

GASR-Net: Geometric Anomaly Synthesis and Task-Oriented Feature Reconstruction for Industrial Anomaly Detection

Luhao Li¹, Xiaoyang Shi², and Shuqiang Gao^{1(✉)}

School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China
School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China

ielhli@163.com

Abstract. In real-world industrial manufacturing, automated visual inspection for surface defect detection is crucial for quality control but remains severely hindered by two critical challenges: the simplified assumption of isotropic anomaly synthesis that fails to capture complex physical defect shapes, and the over-smoothing effect in generative models that triggers unacceptable false alarms in textured backgrounds. To address these persistent industrial limitations, we propose a novel network, GASR-Net (Geometric Anomaly Synthesis and Task-Oriented Feature Reconstruction for unsupervised image anomaly detection and localization). Our architecture tackles the gap between synthesized noise and realistic flaw shapes by introducing a Geometric-Aware Deformable Noise Modulator (GAND-M). By utilizing baseline normal features to dynamically predict deformation offsets, the module warps isotropic Gaussian noise into structurally contiguous and geometric-diverse pseudo-anomaly flows. Simultaneously, to overcome the persistent issue of background over-smoothing and the resulting high false positive rates, we design a Task-Oriented Single-Step Feature Reconstructor based on a lightweight U-Net topology. Unlike conventional generative methods that perform blind standalone restoration, our constructor is driven by a joint feedback loop from the downstream discriminator, adaptively forcing the network to thoroughly eliminate abnormal artifacts while strictly preserving pixel-level background fidelity. Extensive experiments on two challenging industrial anomaly detection benchmarks, MVTec AD and VisA, demonstrate that GASR-Net achieves state-of-the-art performance in both detection and localization accuracy.

Keywords: Unsupervised Anomaly Detection, Geometric Anomaly Synthesis, Deformable Noise Modulation, Feature Reconstruction, Task-Oriented Feedback Loop

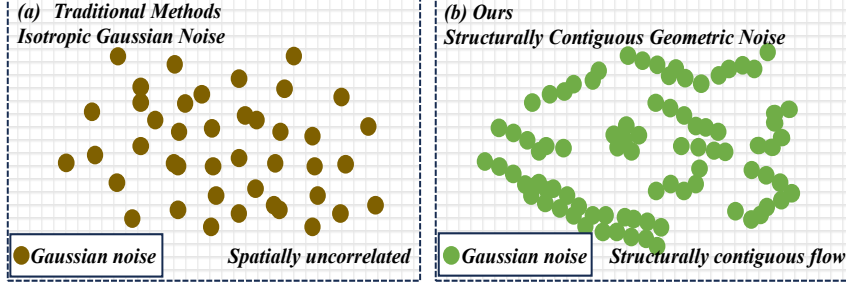


Fig. 1. Conceptual comparison of pseudo-anomaly synthesis paradigms in the feature space. (a) Traditional Methods: Conventional discriminative frameworks inject uncorrelated, isotropic Gaussian noise that is uniformly dispersed and spatially uncorrelated, failing to capture the physical topologies of real-world defects. (b) Ours: By adaptively modulating the raw noise with predicted spatial-context offsets and a dynamic spatial mask, our modulator successfully synthesizes a structurally contiguous flow that naturally mimics the geometric characteristics of realistic industrial flows.

1 Introduction

In real-world industrial manufacturing [23,26], guaranteeing product quality and surface integrity is paramount for maintaining production efficiency and operational safety. Automated visual inspection systems have emerged as a pivotal technology for substituting manual labor in identifying and localizing surface anomalies, such as cracks, scratches, and structural deformities. However, in practical industrial scenarios, collecting a comprehensive dataset containing all potential defective patterns is severely restricted by the rare occurrence of flaws and the unpredictable nature of unexpected anomalies. Consequently, recent research focuses predominantly on unsupervised anomaly detection (UAD) [3], where models are trained exclusively on anomaly-free samples to establish a baseline for normal appearance.

To achieve robust anomaly isolation, current unsupervised paradigms can be broadly categorized into three dominant methodologies: generative methods, reconstruction-based methods, and feature-based discriminative methods. Generative methods typically utilize frameworks like Generative Adversarial Networks (GANs) [24], Variational Autoencoders (VAEs) [18], or Normalizing Flows [0] to model the exact probability density function of normal samples in a latent space, identifying anomalies as statistical outliers with low probability densities. Reconstruction-based methods operate under the assumption that a network trained solely on normal data cannot successfully reconstruct anomalous regions. They pass the input through a bottleneck layer (such as an Autoencoder) [2,12] and measure the pixel-level or patch-level pixel discrepancies between the original image and its anomaly-free reconstruction. Feature-based discriminative methods [20,27], on the other hand, circumvent the complex image generation process by shifting the task into a pre-trained feature space [4,6]. They artificially inject pre-defined noise into normal feature maps to simulate proxy

anomalies [27], thereby transforming the unsupervised problem into a standard, highly efficient self-supervised binary classification task.

Despite their outstanding contributions, each of these prevailing paradigms suffers from intrinsic architectural limitations when applied to precision industrial inspection. To begin with, generative approaches often require unstable adversarial training or computationally intensive bi-directional mapping, making them brittle when handling complex structural variations. Concurrently, traditional reconstruction-based models, including recent advanced single-step denoising diffusion models [17, 28], inevitably face a severe trade-off between defect erasure and texture preservation. In order to thoroughly eliminate large or heterogeneous flaws, these architectures often induce severe over-smoothing effects on non-defective, highly-textured backgrounds, which inadvertently introduces unacceptable rates of false alarms. Furthermore, while existing feature-based discriminative methods achieve remarkable inference speed, their reliance on isotropic Gaussian noise synthesis fails to simulate structurally contiguous, geometric-diverse fluid manifolds such as micro-scratches or line cracks inherently found in realistic physical defects. As visualized in Fig. 1(a), this core discrepancy severely restricts the discriminator’s capability to generalize to finegrained local flaws due to the lack of realistic spatial priors.

To break these architectural bottlenecks and unify the merits of both worlds, we propose a novel network, termed GASR-Net. Instead of executing standalone, unguided operations, our network establishes an elegant, loop-cooperative mechanism directly within the low-dimensional feature space. To eliminate the gap between idealized noise and realistic geometric anomalies, we first design the Geometric-Aware Deformable Noise Modulator (GAND-M). As conceptually contrasted in Fig. 1, conventional feature-space methods typically inject uncorrelated, isotropic Gaussian noise that is uniformly dispersed across the entire feature space, leading to a severe discrepancy with realistic flaw topologies as illustrated in Fig. 1(a). In sharp contrast, our proposed GAND-M utilizes multiscale dilated convolutions and an offset predictor to dynamically predict topological displacements from the baseline features. This architecture warp-shapes raw Gaussian noise into structurally realistic and geometric-diverse pseudo-anomaly flows bounded by a self-generated spatial mask, which successfully simulates the intricate geometric patterns shown in Fig. 1(b) and thereby forces the downstream discriminator to learn highly robust, fine-grained decision boundaries. Simultaneously, to overcome the persistent issue of false alarms triggered by background over-smoothing, we introduce a Task-Oriented Single-Step Feature Reconstructor. Configured with a lightweight U-Net topology, this reconstructor purifies the heterogeneous anomaly manifolds through a single forward pass. Crucially, the reconstruction behavior is explicitly driven by a joint feedback loop from the downstream discriminator. Guided by a dedicated spatial fidelity constraint, it adaptively forces the network to thoroughly wash out abnormal artifacts within designated defective regions while strictly preserving pixel-level background fidelity elsewhere.

The main contributions of this work are summarized as follows:

- We propose GASR-Net, a loop-cooperative feature-space paradigm for unsupervised anomaly detection, which seamlessly bridges the high efficiency of discriminative networks with the robust prior modeling of generative networks.
- We propose the Geometric-Aware Deformable Noise Modulator (GAND-M), which utilizes deformable convolutions and spatial-context offsets to synthesize pseudo-anomalies with realistic physical shapes and topological manifolds.
- We introduce a Task-Oriented Single-Step Feature Reconstructor driven by joint downstream discriminator guidance, which effectively suppresses background over-smoothing and eliminates false alarms without introducing iterative inference latency.
- Extensive experiments on two public industrial anomaly detection benchmarks, MVTec AD and VisA, demonstrate that our proposed GASR-Net achieves state-of-the-art performance in both anomaly detection and localization accuracy.

2 Related Work

In the section, we review recent advances in unsupervised image anomaly detection and localization, which are primarily trained on anomaly-free samples to isolate unexpected patterns. According to their core technological mechanics, these paradigms can be broadly clustered into three dominant categories: generative methods, reconstruction-based methods, and feature-based discriminative methods.

2.1 Generative Methods

Generative frameworks primarily leverage probabilistic density estimation or mathematical mapping to model the exact statistical distribution of normal industrial samples. Early works typically employ GANs or VAEs to establish an implicit manifold boundary of normal semantics. Under this regime, any input instance that falls into low-probability regions of the learned latent space is mathematically categorized as an outlier. More recently, Normalizing Flows (NF) have garnered widespread attention due to their capability to perform exact and tractable likelihood estimation via invertible bi-directional mappings. For instance, CFLOW-AD [13] incorporates conditional normalizing flows paired with positional encodings to capture multi-scale spatial distributions. Similarly, CS-Flow [17] utilizes a fully convolutional flow architecture to process multi-scale feature maps simultaneously, thus capturing global and local cross-scale dependencies. Despite their solid theoretical foundations in probability estimation, generative methods often suffer from severe limitations in precision industrial inspection. Adversarial training in GANs is inherently unstable and prone to mode collapse, while Normalizing Flows incur heavy computational overhead and require bidirectional structural constraints, making them extremely brittle when handling highly intricate background texture variations.

2.2 Reconstruction-based Methods

Reconstruction-based paradigms operate under the direct assumption that a network trained exclusively on normal data possesses a generalized prior for anomaly-free semantics, and consequently, will fail to faithfully reconstruct abnormal artifacts. At inference, anomalies are localized by evaluating the pixel-level or feature-level discrepancies between the input image and its corresponding normal reconstruction. Classic Autoencoders (AEs) and U-Net variants serve as the foundational backbones for this category. To prohibit the network from directly copying anomalous pixels through skip connections, modern methods incorporate denoising diffusion models [29] or memory-augmented banks [12] to restrict latent capacities. For instance, DDAD [7] introduces conditioned or norm-guided denoising processes to execute single-step reconstruction, aiming to rapidly synthesize anomaly-free references. However, a persistent dilemma for traditional reconstruction-based models is the severe and intractable trade-off between defect erasure and texture preservation. If the network’s bottleneck capacity is too high, it suffers from an undesired generalization capability that inadvertently reconstructs the anomalous regions perfectly. Conversely, if the capacity is overly restricted to eliminate large or heterogeneous flaws, the model suffers from severe over-smoothing effects on non-defective, highly-textured backgrounds. This degradation of spatial context fidelity fundamentally leads to a proliferation of false alarms in real-world industrial inspection scenarios.

2.3 Feature-based Discriminative Methods

To circumvent the computationally intensive and fragile image-generation process, feature-based discriminative methods shift the anomaly detection task directly into a pre-trained feature space (e.g., via a frozen ResNet [9,16] or ViT backbone [10]). This category can be divided into embedding-matching methods and self-supervised synthetic methods. Embedding-matching models, such as PatchCore [25], establish a rigorous memory bank of nominal feature representations based on nearest neighbor pattern classification [5] and utilize localized nearest-neighbor search for anomaly segmentation. Alternatively, self-supervised discriminative methods transform the unsupervised task into a highly efficient binary classification problem. They artificially inject pre-defined noise into normal feature maps to serve as negative proxy samples, thereby training a tight decision boundary. For instance, SimpleNet [20] adds direct Gaussian noise to normal features to train a lightweight multi-layer perceptron (MLP) discriminator. Although feature-based discriminative paradigms exhibit remarkable inference speeds and high image-level AUROC scores, their synthesis schemes remain

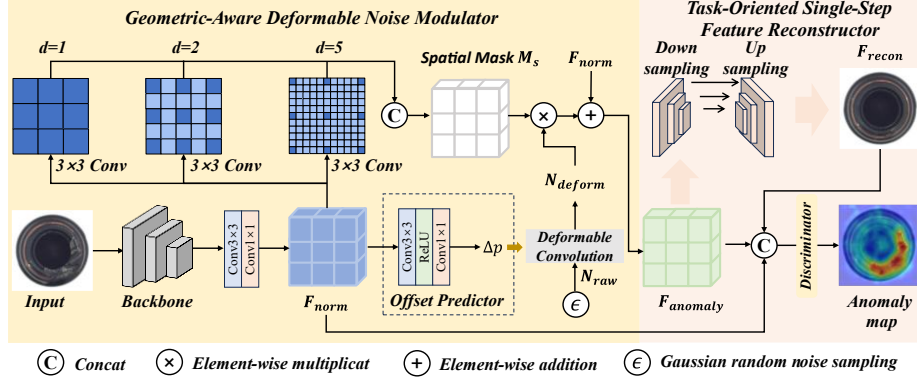


Fig. 2. The overall architectural pipeline of the proposed GASR-Net. Hierarchical features extracted from the frozen backbone are first projected into a unified embedding F_{norm} . The GAND-M warps raw Gaussian noise into structural pseudo-anomaly features $F_{anomaly}$ via predicted spatial offsets Δp and a spatial mask M_s . Concurrently, the Task-Oriented Single-Step Feature Reconstructor purges abnormal artifacts through a single forward pass to generate F_{recon} . Finally, the combined multi-channel features are fed into the downstream Discriminator for joint classification and fidelity optimization.

highly idealized. Existing methods heavily rely on simple isotropic Gaussian noise injection. This pixel-independent, uncorrelated perturbation strategy completely fails to simulate the structurally contiguous, geometric-diverse fluid manifolds (such as line scratches, micro-cracks, or continuous structural deformations) inherently found in physical industrial defects. Consequently, the downstream discriminator is inherently bounded by loose and isotropic decision boundaries, leaving it sub-optimal when generalizing to fine-grained local flaws.

3 Method

The comprehensive architectural workflow of our proposed network is schematically illustrated in Fig. 2. In the following subsections, we introduce the exact mathematical formulation of the geometric synthesis phase, the task-oriented single-step characteristic reconstruction process, and the joint loss optimization strategy in a progressive manner.

3.1 Overall Architecture

Given a query industrial image, our framework aims to concurrently predict an image-level anomaly score and a pixel-level anomaly localization map. To establish a robust baseline descriptor capable of distilling intricate background textures, we employ a frozen DINOv2 [22] backbone as our foundational multiscale feature extractor. Specifically, we extract hierarchical patch tokens from the 6-th, 12-th, and 18-th intermediate

Transformer blocks of the frozen backbone, capturing progressive feature granularities ranging from local micro-textures to global semantic contexts.

To resolve the spatial and dimensional misalignments across these extracted multi-layer representations, the raw feature maps are first spatially aligned via bilinear interpolation and combined along the channel dimension via a concatenation operator. Subsequently, the concatenated representations are passed through a 3×3 convolutional layer followed by a 1×1 convolutional layer to compress the channel capacity, projecting them directly into the unified baseline normal feature map F_{norm} . This unpolluted embedding serves as the nominal anchor reference for the entire network. As schematically illustrated in Fig. 2, F_{norm} is concurrently streamed into two parallel cooperative tracks: the upper anomaly synthesis path controlled by the Geometric-Aware Deformable Noise Modulator (GAND-M) to formulate realistic negative proxy samples, and the downstream discrimination module to establish fine-grained decision boundaries.

3.2 Geometric-Aware Deformable Noise Modulator (GAND-M)

To eliminate the structural gap between idealized statistical perturbations and realistic physical defects, the upper cooperative loop of GASR-Net deploys the GAND-M block. Driven by the contextual guidance of the baseline feature map, this module is specifically engineered to warp raw isotropic Gaussian noise into structurally contiguous, geometric-diverse pseudo-anomaly flows.

As schematically diagrammed in Fig. 2, the GAND-M component concurrently branches into a multi-scale spatial perception track and a geometric coordinate derivation track. The multi-scale perception track takes the baseline normal feature map F_{norm} as input and utilizes three parallel convolutional layers configured with distinct dilation rates $d \in \{1, 2, 5\}$ and a kernel size of 3×3 . This multi-resolution setup adaptively captures both highly localized micro-textures and long-range contextual environments. To aggregate these multi-scale visual cues, the parallel output feature maps are directly concatenated along the channel dimension. A subsequent convolutional projection layer paired with a Sigmoid activation function is applied to compact the fused representation, dynamically generating a continuous spatial mask M_s that strictly bounds the region of interest for potential defect simulation.

Concurrently, the geometric coordinate derivation track passes the nominal anchor F_{norm} through an Offset Predictor, which comprises a 3×3 convolutional layer, a $ReLU$ activation, and a 1×1 convolutional layer, to dynamically estimate dense topological displacement fields denoted as Δp . Symmetrically, a raw standard Gaussian noise map $N_{raw} \in R^{C \times H \times W}$ is sampled from a standard normal distribution $N_{raw} \sim \mathcal{N}(0, \mathbf{I})$. Crucially, to inject structural coherence and physical shapes into the randomized perturbations, a Deformable Convolutional layer takes the raw noise N_{raw} as its direct input, executing precise coordinate grid shifts under the explicit steering of the predicted offset field Δp . For each grid position p_0 on the feature map, the deformable modulation generating the distorted noise representation N_{deform} is mathematically defined as:

$$N_{deform}(p_0) = \sum_{n=1}^N w_n \cdot N_{raw}(p_0 + p_n + \Delta p_n(p_0)), \quad (1)$$

where p_n enumerates the pre-defined regular grid coordinates of the 3×3 deformable kernel containing $N = 9$ sampling points, and w_n represents the learnable convolutional weights. Bilinear interpolation is implemented to sample the exact values from the fractional coordinates generated by the non-integer displacement $\Delta p_n(p_0)$.

Finally, the structurally distorted noise feature N_{deform} is gated by the self-generated spatial mask M_s , and the modulated pseudo-anomaly flow is subsequently injected back into the baseline normal embedding F_{norm} to construct the comprehensive synthesized anomaly feature map $F_{anomaly} \in R^{C \times H \times W}$:

$$F_{anomaly} = F_{norm} \oplus (N_{deform} \otimes M_s), \quad (2)$$

where $\otimes$ and $\oplus$ denote the element-wise multiplication and element-wise addition operators. By Title-guided steering and explicitly driving the topological deformation from the contextual cues of F_{norm} , the resulting negative proxy representation $F_{anomaly}$ exhibits natural boundaries and realistic physical structures, thereby providing an optimized decision margin for the downstream discriminator.

3.3 Task-Oriented Single-Step Feature Reconstructor

To conquer the chronic background over-smoothing dilemma prevalent in conventional generative and reconstruction models, the lower loop of GASR-Net incorporates a Task-Oriented Single-Step Feature Reconstructor. Instead of executing an unguided, blind pixel-level restoration, this reconstructor operates directly within the low-dimensional feature space using a lightweight, symmetrical U-Net topology. As depicted in the right bounding box of Fig. 2, the Reconstructor takes the synthesized pseudo-anomaly representation $F_{anomaly}$ as its vertical input. Specifically, the network architecture is built upon a hierarchical encoder-decoder framework, where the encoder path sequentially compresses the spatial dimensions while expanding the channel width through down-sampling layers to extract compact, high-level nominal semantics. Symmetrically, the decoder path implements progressive up-sampling layers to restore the spatial resolution step-by-step. To mitigate the loss of spatial fidelity during bottleneck transitions, horizontal skip connections are utilized to transfer the uncorrupted structural details directly from the encoder stages, concatenating them with the corresponding up-sampled decoder feature maps.

Crucially, our reconstruction mechanism is executed via a single forward pass without introducing any iterative inference latency (such as the multi-step reverse sampling found in diffusion models). The output of the final decoder layer yields the purified, anomaly-free reconstructed representation $F_{recon} \in R^{C \times H \times W}$:

$$F_{recon} = \mathcal{R}(F_{anomaly}), \quad (3)$$

Table 1. Quantitative Results on MVTec AD. Comparison of I-AUROC, P-AUROC, P-PRO, and P-AP. All metrics are in %. Best results are bold.

Method	I-AUROC	P-AUROC	P-PRO	P-AP
UniAD (NIPS 22)	96.5	96.8	90.7	43.4
PatchCore (CVPR 22)	99.1	98.1	93.5	56.1
RD4AD (CVPR 22)	98.5	97.8	93.9	58.0
DeSTSeg (CVPR 23)	89.2	93.1	64.8	54.3
DMAD (AAAI 23)	99.6	98.3	92.9	49.6
SimpleNet (CVPR 23)	99.5	97.6	90.7	54.8
DiAD (AAAI 24)	97.2	96.8	90.7	52.6
MambaAD (NIPS 24)	98.6	97.7	93.1	56.3
OmiAD (ICML 25)	98.8	97.7	93.2	52.6
Ours	99.7	98.3	94.1	63.6

where $\mathcal{R}(\cdot)$ denotes the mapping function of the single-step reconstructor. Rather than acting as a standalone module, the reconstruction behavior is strictly optimized via a joint feedback loop under the direct constraint of a downstream discriminator, forcing the network to selectively wash out abnormal artifacts within anomalous manifolds while maintaining strict pixel-level fidelity for normal textured backgrounds.

3.4 Discriminator and Joint Loss Optimization

The rightmost stage of GASR-Net establishes a multi-channel joint discrimination scheme. To accurately isolate real defects from pseudo-anomalies and reconstruction artifacts, we pass three distinct feature flows into the system: the nominal baseline F_{norm} , the synthesized anomaly $F_{anomaly}$, and the reconstructed representation F_{recon} .

The three parallel feature maps are first aligned and concatenated along the channel dimension. The channel-fused tensor is subsequently fed into a multilayer convolutional Discriminator $\mathcal{D}(\cdot)$, which scores each local patch position to generate a continuous anomaly map. The learning process of the entire network is jointly optimized via a combined objective function consisting of a discriminative classification loss $\mathcal{L}_{disc}$ and a spatial feature reconstruction fidelity loss $\mathcal{L}_{fid}$:

$$\mathcal{L}_{total} = \mathcal{L}_{disc} + \lambda \cdot \mathcal{L}_{fid}, \quad (4)$$

where λ represents a trade-off hyperparameter balancing the two terms.

The discriminative classification loss $\mathcal{L}_{disc}$ is formulated via a standard binary cross-entropy (BCE) objective, forcing the discriminator to output a nominal score of 0 for clean features and an anomaly score of 1 for synthesized defective areas:

Table 2. Quantitative Results on VisA. Comparison of I-AUROC, P-AUROC, P-PRO, and P-AP. All metrics are in %. Best results are bold.

Method	I-AUROC	P-AUROC	P-PRO	P-AP
UniAD (NIPS 22)	88.8	98.3	85.5	33.7
PatchCore (CVPR 22)	95.1	98.8	91.2	40.1
RD4AD (CVPR 22)	96.0	90.1	70.9	27.7
DeSTSeg (CVPR 23)	88.9	96.1	67.4	39.6
DMAD (AAAI 23)	95.5	98.6	91.3	41.0
SimpleNet (CVPR 23)	96.8	97.8	88.7	36.3
DiAD (AAAI 24)	86.8	96.0	75.2	26.1
MambaAD (NIPS 24)	94.3	98.5	91.0	39.4
OmiAD (ICML 25)	95.3	98.9	89.2	40.4
Ours	97.2	99.0	92.3	48.3

$$\mathcal{L}_{disc} = -\sum_{h,w} \left[\log(1 - \mathcal{D}(F_{norm})^{(h,w)}) + \log(\mathcal{D}(F_{anomaly})^{(h,w)}) \cdot M_s^{(h,w)} \right], \quad (5)$$

where (h, w) iterates over the spatial coordinates, and the classification penalty for $F_{anomaly}$ is adaptively weighted by the spatial mask M_s to ensure optimization is concentrated precisely within the synthesized flaw topologies.

Concurrently, to eliminate false alarms caused by background texture blurring, the spatial feature reconstruction fidelity loss $\mathcal{L}_{fid}$ penalizes any feature discrepancy between the nominal anchor F_{norm} and the reconstructed feature F_{recon} using a $\mathcal{L}_1$ loss function:

$$\mathcal{L}_{fid} = \frac{1}{C \times H \times W} \sum_{c,h,w} \mathcal{H}(F_{norm}^{(c,h,w)} - F_{recon}^{(c,h,w)}), \quad (4)$$

where $\mathcal{H}(\cdot)$ denotes the Huber loss function, which limits gradient explosions during the early training phase and guarantees pixel-level semantic preservation for normal structures.

During testing, the single-step reconstructor and discriminator cooperate seamlessly. The final pixel-level anomaly localization map M is directly computed by evaluating the discriminator’s response to the concatenated test representation, bypassing any hand-crafted post-processing pipelines.

Table 3. Component Effectiveness. Effect of GAND-M and Reconstructor on MVTec AD. All metrics are in %.

GAND-M	Reconstructor	I-AUROC	P-AUROC	P-PRO	P-AP
✓		99.5	98.1	93.8	63.1
	✓	99.6	97.4	93.7	62.5
✓	✓	99.7	98.3	94.1	63.6

4 Experiments

In the section, we conduct extensive experiments on two widely used industrial anomaly detection benchmarks to evaluate the performance of our proposed GASR-Net. We present the detailed experimental setup, followed by a comprehensive quantitative comparison with state-of-the-art (SOTA) methodologies. Finally, we provide in-depth ablation studies and qualitative visualizations to dissect the core contributions of each individual component.

4.1 Experimental Setup

Datasets. To validate the generalization and robustness of our method, we evaluate our paradigm on two challenging public benchmarks, namely MVTec AD [1] and VisA [33]. The MVTec AD dataset is a standard benchmark representing realistic surface inspection, comprising 15 distinct categories with 5 texture types and 10 object types. It provides 3,629 anomaly-free images for training and 1,725 images for testing, along with pixel-level ground truth annotations for various realistic flaws such as scratches, cracks, and structural deformations. Symmetrically, the VisA dataset serves as a more rigorous and complex industrial benchmark containing 12 categories. It consists of 10,821 high-resolution images characterized by challenging multi-instance layouts, extremely subtle flaws, and severe spatial misalignments.

Evaluation Metrics. Following the standard evaluation protocols in unsupervised anomaly detection, we adopt four complementary metrics to measure accuracy at different granularities. For image-level anomaly detection, the Area Under the Receiver Operating Characteristic curve (I-AUROC) is calculated. For pixel-level anomaly localization, we employ standard pixel-level AUROC (P-AUROC), Per-Region Overlap (P-PRO), and Pixel-level Average Precision (P-AP). Crucially, P-PRO and P-AP are heavily preferred for precision physical inspection, as they penalize false positives more severely in complex textures and prevent overly optimistic evaluations on micro-defects. All metrics are reported in percentages (%).

Implementation Details. Our proposed GASR-Net is implemented using the PyTorch framework and trained on a single NVIDIA GeForce RTX 4090 GPU. Images from both benchmarks are resized to 518×518 pixels for efficient batch processing. We utilize the frozen DINOv2-Large as our backbone feature extractor, capturing

hierarchical characteristics from the 6-th, 12-th, and 18-th intermediate Transformer blocks. For the GAND-M module, the parallel dilated convolutions are configured with dilation rates $d \in \{1, 2, 5\}$ and a kernel size of 3×3 . During training, the entire network is optimized for 50 epochs using the Adam optimizer [21] with a constant learning rate of 2×10^{-4} and a batch size of 8. The balancing trade-off hyperparameter λ in the joint loss objective is set to 1.0.

4.2 Quantitative Results

We benchmark GASR-Net against various state-of-the-art methods. These include classic paradigms such as UniAD [30], PatchCore [25], RD4AD [8], DeST-Seg [32], and DMAD [19], SimpleNet [20], as well as the most recent advanced approaches including DiAD [15], MambaAD [14], and OmiAD [11].

Evaluation on MVTec AD. As reported in Table 1, our GASR-Net demonstrates superior capability across all critical metrics on the MVTec AD benchmark. Specifically, our model achieves a state-of-the-art I-AUROC of 99.7% and a P-AUROC of 98.3%. More importantly, GASR-Net yields a significant performance gain on localization-sensitive metrics, delivering a P-PRO of 94.1% and an absolute boost in P-AP to 63.6% (outperforming the second-best method, RD4AD, by an impressive margin of 5.6%). This remarkable leap in P-AP strongly demonstrates that by modulating raw perturbations into structurally contiguous flows via our GAND-M, the model forces the downstream discriminator to establish tight, fine-grained decision boundaries for realistic flaws rather than loose isotropic margins.

Evaluation on VisA. The quantitative comparisons on the highly challenging VisA dataset are summarized in Table 2. Thanks to the severe multiinstance interference and complex spatial variations inherent in VisA, conventional models experience substantial performance degradation. In contrast, our proposed GASR-Net showcases excellent robustness, dominating the leaderboard with 97.2% I-AUROC, 99.0% P-AUROC, 92.3% P-PRO, and a groundbreaking 48.3% P-AP. Crucially, while previous single-step reconstruction models (such as DiAD) suffer from background texture blurring that causes high false positive rates, our Task-Oriented Feature Reconstructor successfully mitigates oversmoothing. Driven by the joint discriminator feedback loop and the strict spatial fidelity constraint, it adaptively purifies defect artifacts within anomalous regions while perfectly preserving background texturing, thereby suppressing false alarms and maintaining precise pixel-level localization.

4.3 Ablation Study

To verify the structural necessity and mutual synergies of our core innovations, we perform comprehensive ablation experiments on the MVTec AD benchmark, with quantitative tracking systematically summarized in Table 3. As evidenced by the empirical statistics, integrating only the GAND-M component delivers a compelling performance baseline with a prominent P-AP score of 63.1%, which explicitly validates

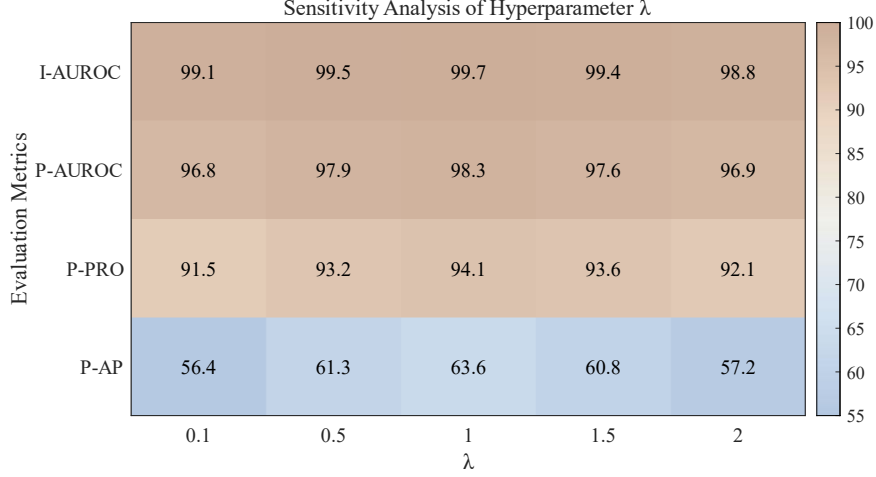


Fig. 3. Hyperparameter sensitivity analysis of the joint loss weight λ on MVTec AD. The localization precision metrics display a distinctive parabolic curve that stabilizes and achieves the optimal performance trade-off precisely at $\lambda = 1$. All metrics are in %.

that warp-shaping raw isotropic noise into structurally contiguous pseudo-anomaly manifolds successfully captures the physical topologies of real-world flaws and sharpens the discriminator’s boundaries. Symmetrically, utilizing the standalone Reconstructor achieves a competitive I-AUROC of 99.6% and a P-PRO of 93.7%, proving that the joint loop-cooperative mechanism and spatial fidelity loss successfully suppress background over-smoothing and minimize false alarms. Ultimately, when both components are simultaneously integrated, our full GASR-Net reaches its peak optimization state, establishing state-of-the-art results across all evaluation axes (99.7% I-AUROC, 98.3% P-AUROC, 94.1% P-PRO, and 63.6% P-AP). This global reciprocity firmly confirms that the organic combination of geometric anomaly synthesis and task-oriented reconstruction provides the most robust and precise closed loop for complex industrial visual inspection.

Sensitivity of Loss Weight. λ . To evaluate the constraint equilibrium between anomaly erasure and feature context preservation, we trace the localization tracking by dynamically varying the balancing hyperparameter $\lambda \in \{0.1, 0.5, 1, 1.5, 2\}$. As visually illustrated in the line trajectories of Fig. 3, the performance metrics strictly adhere to a rigorous parabolic curve peaking at the default setting of $\lambda = 1$ (99.7% I-AUROC, 98.3% P-AUROC, 94.1% P-PRO, and 63.6% P-AP). Specifically, a lower coefficient (e.g., $\lambda = 0.1$) relaxes the spatial fidelity constraint, causing the single-step feature reconstructor to induce over-smoothing artifacts on complex textured backgrounds. This degradation introduces excessive false positives, dropping the P-PRO to 91.5% and P-AP to 56.4%. Conversely, setting an excessively high weight (e.g., $\lambda = 2$) forces the reconstruction fidelity term to dominate the optimization gradients, thereby

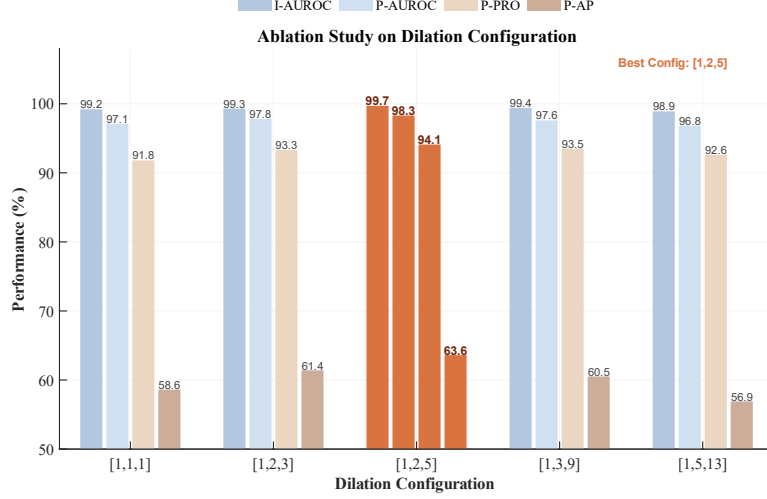


Fig. 4. Ablation study on multi-scale dilation rate configurations within the GANDM block on MVTec AD. The evaluation histograms demonstrate that our baseline setting of [1, 2, 5] yields the highest structural alignment with realistic anomalies. All metrics are in %. suppressing the discriminator’s capability to converge on tight decision boundaries for structural anomalies. Consequently, the standard assignment of $\lambda = 1$ guarantees the optimal mathematical Nash equilibrium for joint training.

Configuration of Dilation Rates. Furthermore, we analyze the multiscale spatial receptive field for geometric anomaly modeling by benchmarking five progressive combinations of dilation rates within the GAND-M block. As demonstrated in the evaluation histograms of Fig. 4, the anomaly localization accuracy exhibits high sensitivity to the pixel-sampling span. Utilizing a standard non-dilated sequence ([1, 1, 1]) constrains spatial perception to localized patches, failing to capture long-range continuous defect topologies and resulting in a suboptimal P-AP of 58.6%. Conversely, aggressively scaling up to over dilated schemes (e.g., [1, 5, 13]) severe the spatial correlations between adjacent features due to excessive sampling sparsity, introducing severe gridding effects and background clutter. In sharp contrast, our designated [1, 2, 5] configuration establishes the optimal golden ratio, seamlessly capturing local micro-deformities via smaller gaps and structurally elongated fluid cracks via expanded spans, thereby achieving the highest alignment with authentic industrial anomalies.

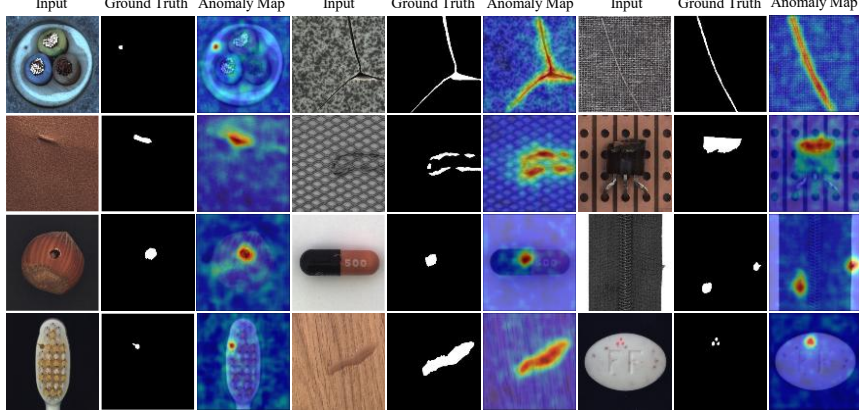


Fig. 5. Qualitative anomaly localization results of the proposed GASR-Net on MVTec AD.

4.4 Qualitative Visualizations

Qualitative Inspection Analysis. To visually demonstrate the high-fidelity anomaly segmentation and precise boundary tracking capabilities of our framework, we illustrate the qualitative localized heatmaps generated by GASR-Net across various challenging structural and textural categories in Fig. 5. As showcased in the localized distributions, our model handles an expansive diversity of industrial anomaly typologies, spanning from high-contrast micro-defects to extremely subtle structural dislocations. Crucially, despite the presence of complex structural backdrops, repetitive patterns, and severe illumination variations, the predicted anomaly maps exhibit an exceptional structural alignment with the manual ground truth annotations. The generated heatmaps are tightly focused on the authentic anomalous manifolds without spilling over into pristine, unpolluted background texturing. This highly discriminative edge delineation strongly confirms that by warp-shaping statistical perturbations into realistic fluid contours via GAND-M and concurrently penalizing over-smoothing via the joint feature reconstructor feedback loop, GASR-Net successfully eliminates background clutter and false alarms, guaranteeing robust and millimeter-level localization precision for practical quality control.

5 Conclusion

In this paper, we presented GASR-Net, a novel loop-cooperative feature-space framework for unsupervised industrial anomaly detection and localization. To eliminate the structural gap between idealized noise and realistic geometric defects, we proposed the Geometric-Aware Deformable Noise Modulator (GAND-M) to warp raw Gaussian noise into structurally contiguous pseudo-anomaly flows. Simultaneously, to resolve the background over-smoothing dilemma prevalent in traditional reconstruction models, a Task-Oriented Single-Step Feature Reconstructor is introduced to purge abnormal artifacts while strictly preserving pixel-level background fidelity under joint

downstream discriminator constraints. Extensive experiments on the MVTec AD and VisA benchmarks demonstrate that GASR-Net successfully achieves state-of-the-art performance across all critical metrics, particularly establishing a prominent superiority in localization precision. Future directions will focus on extending this paradigm to multimodal zero-shot scenarios and optimizing the model architecture for resourceconstrained embedded edge devices.

6 References

1. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 9592–9600 (2019)
2. Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.: Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011* (2018)
3. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. *ACM computing surveys (CSUR)* 41(3), 1–58 (2009)
4. Cohen, N., Hoshen, Y.: Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357* (2020)
5. Cover, T., Hart, P.: Nearest neighbor pattern classification. *IEEE transactions on information theory* 13(1), 21–27 (1967)
6. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: *International conference on pattern recognition*. pp. 475–489. Springer (2021)
7. Dehaene, D., Frigo, O., Combrexelle, S., Eline, P.: Iterative energy-based projection on a normal data manifold for anomaly localization. *arXiv preprint arXiv:2002.03734* (2020)
8. 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*. pp. 9737–9746 (2022)
9. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A largescale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. pp. 248–255. Ieee (2009)
10. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020)
11. 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. In: *Forty-second International Conference on Machine Learning* (2025)
12. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 1705–1714 (2019)
13. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 98–107 (2022)
14. He, H., Bai, Y., Zhang, J., He, Q., Chen, H., Gan, Z., Wang, C., Li, X., Tian, G., Xie, L.: Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *Advances in Neural Information Processing Systems* 37, 71162–71187 (2024)

15. He, H., Zhang, J., Chen, H., Chen, X., Li, Z., Chen, X., Wang, Y., Wang, C., Xie, L.: A diffusion-based framework for multi-class anomaly detection. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 8472–8480 (2024)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
17. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33, 6840–6851 (2020)
18. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013)
19. Liu, W., Chang, H., Ma, B., Shan, S., Chen, X.: Diversity-measurable anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12147–12156 (2023)
20. 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. pp. 20402–20411 (2023)
21. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017)
22. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023)
23. Pang, G., Shen, C., Cao, L., Hengel, A.V.D.: Deep learning for anomaly detection: A review. *ACM computing surveys (CSUR)* 54(2), 1–38 (2021)
24. Perera, P., Nallapati, R., Xiang, B.: Ocgan: One-class novelty detection using gans with constrained latent representations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2898–2906 (2019)
25. 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. pp. 14318–14328 (2022)
26. Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., Dietterich, T.G., Müller, K.R.: A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE* 109(5), 756–795 (2021)
27. Schlüter, H.M., Tan, J., Hou, B., Kainz, B.: Natural synthetic anomalies for selfsupervised anomaly detection and localization. In: *European Conference on Computer Vision*. pp. 474–489. Springer (2022)
28. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020)
29. Song, Y., Ermon, S.: Improved techniques for training score-based generative models. *Advances in neural information processing systems* 33, 12438–12448 (2020)
30. You, Z., Cui, L., Shen, Y., Yang, K., Lu, X., Zheng, Y., Le, X.: A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems* 35, 4571–4584 (2022)
31. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677* (2021)
32. Zhang, X., Li, S., Li, X., Huang, P., Shan, J., Chen, T.: Destseg: Segmentation guided denoising student-teacher for anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3914–3923 (2023)

33. Zou, Y., Jeong, J., Pemula, L., Zhang, D., Dabeer, O.: Spot-the-difference selfsupervised pre-training for anomaly detection and segmentation. In: European conference on computer vision. pp. 392–408. Springer (2022)