 2.37)	1.1883	31.61 (Δ − 5.47)
	ResNet50	**94.65**	**94.54**	**95.28**	**1.5319**	**26.14**
	MobileNetV2	85.70 (Δ + 8.95)	89.30 (Δ + 5.24)	93.60 (Δ + 1.68)	0.7827	38.07 (Δ − 11.93)
	ConvNet	91.85 (Δ + 2.80)	91.33 (Δ + 3.21)	92.46 (Δ + 2.82)	1.3153	90.20 (Δ − 64.06)
Diffusion Noise Schedule	**Linear β**	**94.65**	**94.54**	**95.28**	**1.5319**	**26.14**
	Cosine β	87.41 (Δ + 7.24)	92.01 (Δ + 2.53)	96.66 (Δ − 1.38)	1.1677	27.99 (Δ − 1.85)
	Quadratic β	89.89 (Δ + 4.76)	91.13 (Δ + 3.41)	93.93 (Δ + 1.35)	1.21	51.80 (Δ − 25.66)
	Constant β	87.42 (Δ + 7.23)	91.99 (Δ + 2.55)	96.67 (Δ − 1.39)	1.16	24.89 (Δ + 1.25)
Reconstruction Strategy	Direct Reconstruction	86.05 (Δ + 8.60)	91.74 (Δ + 2.80)	90.38 (Δ + 4.90)	2.06	23.22 (Δ + 2.92)
	Diffusion Reconstruction	**94.65**	**94.54**	**95.28**	**1.5319**	**26.14**
	Multi-scale Reconstruction	85.96 (Δ + 8.69)	88.61 (Δ + 5.93)	95.19 (Δ + 0.09)	1.0076	22.97 (Δ + 3.14)
Loss Function	MSE	87.41 (Δ + 7.24)	91.98 (Δ + 2.56)	96.67 (Δ − 1.39)	1.1664	21.80 (Δ + 4.34)
	L1 Loss	87.42 (Δ + 7.23)	91.98 (Δ + 2.56)	96.65 (Δ − 1.37)	1.1664	26.00 (Δ + 0.14)
	Perceptual Loss	87.42 (Δ + 7.23)	91.99 (Δ + 2.55)	96.66 (Δ − 1.38)	1.1665	24.35 (Δ + 1.79)
	MSE + Perceptual	**94.65**	**94.54**	**95.28**	**1.5319**	**26.14**

Δ is calculated as the reference value minus the current value within each component group (ResNet50, Linear β, Diffusion Reconstruction, and MSE + Perceptual). AUROC/PRO differences are expressed in percentage points and inference-time differences in milliseconds. The boldfaced components indicate the configuration adopted in the proposed method.

5. Discussion

Experimental results demonstrate that FPC-LDR achieves strong performance across the principal metrics on MVTec AD and maintains leading image-level classification accuracy in cross-dataset evaluation on VisA, confirming its detection performance, generalization capability, and robustness. For industrial defect inspection, its unsupervised training

paradigm eliminates the need for annotated defect samples and can accommodate a range of surface anomalies, including scratches, mildew, damage, and foreign materials. This method therefore has the potential to reduce missed detections in manual inspection, improve production-line efficiency, and prevent defective products from entering downstream processes, thereby providing reliable technical support for industrial quality control [56].

The principal advantages of FPC-LDR are threefold. First, its complementary dual-branch architecture integrates memory-bank-based discrimination with diffusion-based reconstruction, improving both image-level anomaly classification and defect-region coverage while mitigating the limitations of each paradigm. Second, its lightweight single-step diffusion reconstruction mechanism substantially reduces the inference latency of conventional diffusion models, supporting potential real-time industrial inspection applications while maintaining competitive accuracy. Third, its adaptive dynamic fusion mechanism requires no manual weight adjustment for new scenarios, supporting stable cross-dataset performance and rapid adaptation to products with different categories and textures.

Nevertheless, this study has two limitations. First, FPC-LDR incurs higher inference latency than lightweight methods such as PaDiM, and further optimization is required for deployment on highly resource-constrained edge devices. Second, the localization accuracy for extremely small and low-contrast defects can still be improved, as illustrated by the representative failure cases below the red dashed line in Figure 7. In real production lines, gradual changes in illumination, material batches, equipment conditions, or process parameters may also shift the normal-data distribution, potentially increase false alarms or missed detections over time.

Future work will therefore focus on three directions: applying model quantization and pruning for embedded edge deployment; improving multi-scale feature enhancement to strengthen the detection and localization of subtle defects; and introducing incremental learning to accommodate dynamic shifts in the distribution of normal production data, thereby extending the method to a broader range of industrial inspection scenarios.

6. Conclusions

To address the scarcity of annotated defect samples and the challenge of balancing detection accuracy with inference efficiency, this study proposes FPC-LDR, an unsupervised dual-branch framework that integrates memory-bank feature matching with lightweight diffusion reconstruction. By exploiting the complementary strengths of discriminative feature matching and generative reconstruction, FPC-LDR enables robust image-level anomaly detection and pixel-level defect localization.

Experiments on MVTec AD demonstrate that FPC-LDR outperforms representative methods, including PatchCore, PaDiM, and SPADE, for image-level AUROC, pixel-level AUROC, and PRO score. Ablation studies further confirm the effectiveness of the main model components. Cross-dataset evaluation on VisA shows that FPC-LDR achieves the highest image-level AUROC among the compared methods while maintaining competitive pixel-level localization performance. In addition, the proposed method achieves a favorable balance between detection performance and computational efficiency, indicating its potential for real-time industrial inspection.

Overall, FPC-LDR provides an effective unsupervised solution for industrial surface defect detection. With an average inference latency of 26.14 ms per image (approximately 38.3 FPS) under the experimental hardware configuration, the proposed method achieves a favorable balance among detection accuracy, localization performance, and computational efficiency, demonstrating its potential for real-time automated quality inspection in intelligent manufacturing.

Author Contributions: Conceptualization, X.T. and J.H.; methodology, X.T., J.H. and J.S.; software, X.T. and X.Z.; validation, X.T., J.H., Y.T., J.S. and X.Z.; formal analysis, X.T., Y.T. and J.S.; investigation, X.T., Y.T. and X.Z.; resources, J.H. and J.S.; data curation, X.T. and Y.T.; writing—original draft preparation, X.T.; writing—review and editing, X.T., J.H., Y.T., J.S. and X.Z.; visualization, X.T., Y.T. and X.Z.; supervision, J.H.; project administration, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this study are publicly available. The source code, pre-trained model weights, and experimental output logs associated with this work will be publicly released upon acceptance of the manuscript. The corresponding repository link will be provided in the final published version.

Acknowledgments: Large language models were used for English language editing and polishing of the manuscript. Specifically, ChatGPT (GPT-6, Medium thinking mode) was used to improve language clarity, grammar, and readability.The final manuscript was reviewed and approved by all authors.

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

References

1. Liu, J.; Xie, G.; Wang, J.; Li, S.; Wang, C.; Zheng, F.; Jin, Y. Deep industrial image anomaly detection: A survey. *Mach. Intell. Res.* **2024**, *21*, 104–135. [CrossRef]
2. Tang, B.; Chen, L.; Sun, W.; Lin, Z. Review of surface defect detection of steel products based on machine vision. *IET Image Process.* **2023**, *17*, 303–322. [CrossRef]
3. Ma, Y.; Yin, J.; Huang, F.; Li, Q. Surface defect inspection of industrial products with object detection deep networks: A systematic review. *Artif. Intell. Rev.* **2024**, *57*, 333. [CrossRef]
4. Liu, M.; Zhang, M.; Chen, X.; Zheng, C.; Wang, H. YOLOv8-LMG: An improved bearing defect detection algorithm based on YOLOv8. *Processes* **2024**, *12*, 930. [CrossRef]
5. Cheng, Y.; Cao, Y.; Yao, H.; Luo, W.; Jiang, C.; Zhang, H.; Shen, W. A comprehensive survey for real-world industrial surface defect detection: Challenges, approaches, and prospects. *J. Manuf. Syst.* **2026**, *84*, 152–172. [CrossRef]
6. Huang, H.; Wang, P.; Pei, J.; Wang, J.; Alexanian, S.; Niyato, D. Deep learning advancements in anomaly detection: A comprehensive survey. *IEEE Internet Things J.* **2025**, *12*, 44318–44342. [CrossRef]
7. Chen, Y.; Zhang, R. Hybrid dual-channel attention CNN and eXtreme Gradient Boosting for industrial process model development and fault diagnosis. *IEEE Internet Things J.* **2025**, *12*, 35649–35661. [CrossRef]
8. Hung, Y.-H. Developing an anomaly detection system for automatic defective products' inspection. *Processes* **2022**, *10*, 1476. [CrossRef]
9. Peng, T.; Zheng, Y.; Zhao, L.; Zheng, E. Industrial product surface anomaly detection with realistic synthetic anomalies based on defect map prediction. *Sensors* **2024**, *24*, 264. [CrossRef] [PubMed]
10. Xu, H.; Xu, S.; Yang, W. Unsupervised industrial anomaly detection with diffusion models. *J. Vis. Commun. Image Represent.* **2023**, *97*, 103983. [CrossRef]
11. Bergmann, P.; Batzner, K.; Fauser, M.; Sattlegger, D.; Steger, C. The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *Int. J. Comput. Vis.* **2021**, *129*, 1038–1059. [CrossRef]
12. Klarák, J.; Andok, R.; Malík, P.; Kuric, I.; Ritomský, M.; Klačková, I.; Tsai, H.-Y. From anomaly detection to defect classification. *Sensors* **2024**, *24*, 429. [CrossRef] [PubMed]
13. Zhang, D.; Zhou, S.; Zheng, Y.; Xu, X. Review on application of machine vision-based intelligent algorithms in gear defect detection. *Processes* **2025**, *13*, 3370. [CrossRef]
14. Hyun, J.; Kim, S.; Jeon, G.; Kim, S.H.; Bae, K.; Kang, B.J. ReConPatch: Contrastive patch representation learning for industrial anomaly detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2024; pp. 2052–2061. [CrossRef]
15. Qu, Z.; Tao, X.; Gong, X.; Qu, S.; Chen, Q.; Zhang, Z.; Wang, X.; Ding, G. Bayesian prompt flow learning for zero-shot anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 30398–30408. [CrossRef]

16. Zhou, Q.; Pang, G.; Tian, Y.; He, S.; Chen, J. AnomalyCLIP: Object-agnostic prompt learning for zero-shot anomaly detection. In Proceedings of the International Conference on Learning Representations (ICLR), Vienna, Austria, 7–11 May 2024.
17. Zhang, T.; Gao, L.; Li, X.; Gao, Y. Diffusion-based zero-shot anomaly detection. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 10131–10138. [CrossRef]
18. Beizaee, F.; Lodygensky, G.A.; Desrosiers, C.; Dolz, J. Correcting deviations from normality: A reformulated diffusion model for multi-class unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 19088–19097. [CrossRef]
19. Jeong, J.; Zou, Y.; Kim, T.; Zhang, D.; Ravichandran, A.; Dabeer, O. WinCLIP: Zero-/few-shot anomaly classification and segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 19606–19616. [CrossRef]
20. 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 (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 14318–14328. [CrossRef]
21. Ishida, K.; Takena, Y.; Nota, Y.; Mochizuki, R.; Matsumura, I.; Ohashi, G. SA-PatchCore: Anomaly detection in dataset with co-occurrence relationships using self-attention. *IEEE Access* **2023**, *11*, 3232–3240. [CrossRef]
22. Kamiya, S.; Yamashita, K.; Hotta, K. ASO-PatchCore: Memory-efficient and fast anomaly detection via automatic sampling optimization. In *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP)*; SciTePress: Setúbal, Portugal, 2026; pp. 423–430. [CrossRef]
23. Wu, S.; Wang, Y.; Liu, X.; Yang, Y.; Wang, R.; Guo, G.; Doermann, D.; Zhang, B. DFM: Differentiable feature matching for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 15224–15233. [CrossRef]
24. Zheng, Z.; Yang, H.; Yin, Z.P. A hierarchical patch feature distribution network for industrial multiscale defect detection. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 19709–19722. [CrossRef]
25. Liu, C.; Huang, Y.; Zhao, Z.; Geng, W.; Luo, T. LGM-YOLO: A context-aware multi-scale YOLO-based network for automated structural defect detection. *Processes* **2025**, *13*, 2411. [CrossRef]
26. Guo, J.; Lu, S.; Zhang, W.; Chen, F.; Li, H.; Liao, H. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 20405–20415.
27. Jiang, Z.; Zhang, Y.; Wang, Y.; Li, J.; Gao, X. FR-PatchCore: An industrial anomaly detection method for improving generalization. *Sensors* **2024**, *24*, 1368. [CrossRef] [PubMed]
28. Fan, S.K.S.; Lai, L.C. A novel diffusion model-based deep learning approach for anomaly detection in computer connector image analysis. *Eng. Appl. Artif. Intell.* **2026**, *166*, 113604. [CrossRef]
29. Chen, L.; Zheng, J.; Xu, F.; Zhu, F.; Abraham, A.; Zeng, H. Enhancing few-shot marble slab surface defect detection: A diffusion framework with knowledge distillation and semantic guidance. *Eng. Appl. Artif. Intell.* **2026**, *172*, 114334. [CrossRef]
30. Sun, C.; Gu, P.; Wan, H.; Ren, M.; Ma, Y. Enhancing 3C product surface defect detection through historical data retrieval and diffusion model-based synthetic data augmentation. *Comput. Ind. Eng.* **2026**, *216*, 111931. [CrossRef]
31. Cai, M.; Zhang, Z.; Wu, G.; Chai, T. XMatchAD: A cross-modal matching perspective on reconstruction-based anomaly detection. *IEEE Trans. Image Process.* **2026**, *35*, 8443–8455. [CrossRef] [PubMed]
32. Li, Q.; Chang, X.; Jiang, Q. FEAD: Frequency-enhanced cross-modal prompt learning for zero-shot anomaly detection. *Pattern Recognit.* **2026**, *180*, 114356. [CrossRef]
33. Shao, Y.; Wang, L.; Li, C.; Chen, P.; Liu, Q. PromptMoE: Generalizable zero-shot anomaly detection via visually-guided prompt mixtures. *Proc. AAAI Conf. Artif. Intell.* **2026**, *40*, 8878–8886. [CrossRef]
34. Defard, T.; Setkov, A.; Loesch, A.; Audigier, R. PaDiM: A patch distribution modeling framework for anomaly detection and localization. In *Pattern Recognition. ICPR International Workshops and Challenges*; Springer: Cham, Switzerland, 2021; pp. 475–489. [CrossRef]
35. Cohen, N.; Hoshen, Y. Sub-image anomaly detection with deep pyramid correspondences. *arXiv* **2020**, arXiv:2005.02357. [CrossRef]
36. Liu, J.; Ma, Z.; Wang, Z.; Zou, C.; Ren, J.; Wang, Z.; Song, L.; Hu, B.; Liu, Y.; Leung, V.C.M. A survey on diffusion models for anomaly detection. *arXiv* **2025**, arXiv:2501.11430. [CrossRef]
37. 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 (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 8330–8339. [CrossRef]
38. Zhang, H.; Wang, Z.; Zeng, D.; Wu, Z.; Jiang, Y.-G. DiffusionAD: Norm-guided one-step denoising diffusion for anomaly detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 7140–7152. [CrossRef] [PubMed]

39. Feng, Y.; Chen, W.; Li, Y.; Chen, B.; Wang, Y.; Zhao, Z.; Liu, H.; Zhou, M. OmiAD: One-step adaptive masked diffusion model for multi-class anomaly detection via adversarial distillation. *Proc. Mach. Learn. Res.* **2025**, *267*, 16604–16633.
40. Akshay, S.; Narasimhan, N.L.; George, J.; Balasubramanian, V.N. A unified latent Schrödinger bridge diffusion model for unsupervised anomaly detection and localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 25528–25538. [CrossRef]
41. 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 (WACV), Waikoloa, HI, USA, 3–8 January 2024; pp. 128–138. [CrossRef]
42. 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 (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 9737–9746. [CrossRef]
43. Wei, C.; Han, H.; Xia, Y.; Ji, Z. TDAD: Self-supervised industrial anomaly detection with a two-stage diffusion model. *Comput. Ind.* **2025**, *164*, 104192. [CrossRef]
44. An, C.; Wu, J.; Zhang, H. F-ADMD: Fast anomaly detection via multi-scale guided denoising diffusion model. *Neurocomputing* **2025**, *654*, 131305. [CrossRef]
45. Li, H.; Chen, Z.; Xu, Y.; Hu, J. Hyperbolic anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 17511–17520. [CrossRef]
46. Chen, Y.; Ding, Y.; Zhao, F.; Zhang, E.; Wu, Z.; Shao, L. Surface defect detection methods for industrial products: A review. *Appl. Sci.* **2021**, *11*, 7657. [CrossRef]
47. Qu, X.; Liu, Z.; Wu, C.Q.; Hou, A.; Yin, X.; Chen, Z. MFGAN: Multimodal fusion for industrial anomaly detection using attention-based autoencoder and generative adversarial network. *Sensors* **2024**, *24*, 637. [CrossRef] [PubMed]
48. Cheng, H.; Luo, J.; Zhang, X. Multimodal industrial anomaly detection via uni-modal and cross-modal fusion. *IEEE Trans. Ind. Inform.* **2025**, *21*, 5000–5010. [CrossRef]
49. Costanzino, A.; Zama Ramirez, P.; Lisanti, G.; Di Stefano, L. Multimodal industrial anomaly detection by crossmodal feature mapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 17234–17243. [CrossRef]
50. Wu, G.; Zhang, Y.; Deng, L.; Zhang, J.; Chai, T. Cross-modal learning for anomaly detection in complex industrial process: Methodology and benchmark. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 2632–2645. [CrossRef]
51. Xie, G.; Wang, J.; Liu, J.; Lyu, J.; Liu, Y.; Wang, C.; Zheng, F.; Jin, Y. IM-IAD: Industrial image anomaly detection benchmark in manufacturing. *IEEE Trans. Cybern.* **2024**, *54*, 2720–2733. [CrossRef] [PubMed]
52. 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 (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 20402–20411. [CrossRef]
53. Xu, R.; Chiu, Y.-T.; Chen, T.-I.; Chew, O.; Chuang, Y.-Y.; Cheng, W.-H. Training-free industrial defect generation with diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–23 October 2025; pp. 24214–24223. [CrossRef]
54. Song, J.; Park, D.; Baek, K.; Lee, S.; Choi, J.; Kim, E.; Yoon, S. DefectFill: Realistic defect generation with inpainting diffusion model for visual inspection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 18718–18727. [CrossRef]
55. Zou, Y.; Jeong, J.; Pemula, L.; Zhang, D.; Dabeer, O. SPot-the-Difference self-supervised pre-training for anomaly detection and segmentation. In *Computer Vision—ECCV 2022 Workshops*; Springer: Cham, Switzerland, 2022; pp. 392–408. [CrossRef]
56. Im, D.; Lee, S.; Lee, H.; Yoon, B.; So, F.; Jeong, J. A data-centric approach to design and analysis of a surface-inspection system based on deep learning in the plastic injection molding industry. *Processes* **2021**, *9*, 1895. [CrossRef]

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.