

Article

Regularized Latent Adaptive Framework for Unsupervised Industrial Anomaly Detection via Multi-Scale Generative–Discriminative Learning

Leqi Chi ¹, Tao Ma ² , Yuhang Lang ^{1,†}, Xinran Lv ^{3,†}, Xingfan Li ^{1,†} and Xiaoguang Li ^{1,4,*}

¹ School of Electronic Information Engineering, Changchun University, Changchun 130000, China; 240402207@mails.ccu.edu.cn (L.C.); 17835209338@163.com (Y.L.); 240402203@mails.ccu.edu.cn (X.L.)

² School of Economics and Management, Harbin University of Science and Technology, Harbin 150000, China; 2420710211@stu.hrbust.edu.cn

³ School of Mechanical and Power Engineering, Harbin University of Science and Technology, Harbin 150000, China; lxxxr0705@163.com

⁴ School of Electrical and Information Engineering, Changchun Guanghua University, Changchun 130000, China

* Correspondence: lixiaoguang@ghu.edu.cn

† These authors contributed equally to this work.

Abstract

Industrial visual inspection relies heavily on unsupervised anomaly detection due to the scarcity of annotated defect samples. However, existing methods struggle to balance global structural consistency and local defect sensitivity, leading to limited accuracy in practical scenarios. To address this challenge, we propose a unified generative–discriminative framework that combines regularized latent space encoding with multi-scale discriminator-guided supervision. Specifically, a dynamic compactness regularization strategy constrains latent representations of normal samples into a compact manifold to suppress anomaly reconstruction, while a multi-scale discriminator provides hierarchical perceptual constraints to enhance fine-grained anomaly localization across different spatial resolutions. Here we show that the proposed method achieves 98.6% image-level and 98.4% pixel-level AUROC on the MVTec AD benchmark, outperforming state-of-the-art approaches. This framework provides a stable and effective solution for real-world industrial quality inspection.

Keywords: unsupervised learning; industrial anomaly detection; generative adversarial networks; attention mechanism; latent space distribution; feature fusion

1. Introduction

Currently, automated surface defect detection with machine vision has become a crucial part of modern industrial quality control, replacing subjective and inefficient manual inspections in key manufacturing sectors. Its applications are extensive, covering areas from semiconductor fabrication and aerospace engineering to automotive manufacturing. However, most systems rely on observing many examples of defects that are precisely labeled. That is a problem [1]. Real defects are rare, varied, and unpredictable. Generating enough labeled data is impossible. This causes the systems to fail to identify new or unusual defects. This limitation necessitates a shift to alternative learning frameworks [2]. In recent years, deep learning-based anomaly detection methods have become popular in various industrial contexts [3]. Despite these advances, achieving robust and generalizable detection remains challenging, particularly when the variability and fine-grained nature of real defects must be captured under strict data constraints [4–6].



Academic Editor: Roberto Teti

Received: 27 May 2026

Revised: 23 June 2026

Accepted: 24 June 2026

Published: 1 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.

Historically, the direct application of deep learning to this domain has been significantly constrained by the scarcity of available anomaly samples in real-world industrial settings [7,8]. In stark contrast to large-scale datasets [9] such as ImageNet, which contains over 14 million images, surface defect detection is often confronted with the small sample challenge, where datasets may comprise merely a few dozen exemplars [10–12]. Numerous unsupervised domain adaptation strategies have been proposed to mitigate this issue by leveraging unlabeled target-domain data to improve generalization [13,14]. Yet, domain adaptation alone cannot fully address the intrinsic modeling gap: capturing the underlying manifold of normal imagery with sufficient fidelity to reliably reveal subtle deviations [15–17]. Motivated by this challenge, researchers explore a series of feature-level solutions that leverage the semantic richness of generative latent spaces to construct a precise and expressive representation of normalcy [18–20]. These explorations further promote the application of generative models in unsupervised anomaly detection.

Overall, unsupervised anomaly detection methods are mainly divided into two categories. Including methods based on reconstruction and feature embedding. While generative reconstruction models intuitively identify anomalies via pixel-level discrepancies, their core limitation is the well-known phenomenon of “anomaly reconstruction”—powerful generators often learn to reproduce anomalous regions faithfully, thereby severely diminishing detection sensitivity. Conversely, feature embedding-based methods offer superior local discriminability but heavily rely on fixed, task-agnostic external priors, which often suffer from domain gaps when applied to specific industrial textures. Therefore, existing methods still struggle to effectively reconcile the inherent contradiction between global structural consistency and local defect sensitivity. A critical challenge remains unresolved: how to organically integrate the global generative consistency of a task-specific reconstruction model with the local discriminative strength of a learned multi-scale feature representation, while simultaneously avoiding reliance on externally pretrained priors.

To address these fundamental challenges, this paper proposes a novel unsupervised anomaly detection framework that leverages StyleGAN2-ADA [21] as a generative backbone while introducing regularized latent representation learning and multi-scale discriminative supervision for industrial anomaly detection. The core motivation of the proposed framework is to reconcile the trade-off between global structural consistency and local defect sensitivity through a coordinated multi-component design. First, a coordinate attention module (CAM) is incorporated into the encoder to enhance spatial feature representation and preserve long-range structural dependencies. Based on these enriched representations, the proposed regularized latent space (RLS) constrains normal samples to form a latent manifold, thereby suppressing anomaly reconstruction and improving latent space separability. To further balance compactness and reconstruction fidelity, a PI-based adaptive controller dynamically adjusts the regularization strength during training. Second, to overcome the limited local sensitivity of conventional discriminators, the multi-scale discriminator (MSD) provides patch-level adversarial supervision at multiple spatial resolutions, enabling enhanced perception of subtle defects. During inference, the anomaly score is computed as the Mahalanobis distance between a test feature and the learned normal distribution for that scale and location. These multi-scale anomaly maps are then fused to produce a final, high-resolution localization result. By modeling the distribution of features learned specifically for the task of discriminating normal reconstructions, the proposed framework effectively bridges the gap between reconstruction fidelity and anomaly sensitivity without relying on external task-agnostic priors.

The main contributions of this paper can be summarized as follows:

1. We propose a new regularized latent space (RLS) encoding module that adds compactness constraints to the encoder generator mapping to balance reconstruction accuracy and anomaly detection capability during training.
2. We design a two-phase generative–discriminative architecture that intelligently repurposes a pretrained generative model into a multi-scale discriminator (MSD), enabling fine-grained, localized anomaly perception.
3. We introduce a probabilistic patch-distribution modeling strategy that fits multivariate Gaussian distributions to the task-specific, multi-scale adversarial features extracted by the discriminator, yielding precise and interpretable defect segmentations.
4. We propose a unified encoding–generation–discrimination anomaly detection paradigm that integrates representation learning, image reconstruction, and anomaly localization within a single framework. Extensive ablation studies validate the effectiveness of the proposed design.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of the recent literature related to unsupervised industrial anomaly detection. Section 3 introduces the overall architecture of our proposed generative–discriminative framework and details its specific algorithmic implementations. Section 4 outlines the experimental procedures and presents a thorough analysis of both quantitative and qualitative results. Finally, Section 5 summarizes our main contributions and concludes the paper.

2. Related Work

2.1. Unsupervised Models for Anomaly Detection

Due to the inherent unpredictability and extreme scarcity of annotated defect samples in real-world manufacturing environments, unsupervised learning has emerged as the dominant paradigm for industrial surface anomaly detection. Over the past few years, extensive research has been dedicated to modeling the distribution of normal data to identify out-of-distribution patterns. Broadly, existing state-of-the-art approaches can be classified into two predominant trajectories based on their core mechanisms: reconstruction-based methods and feature embedding-based methods. In this section, we provide a brief review of the recent advancements and inherent limitations within these two main lines of research.

Reconstruction-based approaches fundamentally operate on the assumption that models trained exclusively on defect-free samples will fail to accurately reconstruct anomalous regions. Classic architectures in this trajectory, such as Variational Autoencoders (VAEs) [22] and GANomaly [23], learn a generative manifold of normal data and subsequently identify anomalies by computing pixel-level discrepancies between the input image and its reconstructed output. DRAEM [24], a typical representative of generative models, utilizes simulated pseudo-anomaly strategies to enhance sensitivity to subtle surface defects. However, its generalization ability is essentially limited by the diversity and authenticity of synthetic defects, which leads to fuzzy localization due to domain differences when encountering anomalies of different scales. Recent studies have shown that combining multi-scale architectures with GANs can effectively improve fine-grained anomaly perception, as demonstrated in multi-scale surface inpainting [25]. However, it is difficult to balance the mask size with local contextual information, and it lacks explicit constraints on the distribution of feature space. Although these methods are very intuitive and adept at capturing the global semantic consistency of target objects, they all have a key bottleneck called anomaly reconstruction. The powerful generalization capabilities of modern deep neural networks enable these generators to often learn an identity mapping that faithfully reproduces even the anomalous regions. Consequently, the expected pixel-level

discrepancies are minimized, severely diminishing the model's sensitivity to fine-grained, localized defects.

To address the limitations of pixel-level reconstruction, feature embedding-based methods have emerged as a dominant alternative. These techniques typically leverage deep neural networks pretrained on large-scale datasets to extract rich, patch-level feature representations. State-of-the-art frameworks then construct a normative distribution of these deep features using various statistical modeling techniques: PatchSVDD [26] employs hyperspheres, SPADE [27] utilizes memory banks, and PaDiM [28] models multivariate Gaussian distributions. Furthermore, advanced methods like DifferNet [29] apply normalizing flows for precise probability density estimation within the feature space [30], detecting anomalies as outliers based on feature distances. While embedding-based methods offer superior local discriminability, their primary drawback lies in a heavy reliance on fixed, external priors [31]. Because the pretrained backbones are not tailored to the specific textural and structural characteristics of industrial surfaces, a significant domain gap often emerges, limiting overall performance. Additionally, the patch-centric nature of these methods frequently neglects the holistic structural integrity of the target object. Recognizing this persistent dichotomy—where reconstruction methods capture global structure but miss local details, and embedding methods excel locally but lack global coherence and task-specific optimization—our work seeks to bridge these two paradigms. By combining the generative fidelity of a task-specific model with the robust discriminative power of a learned, multi-scale feature space, our framework achieves highly precise localization without depending on external pretraining [32,33].

2.2. Attention Mechanisms for Feature Enhancement

In unsupervised anomaly detection, extracting discriminative and robust feature representations is crucial for distinguishing subtle defects from complex background variations. To enhance feature expressiveness, attention mechanisms have been widely integrated into deep learning architectures, allowing models to dynamically focus on informative regions while suppressing irrelevant noise. Early approaches, such as squeeze-and-excitation (SE) networks or convolutional block attention module (CBAM) [34], primarily rely on channel-wise re-weighting or simplistic spatial pooling. However, these methods often compress feature maps excessively, leading to a loss of fine-grained spatial details that are critical for accurate defect localization. Conversely, while global self-attention modules or vision transformers (ViT) can effectively capture long-range dependencies, they introduce prohibitive computational costs. To address this dilemma, the coordinate attention module (CAM) [35] has emerged as a highly efficient alternative.

Unlike conventional channels or self-attention mechanisms, CAM introduces direction-aware encoding along both the horizontal and vertical axes, enabling the network to retain precise location information while maintaining low computational overhead. This is particularly beneficial for structured industrial defects such as elongated scratches and linear cracks, where spatial continuity is essential. Moreover, compared to global self-attention [36], it offers a superior trade-off between spatial sensitivity and efficiency, making it suitable for deployment without high computational cost. And coordinate attention provides an optimal balance between spatial information retention, computational complexity, and localization accuracy for this task. Additionally, the latent space regularization constraint promotes compactness within the encoded manifold, further enhancing mapping stability and anomaly separability.

In summary, although existing latent space regularization approaches have demonstrated effectiveness in improving representation quality, most rely on fixed regularization objectives that cannot adapt to the evolving latent distribution during training. Motivated

by these limitations, we propose a regularized latent adaptive framework. By integrating latent space regularization, multi-scale discriminative learning, and embedding-based anomaly localization within a unified generative–discriminative paradigm, our method effectively addresses these challenges and provides a comprehensive solution for industrial inspection.

3. Method

In this paper, our framework is built upon StyleGAN2-ADA, a choice motivated by its unique architectural advantages that are highly synergistic with our proposed innovations. Beyond its state-of-the-art performance in few-shot learning, it features a highly disentangled latent space $W+$, which provides an ideal manifold for our regularized latent space encoding to effectively enforce compactness and separability. Furthermore, its hierarchical, multi-resolution synthesis process offers a natural structural foundation for our multi-scale discriminator to perform fine-grained, patch-level analysis at native spatial scales. For a detailed step-by-step description of the training algorithm and hyperparameter updates, we provide the pseudo-code of the algorithm in the Supplementary Materials.

This section describes the proposed unsupervised anomaly detection framework, which incorporates two main innovations: (1) a regularized latent space (RLS) encoding module, and (2) a multi-scale discriminator (MSD) that serves a dual role. During training, it provides adversarial supervision; during inference, its learned multi-scale features are utilized for localization via probabilistic distribution modeling. The overall anomaly detection architecture is illustrated in Figure 1.

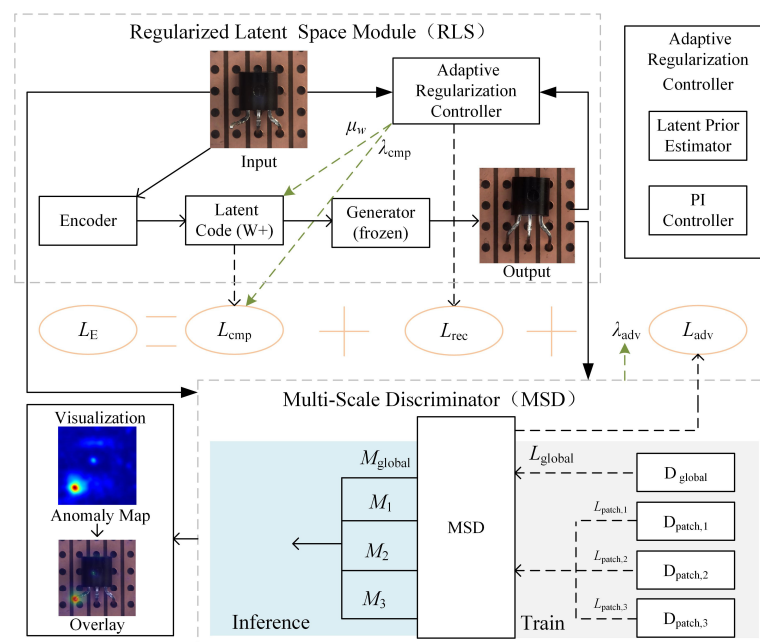


Figure 1. Overview of the proposed anomaly detection architecture. Solid arrows indicate the forward data flow; dashed black lines denote the computation of loss components; and dashed green lines represent the adaptive control signals for parameter updates.

3.1. The Architecture of Regularized Latent Space Encoding

This section details the architecture and regularization strategies of the encoder, the overall structure of which is illustrated in Figure 2. The encoder projects an input image into the latent space $W+$. Let G denote the generator pretrained only on normal images, learning the manifold distribution $X_{\text{normal}} \subset \mathbb{R}^{C \times H \times W}$. To obtain semantically meaningful and discriminative latent representations for anomaly detection, we propose an encoding

design that integrates attention-driven feature extraction with explicit latent regularization. The encoder incorporates a coordinate attention module that embeds positional information into channel attention to capture long-range spatial dependencies, thereby enhancing the model's ability to capture fine-grained spatial and contextual features. This design enables the encoder to capture long-range spatial dependencies and fine-grained contextual structures, thereby improving the separation between normal and abnormal latent manifolds. Additionally, the regularization constraint encourages compactness within the latent distribution, promoting stable mapping and enhanced anomaly separability.

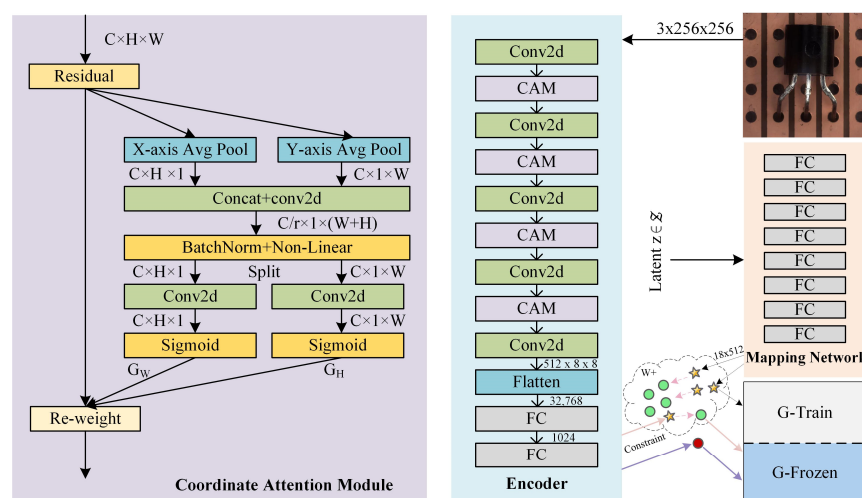


Figure 2. The network structure of the encoder for the latent space.

In this section, let $w(x) = E(x) \in \mathbb{R}^{18 \times 512}$ denote the latent code of an image in the latent space $W+$. For simplicity, these latent codes are represented as vectors in a high-dimensional space, and all operations such as addition and subtraction are performed element-wise.

To ensure that latent representations of normal samples remain compact and separable from anomalies, a compactness loss is imposed:

$$L_{\text{cmp}} = E_{x \sim X_{\text{normal}}} \|E(x) - \mu_w\|_2^2 \quad (1)$$

where μ_w denotes the global centroid of the normal latent manifold. This centroid is dynamically estimated via an exponential moving average (EMA) over the mini-batches. At each training step t , given a mini-batch of normal samples B_t , the centroid is updated as follows:

$$\mu_w(t) = (1 - \alpha)\mu_w^{(t-1)} + \alpha E_{x \sim B_t}[w(x)] \quad (2)$$

where $E_{x \sim B_t}[w(x)]$ denotes the empirical mean of the latent codes within the mini-batch B_t , and $\alpha \in (0, 1)$ is the momentum parameter.

A reconstruction constraint is also applied to maintain the generator fidelity:

$$L_{\text{rec}} = E_{x \sim X_{\text{normal}}} \|x - \hat{x}\|_1 \quad (3)$$

where $\hat{x}$ is defined as $\hat{x} = G(E(x))$.

To dynamically balance reconstruction fidelity and latent compactness, we introduce an adaptive adjustment scheme for the compactness weight λ_{cmp} . Inspired by control theory [37], we formulate this adjustment as a discrete proportional integral (PI) control problem. Specifically, the controller updates λ_{cmp} based on the deviation of the current reconstruction error, $L_{\text{rec}}(t)$, from a predefined target threshold, τ .

The error term at training step t is defined as:

$$e(t) = L_{\text{rec}}(t) - \tau \quad (4)$$

To prevent the issue of integral wind-up—where the unbounded accumulation of past errors can lead to instability—we approximate the integral term using an EMA with a momentum parameter $\beta \in (0, 1)$:

$$I(t) = \beta \cdot I(t-1) + (1 - \beta) \cdot e(t) \quad (5)$$

Here, this acts as a leaky integrator, effectively maintaining a history of recent errors while gradually decaying the influence of older ones, ensuring bounded and stable control dynamics.

The final PI update rule for λ_{cmp} is then formulated as:

$$\lambda_{\text{cmp}}(t+1) = \lambda_{\text{cmp}}(t) + \eta_P \cdot e(t) + \eta_I \cdot I(t) \quad (6)$$

where η_P and η_I are the proportional and integral gains. This dual-term feedback control ensures a smooth and stable adaptation λ_{cmp} , mitigating oscillations and robustly guiding the training convergence.

The total regularization objective for the encoder, L_{RLS} , is formulated as a dynamic weighted sum, where the PI-controlled $\lambda_{\text{cmp}}(t)$ adaptively balances the two competing objectives:

$$L_{\text{RLS}} = L_{\text{rec}} + \lambda_{\text{cmp}}(t) L_{\text{cmp}} \quad (7)$$

This formulation establishes reconstruction as the baseline objective and employs the weight $\lambda_{\text{cmp}}(t)$ to dynamically regulate the intensity of the compactness constraint. This mechanism ensures a robust equilibrium, preventing the model from prioritizing reconstruction fidelity at the expense of latent separability, or vice versa.

3.2. The Architecture of Multi-Scale Discriminator

To address the specific challenges of industrial anomaly detection, where anomalies can manifest as both large-scale structural deviations and subtle textural inconsistencies, a standard single-output discriminator is insufficient. We therefore propose a multi-scale discriminator, designed to enforce realism across a hierarchy of feature resolutions. Figure 3 illustrates the structure of the discriminator.

The discriminator is composed of two primary components: a deep residual backbone and a set of parallel projection heads. The backbone is constructed by stacking a series of StyleGAN2-based residual D Blocks. Figure 3, on the left, progressively downsamples the input image from 256×256 to 4×4 , extracting increasingly abstract features at each stage. This design ensures powerful and stable feature representation. To achieve a better balance between global realism and local texture sensitivity, while the global branch evaluates overall image realism, the patch-level heads focus on fine-grained texture inconsistencies, thereby enhancing the model's capability in local anomaly perception. Building upon this backbone, we introduce three projection heads, denoted as $D = \{D_1, D_2, D_3\}$. As illustrated in Figure 3, right, these heads branch off from the outputs of intermediate D Blocks, receiving features at resolutions of 64×64 , 32×32 , and 16×16 , respectively. Each head, implemented as a simple 3×3 convolutional layer, projects the multi-channel feature map into a single-channel, patch-wise realism map. This forces the generator to produce plausible textures and local structures at multiple scales. Concurrently, the main branch of the backbone continues to the final block, which processes the most abstract 4×4 features

to produce a single, global realism score, ensuring the overall coherence and semantic correctness of the generated image.

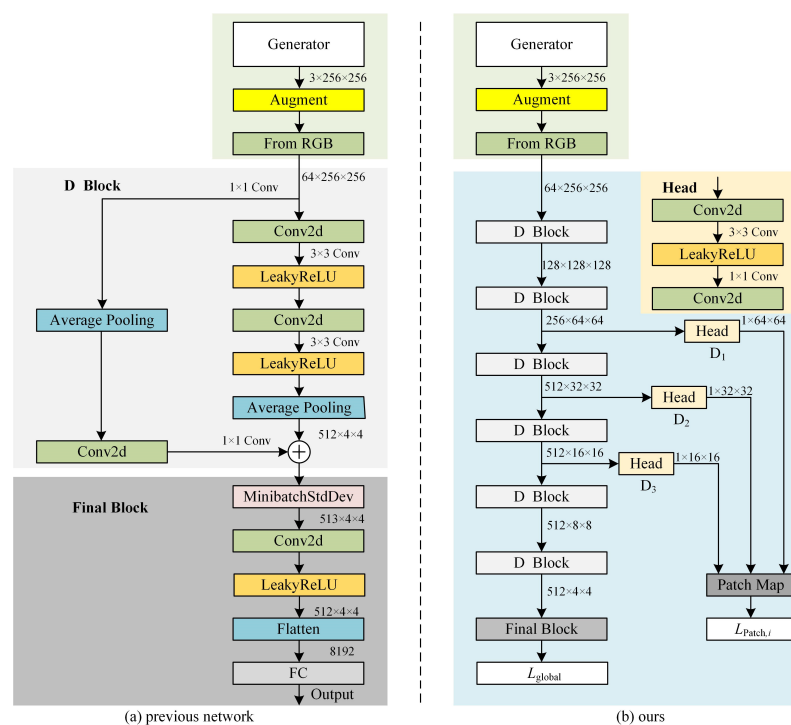


Figure 3. The network structure of the multi-scale discriminator.

Following the principles of Geometric GAN [38] optimization, the Hinge loss is adopted to stabilize discriminator optimization and improve discrimination between real and reconstructed samples. The discriminator is optimized using the following objective:

$$\ell_{\text{hinge}}(D(x), t) = E[\max(0, 1 - tD(x))] \quad (8)$$

where $D(x)$ denotes the discriminator output and $t \in \{+1, -1\}$ indicates the label for real and generated samples, respectively. This formulation enforces a margin-based separation between real and synthetic data distributions, stabilizing the gradient dynamics during adversarial learning.

The total discriminator loss L_D combines the contributions of the global path and the three patch-wise heads, with adjustable weighting coefficients λ_i :

$$L_D = L_{\text{global}} + \sum_{i=1}^n \lambda_i L_{\text{patch},i} \quad (9)$$

L_{global} is the hinge loss applied to the discriminator's global realism score. The first term penalizes the discriminator for assigning low scores to real images, while the second term penalizes high scores assigned to reconstructed images $\hat{x}$.

$$L_{\text{global}} = E_{x \sim P_{\text{data}}}[\max(0, 1 - D_{\text{global}}(x))] + E_{\hat{x} \sim P_G}[\max(0, 1 + D_{\text{global}}(\hat{x}))] \quad (10)$$

For each scale, the discriminator outputs a spatial realism map D_i , and the loss is averaged across all pixel locations:

$$L_{\text{patch},i} = \text{mean}_{u,v}[\ell_{\text{hinge}}(D_i(x)_{u,v} + 1) + \ell_{\text{hinge}}(D_i(\hat{x})_{u,v} - 1)] \quad (11)$$

This multi-scale hinge loss encourages each discriminative head to focus on its own receptive field, resulting in complementary supervision between global and local branches. The adversarial component is formulated as:

$$L_{\text{adv}} = -E_{\hat{x} \sim p_G} \left[D_{\text{global}}(\hat{x}) + \sum_{i=1}^3 \lambda_i \cdot \text{mean}_{u,v} D_i(\hat{x})_{u,v} \right] \quad (12)$$

The encoder is optimized jointly with a fixed generator using the RLS reconstruction loss and an adversarial objective that compels them to deceive the MSD at all scales:

$$L_E = L_{\text{RLS}} + \lambda_{\text{adv}} L_{\text{adv}} \quad (13)$$

Among them, $\lambda_{\text{adv}} \in (0, 1)$, there is an artificially regulated balance coefficient between implicit regularization and adversarial training.

The overall optimization objectives of the proposed adversarial framework are summarized as follows:

$$\min_E \max_D V(E, D) \quad (14)$$

where $V(E, D)$ denotes the adversarial value function that measures the discrepancy between real samples and their reconstructed counterparts. This objective encourages the encoder to produce latent codes that lead to realistic reconstructions indistinguishable from normal samples, while the discriminator learns to identify subtle deviations, thus establishing a balanced adversarial dynamic. This joint formulation enables the model to maintain global generative stability while improving local discriminative sensitivity, providing robust and fine-grained anomaly localization under unsupervised conditions.

3.3. Modeling the Feature Distribution

To achieve high-resolution anomaly localization, we adopt a probabilistic feature modeling strategy inspired by PaDiM [28]. However, a crucial distinction of our work is that we do not rely on a generic, pretrained backbone. Instead, we leverage the powerful, task-specific feature representations learned by our adversarially trained multi-scale discriminator. During a preliminary step after training, we build a statistical model of the normal feature distribution. All normal images from the training set are passed through our trained Encoder and MSD. For each of the n discriminator branches, D_i , we extract the intermediate feature maps and compute the patch-wise mean and covariance matrix for the features at each spatial location. This results in a set of n multivariate Gaussian models, each corresponding to a specific scale of normalcy.

During inference, a test image x is fed through the same E-MSD pipeline. For each D_i , we extract its feature map and compute the Mahalanobis distance of each patch-wise feature vector with respect to the corresponding pre-computed Gaussian model for that location and scale. This process yields independent, scale-specific anomaly score maps:

$$M_i(x)_{u,v} = \sqrt{(f_i(x)_{u,v} - \mu_{i,u,v})^T \Sigma_{i,u,v}^{-1} (f_i(x)_{u,v} - \mu_{i,u,v})} \quad (15)$$

where $f_i(x)_{u,v}$ is the feature vector at location u, v from the branch D_i , and $(\mu_{i,u,v}, \Sigma_{i,u,v})$ are the learned statistics.

The scale-specific anomaly maps are upsampled to the input image resolution and fused via averaging to produce the final, high-resolution anomaly map. This strategy leverages the statistical modeling of discriminator features, rather than just relying on reconstruction error, yielding interpretable and granular anomaly localization. Because

E-MSD is trained under a regularized latent constraint, the normal features become more compact, improving the Mahalanobis separation between normal and anomalous patches.

3.4. Evaluation Metrics

In this study, multiple quantitative metrics are employed to comprehensively evaluate the performance of the proposed framework in anomaly detection accuracy. For anomaly detection performance, we adopt the area under the receiver operating characteristic curve (AUROC) [39] at both the image- and pixel-level. The AUROC is a threshold-independent performance indicator that measures the trade-off between the true positive rate (TPR) and the false positive rate (FPR).

The TPR and FPR are defined as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (17)$$

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. The AUROC value is calculated by integrating the TPR over the FPR range:

$$\text{AUROC} = \int_0^1 \text{TPR}(\text{FPR})d(\text{FPR}) \quad (18)$$

In this study, the AUROC is applied at two different granularities.

Image-Level AUROC: This metric assesses the model's capability to classify an entire image as either normal or anomalous. For each test image, a single anomaly score is assigned. It reflects the framework's reliability in binary industrial quality grading.

Pixel-Level AUROC: This metric evaluates the precision of anomaly localization. Each pixel in the anomaly map is treated as an individual classification sample and compared against the corresponding pixel in the ground-truth mask. It indicates the model's ability to accurately segment defective regions from the background, regardless of the anomaly's shape or size.

4. Experiments

4.1. Experiment Details

4.1.1. Datasets

The MVTec AD dataset [40] serves as a comprehensive and standard benchmark in this field, containing 5354 high-resolution color images across 15 subcategories. Each subcategory consists exclusively of defect-free images for training, and test images containing various defect types. The dataset, as shown in Figure 4, spans a wide range of industrial products, providing a diverse set of challenges for evaluating the generative fidelity and reconstruction stability of models.

4.1.2. Experimental Environment and Hyperparameter Settings

All experiments were conducted with an Intel® Core™ i7 CPU and a single NVIDIA GeForce RTX 4090 GPU. The implementation was developed in Python 3.9.1 using PyTorch 2.1.0 and CUDA 12.1. All input images from the MVTec AD dataset were resized to a uniform resolution of 256×256 pixels.

Following the official configuration, our framework's generator is built upon its architecture. This choice provides a powerful pretrained manifold of normal data and leverages its sophisticated synthesis mechanism, which utilizes weight demodulation instead of the original AdaIN to inject style information, thereby preventing normalization

artifacts and ensuring high-fidelity image reconstruction. Unlike the original discriminator, R1 regularization is not applied to the global discriminator branch in our framework. Since the global discriminator is expected to learn anomaly-sensitive feature representations in addition to adversarial discrimination, excessive gradient smoothing may weaken its ability to capture subtle abnormal patterns. Therefore, the stability of the global branch is primarily maintained through latent space regularization and adaptive augmentation. However, R1 regularization is retained for the scale-specific discriminator heads to stabilize local adversarial learning.

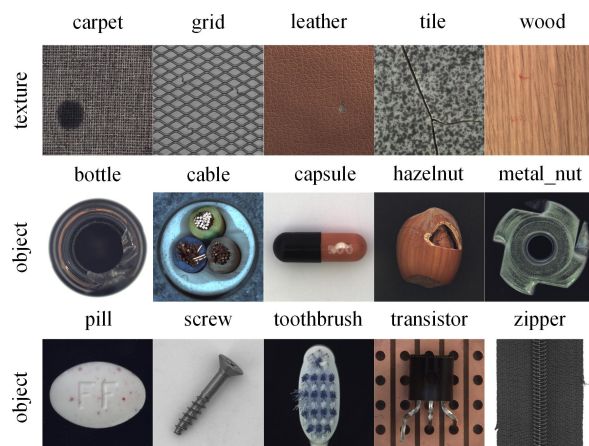


Figure 4. Representative images of the 15 categories in the MVTec AD dataset, including 10 object categories and five texture categories, which are used to evaluate the performance of the proposed unsupervised anomaly detection framework.

Our training strategy involves a two-stage process. First, the generator is pretrained on anomaly-free images and then remains frozen to preserve the learned manifold of normal data. In the second stage, the pre-trained generator was frozen to preserve the learned normal image manifold, while the encoder and the multi-scale discriminator were jointly optimized. The AdamW optimizer was adopted with $\beta_1 = 0.0$, $\beta_2 = 0.99$, a learning rate of 1×10^{-4} , and a weight decay of 0.01. Training was performed for 100 epochs with a batch size of 16. R1 regularization ($\gamma = 10$) was applied to each scale-specific discriminator. All experiments were conducted with over 5 random seeds for reproducibility.

4.1.3. Adaptive Discriminator Augmentation Strategy

We employ adaptive discriminator augmentation (ADA) strategically across the two training phases. During the generator pretraining stage, standard ADA is applied to enhance the generator's robustness to photometric variations. However, for the subsequent encoder and optimization, we adopt a geometry-preserving augmentation policy. Specifically, destructive geometric transformations that alter the structural integrity of defects—such as large rotations, cropping, and cutout—are strictly disabled to ensure spatial consistency between training and inference. Instead, we utilize a constrained set of augmentations, including non-destructive geometric shifts and photometric perturbations. This tailored strategy prevents the model from learning invariance to critical defect geometries while maintaining sufficient regularization to stabilize adversarial training.

Considering the task requirements of industrial anomaly detection, excessive enhancement is harmful to model training, and the strategy of the original ADA has been optimized. As shown in Figure 5, as P increases from 0 to 0.5, texture deformation and color perturbation become progressively stronger, introducing richer variations while maintaining the overall structural consistency of normal samples.

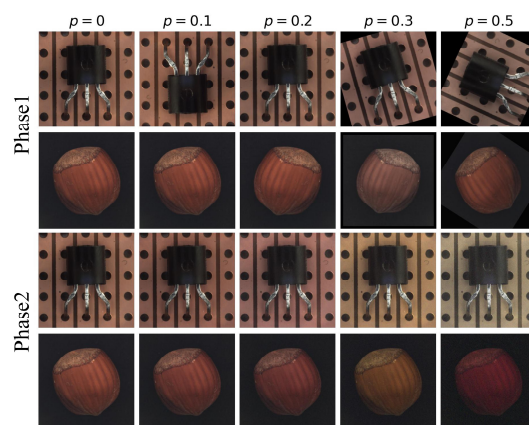


Figure 5. The visual effects of adaptive data augmentation on sample images.

4.2. Anomaly Detection Results

Figure 6 visualizes the anomaly localization results of representative objects in the MVTec AD benchmark. The figure illustrates the input images, ground-truth masks, predicted anomaly maps, and their overlaps, respectively. The visualizations demonstrate that our model produces highly interpretable heatmaps with a sharp contrast between normal and anomalous regions. Whether dealing with distinct structural defects on objects or subtle textural variations on surfaces, our model precisely localizes the anomalous regions at a fine-grained scale without triggering false positives in normal areas. More comprehensive visualization results are available in the Supplementary Materials.

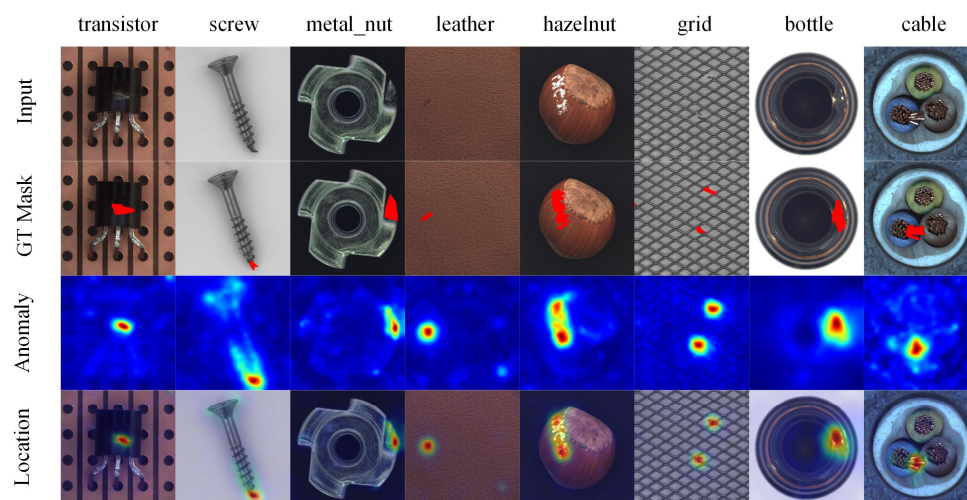


Figure 6. Qualitative anomaly localization results produced by the proposed method on representative samples from the MVTec AD dataset. From top to bottom, each group shows the input image, the ground-truth mask, the predicted anomaly map, and the corresponding localization result.

4.3. Comparison with Existing Methods

To validate the effectiveness of the proposed detection framework, we conducted extensive quantitative experiments, comparing our method with several state-of-the-art unsupervised anomaly detection approaches, including vis.expl.VAE [22], GANomaly [23], PatchSVDD [26], SPADE [27], PaDiM [28], and DifferNet [29]. All methods were evaluated under the same experimental settings to ensure fairness and reproducibility. For the discriminator loss, all patch-head weights were uniformly set to 1.0 to ensure that each scale contributes equally to the adversarial objective without introducing a preferential bias.

To quantitatively validate the effectiveness of the proposed framework, we compare its image-level anomaly detection performance against several state-of-the-art unsupervised

baseline methods. As summarized in Table 1, our method establishes a significant performance with an average image-level AUROC of 98.6%, outperforming both reconstruction-based models and embedding-based approaches. This superior performance is further illustrated in Figure 7, where our method clearly forms the outermost bounding polygon, indicating not only peak accuracy but also remarkable stability across diverse categories.

Table 1. Anomaly detection performance (image-level AUROC [%]) on the MVTec AD dataset. “*” denotes a result with a backbone specifically selected for the task of image-level anomaly detection, which we could not reproduce. And the results for “Ours” are averaged over three independent runs, and the standard deviation reflects the variance across five random seeds.

Category	GANomaly	PatchSVDD	PaDiM	DifferNet	PaDiM *	Ours
bottle	89.9	98.8	-	99.4	-	98.7 $\pm$ 0.3
cable	77.7	90.4	-	96.1	-	98.4 $\pm$ 0.5
capsule	73.2	77.1	-	89.9	-	98.8 $\pm$ 0.4
carpet	69.8	92.6	-	92.8	-	98.6 $\pm$ 0.2
grid	73.8	95.1	-	89.5	-	98.4 $\pm$ 0.6
hazelnut	79.5	93.3	-	99.1	-	98.8 $\pm$ 0.4
leather	84.0	91.2	-	97.1	-	98.1 $\pm$ 0.3
metal nut	70.3	94.5	-	96.1	-	98.8 $\pm$ 0.3
pill	74.2	86.1	-	89.7	-	97.4 $\pm$ 0.5
screw	74.6	81.3	-	96.3	-	98.9 $\pm$ 0.2
tile	79.4	97.8	-	99.2	-	98.2 $\pm$ 0.4
toothbrush	65.1	99.7	-	98.6	-	98.5 $\pm$ 0.3
transistor	79.6	91.2	-	93.7	-	99.1 $\pm$ 0.2
wood	84.7	96.8	-	99.6	-	98.7 $\pm$ 0.4
zipper	74.6	97.9	-	96.7	-	98.9 $\pm$ 0.3
average	76.7	92.3	95.2	95.6	97.9 *	98.6 $\pm$ 0.4

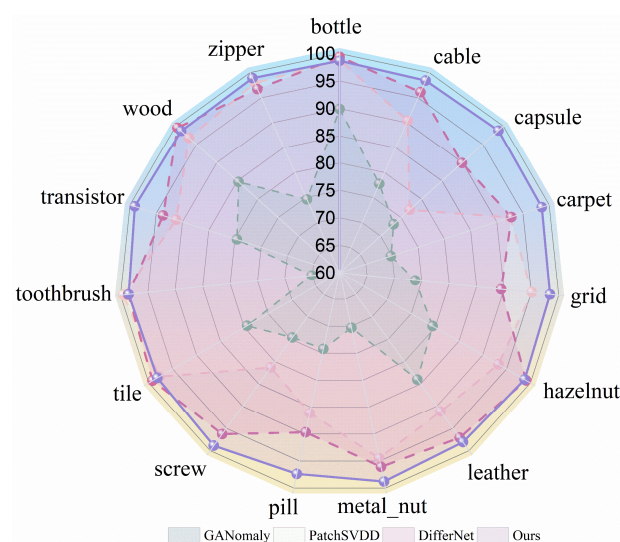


Figure 7. The radar chart provides a category-wise comparison of image-level AUROC performance, demonstrating the effectiveness in anomaly detection across different categories.

To further evaluate the fine-grained defect localization capability of our framework, we compare the pixel-level AUROC scores against established baselines, as presented in Table 2. Our method achieves a state-of-the-art average score of 98.4%, outperforming strong feature-embedding methods such as PaDiM (97.3%) and SPADE (96.0%). Beyond the highest overall accuracy, the most striking advantage of our approach is its extraordinary

robustness across diverse defect types, which is vividly illustrated by the line chart in Figure 8. As observed, the performance curve of our method consistently operates at the uppermost boundary with minimal fluctuation across all 15 categories. In stark contrast, competing baselines exhibit severe variance and sharp performance drops on challenging categories. This exceptional cross-category stability is a direct benefit of integrating our multi-scale discriminator with probabilistic patch distribution modeling. By extracting and statistically modeling multi-scale adversarial features, our framework effectively eliminates the reliance on generic pretrained backbones. It accurately captures localized textural inconsistencies and fine-grained spatial deviations, thereby yielding highly precise and robust pixel-level segmentations.

Table 2. Anomaly detection performance (pixel-level AUROC [%]) on the MVTec AD dataset.

Category	vis.expl.VAE	PatchSVDD	SPADE	PaDiM	Ours
bottle	86.7	98.1	98.3	98.1	98.8 ± 0.3
cable	89.1	96.8	97.1	96.6	98.7 ± 0.4
capsule	74.0	95.7	98.2	98.3	98.8 ± 0.3
carpet	78.2	92.5	96.8	99.1	98.9 ± 0.2
grid	75.6	96.1	93.6	97.0	98.4 ± 0.4
hazelnut	97.4	97.5	98.3	98.2	98.8 ± 0.3
leather	95.1	97.4	97.7	98.4	98.6 ± 0.4
metal nut	94.2	98.0	98.1	97.2	97.9 ± 0.5
pill	83.2	95.2	96.5	95.4	98.2 ± 0.4
screw	96.5	95.9	98.9	98.7	98.8 ± 0.3
tile	80.1	91.6	87.1	93.1	95.8 ± 0.5
toothbrush	94.3	98.1	98.2	98.8	98.7 ± 0.3
transistor	93.9	97.3	94.4	98.5	98.9 ± 0.3
wood	77.0	90.6	89.2	94.5	97.9 ± 0.4
zipper	78.1	95.5	96.9	98.0	98.3 ± 0.4
average	86.2	95.8	96.0	97.3	98.4 ± 0.4

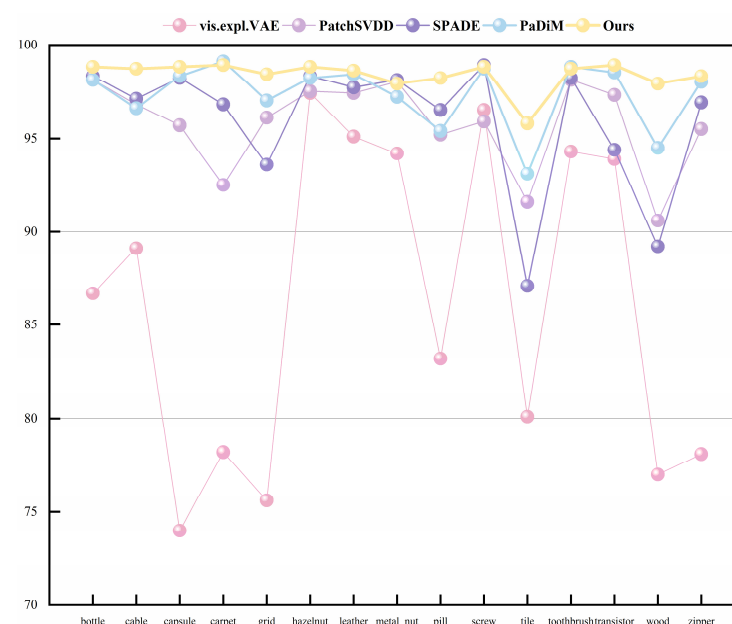


Figure 8. The line chart provides a category-wise comparison of pixel-level AUROC performance, demonstrating the effectiveness in anomaly localization across different categories.

4.4. Ablation Study

4.4.1. Effect of Adaptive Regularization Controller

To evaluate the efficacy of the adaptive regularization controller, we first establish the rationale for its key parameters: the proportional and integral gains (η_P , η_I), and the target threshold τ . These parameters were selected to ensure a critically damped response, preventing potential oscillations in the regularization weight λ_{cmp} while maintaining stable convergence. The selected configuration ($\eta_P = 1 \times 10^{-2}$, $\eta_I = 1 \times 10^{-3}$, $\tau = 0.8 \times \text{initial mean error}$) consistently yielded stable training dynamics across different categories of the dataset. The τ serves as the operating setpoint of the controller, ensuring that latent compactness is progressively emphasized only after the reconstruction error approaches the desired target level, thereby maintaining a balanced trade-off between reconstruction fidelity and anomaly separability.

Unlike the fixed-weight and original proportional strategies, the dynamic PI controller adaptively adjusts λ_{cmp} to balance reconstruction fidelity and anomaly sensitivity. Table 3 compares the performance of different strategies for weighting the compactness loss L_{cmp} .

Table 3. Comparison of the regularization controller parameter.

Controller Type	Configuration	Image-AUROC	Pixel-AUROC
No L_{cmp}	$\lambda_{\text{cmp}} = 0$	95.1	93.9
Fixed Weight	$\lambda_{\text{cmp}} = 0.5$	97.0	96.2
P-Controller	$\lambda_{\text{cmp}}(t)$	98.1	97.5
Ours	$\lambda_{\text{cmp}}(t)$	98.6	98.4

For the best fixed-weight baseline, the compactness coefficient λ_{cmp} was set to 0.5, following the configuration [41,42], evaluated $\lambda_{\text{cmp}} \in (0.1, 1.0)$ on the transistor set. We observed that smaller values, such as 0.1, failed to effectively cluster normal samples, while larger values, such as 1.0, overly constrained the latent space, degrading reconstruction fidelity. Consistent with prior work that empirically selects a balanced mid-range weight, this setting imposes a sufficient penalty to ensure latent compactness without overwhelming the reconstruction objective, thus avoiding posterior collapse, and provides the best trade-off between accurate reconstruction and anomaly discriminability, and is thus adopted as our default fixed weight.

In terms of tuning cost, the fixed-weight strategy is the most expensive, requiring an extensive grid search over λ_{cmp} for each new dataset. In contrast, the PI-controller, while introducing three new hyperparameters (τ , η_P , η_I), can be effectively managed with a simple set of heuristics. Therefore, while its initial setup cost is moderate, it is significantly more practical and generalizable than the exhaustive search required by the fixed-weight baseline. The P-controller offers the lowest tuning cost, but at the expense of performance and stability. The results demonstrate that introducing the proposed PI control significantly improves both image- and pixel-level AUROC scores, achieving the most stable convergence throughout training.

4.4.2. Effect of Multi-Scale Discriminator

The introduction of the multi-scale discriminator significantly improves fine-grained defect localization. By aggregating discriminative responses from small, medium, and large receptive fields, it enhances sensitivity to subtle texture changes while maintaining robustness to scale variations. Table 4 lists the results of different scale discriminators. In terms of average pixel-AUROC, our method only improves the score by 1.7% over the main discriminator.

Table 4. Comparison of pixel-AUROC for different scale discriminators in the MVTec AD transistor data.

Model Configuration				Avg Pixel-AUROC	Bent Only	Cut Only	Damaged Case Only	Good Only	Misplace Only
D_{global}	D_1	D_2	D_3						
✓	-	-	-	97.2	96.8	96.4	97.2	98.3	97.4
-	✓	✓	✓	97.8	98.0	97.6	97.9	98.6	97.1
✓	✓	-	-	98.1	98.1	98.2	97.6	98.7	97.8
✓	-	✓	-	98.6	98.9	98.5	98.2	99.0	98.4
✓	-	-	✓	98.0	98.3	97.8	97.4	98.8	97.9
✓	✓	✓	✓	98.9	98.5	98.8	98.5	100	98.7

Overall, the quantitative results verify that our method not only achieves state-of-the-art detection accuracy but also maintains stable training dynamics under limited normal data, which is essential for real-world industrial inspection scenarios.

5. Conclusions

In this work, we presented a regularized latent adaptive framework, a unified unsupervised anomaly detection framework that combines regularized latent encoding with multi-scale discriminative learning. By integrating compactness regularization into the latent space and enforcing detailed realism constraints through multi-scale discrimination, our proposed framework achieves a balanced synergy between generative reconstruction and discriminative localization. Extensive tests on the MVTec AD benchmark showed that our model not only outperforms existing GAN-based and embedding-based methods but also demonstrates excellent training stability and robustness under conditions with limited data. These results confirm that combining geometric regularization with adversarial hierarchy offers an effective approach for bridging the longstanding gap between reconstruction quality and anomaly sensitivity in industrial visual inspection. Beyond mere numerical improvements, the importance of our framework is in its conceptual contribution: it redefines unsupervised anomaly detection as a dual-space consistency challenge—where the latent manifold of normalcy must stay geometrically tight, while the perceptual space must maintain multi-scale realism.

While the proposed framework demonstrates significant efficacy, certain limitations warrant acknowledgment. First, the evaluation is primarily centered on the MVTec AD benchmark; thus, its generalization across a broader spectrum of industrial datasets and real-world scenarios requires further validation. Additionally, the current reliance on a regularized latent representation may constrain adaptability in the face of highly complex multi-instance distributions or non-stationary production environments. Furthermore, although the proposed framework achieves competitive detection performance, its computational efficiency and real-time inference capability have not yet been systematically investigated. Future work will therefore focus on optimizing the inference pipeline and conducting comprehensive FPS evaluations to facilitate deployment in practical industrial inspection systems. Nevertheless, this work successfully transcends the conventional reconstruction-versus-discrimination dichotomy. It suggests that the next frontier in anomaly detection lies not merely in increasing network depth or dataset scale, but in developing refined regularized representations that synergize generative and discriminative objectives within a unified optimization framework.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s26134151/s1>. Figure S1. The visual latent distributions in MVTec AD transistor data. Figure S2. Explanatory analysis and curve graphs of four categories in

the MVTec AD dataset. Figure S3. Visual display of a histogram in the MVTec AD transistor data. Figure S4. The proposed method produces qualitative anomaly localization results on supplementary samples in the MVTec AD dataset. From top to bottom, each group displays the input image, ground truth mask, predicted anomaly map, and corresponding localization results. Algorithm S1. Training procedure. Algorithm S2. Inference and anomaly localization procedure. Table S1. Ablation study on different augmentation strategies during the joint training phase. Performance is reported as pixel-AUROC (%) on the MVTec AD dataset with 5 random seeds for reproducibility. Table S2. Evaluation of reconstruction fidelity on normal samples with 5 random seeds for reproducibility. References [42–46] are cited in the supplementary materials.

Author Contributions: L.C. independently completed the writing of this paper. T.M., Y.L., X.L. (Xinran Lv), and X.L. (Xingfan Li) performed the investigation and provided many valuable comments. X.L. (Xiaoguang Li) guided the research, supervised the project, and provided the research resources and funding. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Scientific Research Project of the Jilin Provincial Development and Reform Commission, Key technology development of industrial product defect detection based on generative AI technology (Grant No. 2024C008-6).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The MVTec dataset is available at <https://www.mvtec.com/research-teaching/datasets/mvtec-ad>. (accessed on 23 June 2026). Additional processed data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank Xiaoguang Li for their valuable guidance on the theoretical research, experiments, and data analysis. During the preparation of this work, the authors used [ChatGPT (GPT-5.5)] in order to improve the language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

RLS	Regularized latent space
MSD	Multi-scale discriminator
CAM	Coordinate attention module
ADA	Adaptive discriminator augmentation
AUROC	Area under the receiver operating characteristic curve

References

1. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [\[CrossRef\]](#)
2. Liu, Y.; Zhang, C.; Dong, X. A survey of real-time surface defect inspection methods based on deep learning. *Artif. Intell. Rev.* **2023**, *56*, 12131–12170. [\[CrossRef\]](#)
3. Tulbure, A.A.; Dulf, E.H. A review on modern defect detection models using DCNNs—Deep convolutional neural networks. *J. Adv. Res.* **2022**, *35*, 33–48. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Lin, Y.; Chang, Y.; Tong, X.; Yu, J.; Liotta, A.; Huang, G.; Song, W.; Zeng, D.; Wu, Z.; Wang, Y.; et al. A survey on RGB, 3D, and multimodal approaches for unsupervised industrial image anomaly detection. *Inf. Fusion* **2025**, *121*, 103139. [\[CrossRef\]](#)
5. Wang, Z.; Li, T.; Qin, Y.; Sheng, B. Contrastive decoupling: Dynamic regularization for enhanced fine-grained image classification. *IEEE Trans. Circuits Syst. Video Technol.* **2026**, *36*, 7588–7599. [\[CrossRef\]](#)

6. 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, 10–17 June 2025; pp. 20405–20415. [\[CrossRef\]](#)
7. Li, Z.; Yan, Y.; Wang, X.; Ge, Y.; Meng, L. A survey of deep learning for industrial visual anomaly detection. *Artif. Intell. Rev.* **2025**, *58*, 279. [\[CrossRef\]](#)
8. Alzarooni, A.; Iqbal, E.; Khan, S.U.; Javed, S.; Moyo, B.; Abdulrahman, Y. Anomaly detection for industrial applications, its challenges, solutions, and future directions: A review. *arXiv* **2025**, arXiv:2501.11310. [\[CrossRef\]](#)
9. Jiang, N.; Ji, H.; Guan, Z.; Pan, Y.; Deng, C.; Guo, Y.; Wong, T.Y. A deep learning system for detecting silent brain infarction and predicting stroke risk. *Nat. Biomed. Eng.* **2025**, *9*, 1907–1919. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Wyatt, J.; Leach, A.; Schmon, S.M.; Willcocks, C.G. AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 649–655. [\[CrossRef\]](#)
11. Krassnig, P.J.; Gruber, D.P. ISP-AD: A large-scale real-world dataset for advancing industrial anomaly detection with synthetic and real defects. *arXiv* **2025**, arXiv:2503.04997. [\[CrossRef\]](#)
12. Xi, R.; Xu, Y.; Wu, S.; Ji, A.; Zhao, B. Bearing fault diagnosis method based on unsupervised pseudo-labels adapted for Locality Preserving Projections. *Nondestruct. Test. Eval.* **2025**, 1–32. [\[CrossRef\]](#)
13. Wang, J.; Lan, C.; Liu, C.; Ouyang, Y.; Qin, T.; Lu, W.; Chen, Y.; Zeng, W.; Yu, P.S. Generalizing to unseen domains: A survey on domain generalization. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 8052–8072. [\[CrossRef\]](#)
14. Zhao, S.; Yue, X.; Zhang, S.; Li, B.; Zhao, H.; Wu, B.; Krishna, R.; Gonzalez, J.E.; Sangiovanni-Vincentelli, A.L.; Seshia, S.A.; et al. A review of single-source deep unsupervised visual domain adaptation. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *33*, 473–493. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Barbato, F.; Toldo, M.; Michieli, U.; Zanuttigh, P. Latent space regularization for unsupervised domain adaptation in semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 2829–2839. [\[CrossRef\]](#)
16. Wang, J.; Lu, X.; Bennamoun, M.; Sheng, B. Non-rigid point cloud registration via anisotropic hybrid field harmonization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 7898–7915. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Bui, M.H.; Tran, T.; Tran, A.; Phung, D. Exploiting domain-specific features to enhance domain generalization. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Dutchess County, NY, USA, 2021; Volume 34, pp. 21189–21201.
18. Li, H.; Pan, S.J.; Wang, S.; Kot, A.C. Domain generalization with adversarial feature learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5400–5409. [\[CrossRef\]](#)
19. Bojanowski, P.; Joulin, A.; Lopez-Paz, D.; Szlam, A. Optimizing the latent space of generative networks. In Proceedings of the International Conference on Machine Learning (ICML), Stockholm, Sweden, 10–15 July 2018; pp. 600–609. Available online: <http://proceedings.mlr.press/v80/bojanowski18a.html> (accessed on 23 June 2026).
20. Huang, W. Novel view synthesis based on similar perspective. *Comput. Animat. Virtual Worlds* **2025**, *36*, e70006. [\[CrossRef\]](#)
21. Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; Aila, T. Training generative adversarial networks with limited data. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Dutchess County, NY, USA, 2020; Volume 33, pp. 12104–12114.
22. Liu, W.; Li, R.; Zheng, M.; Karanam, S.; Wu, Z.; Bhanu, B.; Radke, R.J.; Camps, O. Towards visually explaining variational autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 8639–8648. [\[CrossRef\]](#)
23. Akcay, S.; Atapour-Abarghouei, A.; Breckon, T.P. GANomaly: Semi-supervised anomaly detection via adversarial training. In Proceedings of the Asian Conference on Computer Vision (ACCV), Perth, Australia, 2–6 December 2018; pp. 622–637. [\[CrossRef\]](#)
24. Zavrtanik, S.; Kristan, M.; Skocaj, 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. 8310–8319. [\[CrossRef\]](#)
25. Liu, W.; Wang, C.; Zhang, Y. Industrial surface defect detection by multi-scale Inpainting-GAN. *Vis. Comput.* **2025**, *41*, 5643–5660. [\[CrossRef\]](#)
26. Yi, J.; Yoon, S. PatchSVDD: Patch-level SVDD for anomaly detection and segmentation. In Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2021; pp. 375–390. [\[CrossRef\]](#)
27. Park, T.; Liu, M.; Wang, T.; Zhu, J.-Y. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2332–2341. [\[CrossRef\]](#)
28. Defard, T.; Setkov, A.; Loesch, A.; Audigier, R. PaDiM: A patch distribution modeling framework for anomaly detection and localization. In Proceedings of the International Workshops and Challenges (ICPR), Virtual, 10–15 January 2021; pp. 475–489. [\[CrossRef\]](#)

29. Rudolph, D.; Wandt, B.; Rosenhahn, B. Same same but different: Semi-supervised defect detection with normalizing flows. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021; pp. 1906–1915. [\[CrossRef\]](#)
30. Kansizoglou, I.; Bampis, L.; Gasteratos, A. Deep feature space: A geometrical perspective. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 6823–6838. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Wu, K.; Zhu, L.; Shi, W.; Wang, W. Automated fabric defect detection using multi-scale fusion MemAE. *Vis. Comput.* **2025**, *41*, 723–737. [\[CrossRef\]](#)
32. Zoph, B.; Ghiasi, G.; Lin, T.Y.; Cui, Y.; Liu, H.; Cubuk, E.D.; Le, Q. Rethinking pre-training and self-training. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Dutchess County, NY, USA, 2020; Volume 33, pp. 3833–3845.
33. Kataoka, H.; Okayasu, K.; Matsumoto, A.; Yamagata, E.; Yamada, R.; Inoue, N.; Kimura, A.; Satoh, Y. Pre-training without natural images. In Proceedings of the Asian Conference on Computer Vision (ACCV), Kyoto, Japan, 30 November–4 December 2020; pp. 583–600. [\[CrossRef\]](#)
34. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [\[CrossRef\]](#)
35. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717. [\[CrossRef\]](#)
36. Shen, Z.; Bello, I.; Vemulapalli, R.; Jia, X.; Chen, C.H. Global self-attention networks for image recognition. *arXiv* **2020**, arXiv:2010.03019. [\[CrossRef\]](#)
37. Tavazoei, M.S. From traditional to fractional PI control: A key for generalization. *IEEE Ind. Electron. Mag.* **2012**, *6*, 41–51. [\[CrossRef\]](#)
38. Lim, J.; Ye, J.C. Geometric GAN. *arXiv* **2017**, arXiv:1705.02894. [\[CrossRef\]](#)
39. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [\[CrossRef\]](#)
40. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9584–9592. [\[CrossRef\]](#)
41. Liu, H.; Zhang, W.; Li, B.; Wu, H.; He, N.; Huang, Y.; Li, Y.; Ghanem, B.; Zheng, Y. AdaptiveMix: Improving GAN training via feature space shrinkage. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 16219–16229. [\[CrossRef\]](#)
42. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Langs, G.; Schmidt-Erfurth, U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Med. Image Anal.* **2019**, *54*, 30–44. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [\[CrossRef\]](#)
44. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Bergmann, P.; Löwe, S.; Fauser, M.; Sattlegger, D.; Steger, C. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP), Prague, Czech Republic, 25–27 February 2019.
46. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.

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.