OSD-IRF: ONE-STEP DIFFUSION ANOMALY DETECTION WITH INVERSE RESIDUAL FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models have demonstrated strong performance in unsupervised industrial anomaly detection (uIAD) by learning the distribution of normal data. These methods typically assume that off-manifold anomalies are more difficult to reconstruct or generate, leading to larger reconstruction errors in the data space or lower probability densities in a tractable latent space. However, their iterative denoising process incurs substantial inference latency. To overcome this limitation, we propose OSD-IRF, a novel one-step diffusion framework based on inverse residual fields (IRF). Our key observation is that anomalies become distinguishable in the IRF space, enabling efficient detection in a single diffusion step without iterative denoising or model distillation. Specifically, we train an unconditional denoising diffusion probabilistic model (DDPM) exclusively on normal data. At inference time, we use the score function learned by the DDPM to estimate the IRF of each test sample. Anomaly detection is then performed by evaluating the IRF likelihood under a Gaussian distribution and comparing it against a predefined threshold. We further leverage the aggregate signal-to-noise ratio (SNR) statistics of the entire training set to automatically determine the probing timestep during inference, thereby eliminating empirical timestep selection in IRF estimation. Extensive experiments on widely used uIAD benchmarks show that OSD-IRF achieves competitive performance in anomaly detection and localization while substantially reducing the inference cost of diffusion-based methods.

1 INTRODUCTION

Industrial anomaly detection (IAD) plays a crucial role in applications such as industrial quality inspection and product safety. Its goal is to automatically identify industrial images that deviate from normal patterns. In real-world production lines, anomalous images are scarce and costly to acquire, making supervised IAD methods challenging to apply. Consequently, unsupervised IAD (uIAD) methods, trained on large amounts of normal image data, have been developed. The success of uIAD heavily relies on learning a compact high-dimensional distribution of normal data, from which anomalous images become distinguishable at test time. Diffusion-based generative models (DGMs) have therefore emerged as powerful tools for uIAD due to their strong capability to handle high-dimensional data and approximate complex distributions. DGMs for uIAD can be mainly divided into reconstruction-based methods and density estimation-based methods.

Reconstruction-based methods typically treat anomalies as noise and leverage a diffusion model trained exclusively on normal images to denoise anomalous inputs. As a result, anomalous features can be reconstructed into normal ones. Anomalies are then detected based on the discrepancy between the input image features and their reconstructed counterparts, as illustrated in Fig. 1(a). Lu et al. (2023a) introduces random noise to overwhelm anomalous pixels and obtains pixel-wise anomaly scores from the intermediate denoising process. He et al. (2024); Mousakhan et al. (2024) use the input image as a condition to guide the denoising process, yielding defect-free reconstructions while preserving normal patterns. Different from Gaussian diffusion models, Wyatt et al. (2022) develops a multi-scale simplex-noise diffusion process that offers control over the target anomaly size. To speed up inference, OmiAD (Feng et al., 2025) constructs a multi-step adaptive masked diffusion process and compresses it into a one-step denoising model through adversarial score distillation. Despite its effectiveness, the additional distillation increases training complexity and may introduce instability.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

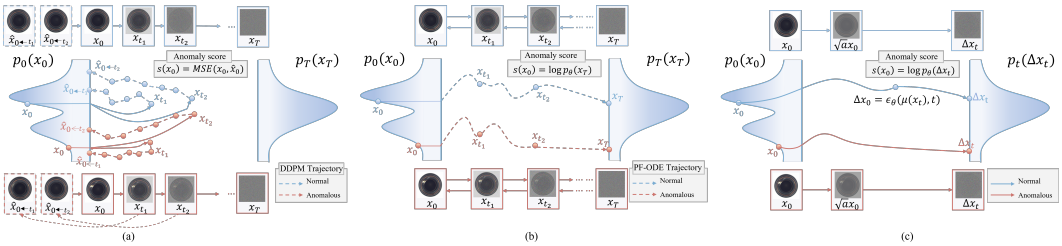


Figure 1: Comparisons of diffusion-based paradigms for uIAD. (a) Reconstruction-based method perturbs an input $\boldsymbol{x}_0$ to an intermediate diffusion state and reconstructs an anomaly-free counterpart $\hat{\boldsymbol{x}}_0$ using a diffusion model trained on normal data. Anomalies are detected from the discrepancy between $\boldsymbol{x}_0$ and $\hat{\boldsymbol{x}}_0$. (b) Density estimation-based method maps $\boldsymbol{x}_0$ to a tractable latent variable $\boldsymbol{x}_T$ through diffusion inversion and evaluates its typicality under the Gaussian distribution. (c) OSD-IRF estimates the inverse residual field (IRF) from the score function learned by a DDPM trained on normal data and uses the resulting IRF response for anomaly detection. The probing timestep $t^\star$ is determined from the aggregate SNR statistics of normal training data. Our method obtains the IRF response with a single diffusion evaluation, without iterative denoising or multi-step inversion.

Density estimation-based methods leverage the denoising process, which implicitly learns the score function of the data distribution (Ho et al., 2020; Song et al., 2021b), to enable likelihood-based anomaly detection, as illustrated in Fig. 1(b). The score function characterizes the gradient of the log-density and therefore contains local distributional information learned from normal data (Song et al., 2021b). For example, DTE (Livernoche et al., 2024) estimates the distribution over time for an input image and uses the mean of this distribution as the anomaly score, while InvAD (Sakai et al., 2025) infers the latent variable through DDIM inversion (Song et al., 2021a) and measures its deviation under the Gaussian distribution. Recent work has also shown that score function learned by diffusion models provide some useful anomaly information (Shin et al., 2023). However, these methods often incur substantial computational overhead at inference time. DTE requires repeated evaluations across multiple reverse diffusion steps, while InvAD maps the input to a tractable Gaussian latent representation through a multi-step inversion procedure. This raises a natural question: can the score function learned by a diffusion model be directly leveraged for anomaly detection, without iterative reverse diffusion or inversion Motivated by this question, we investigate a simple yet effective approach to fast uIAD inference by directly exploiting the learned score function.

To this end, we propose OSD-IRF, a one-step diffusion-based anomaly detection framework built upon inverse residual fields (IRFs). OSD-IRF estimates the IRF by evaluating the aggregated score function constructed from the posterior Gaussian distribution of the noisy state $\boldsymbol{x}_{t^\star}$. The score function is pre-trained exclusively on normal data using DDPM (Ho et al., 2020). The probing timestep $t^\star$ is adaptively selected based on summary signal-to-noise ratio (SNR) statistics of the normal training data, which are shown in Sec. 3.4 to be strongly correlated with uIAD performance. Our key observation is that normal and anomalous patterns induce distinct IRF responses, as illustrated in Fig. 1(c). We further substantiate this observation through a controlled toy-data analysis in Sec. 3.3. Once $t^\star$ is determined, anomaly detection can be performed directly from the IRF response in a single diffusion step, eliminating the iterative denoising, inversion, and model distillation required by prior approaches. The novelty and contributions of our method are summarized as follows:

- We propose IRF for robust uIAD inference, derived from the aggregated score function evaluated at an adaptively selected probing timestep. The probing timestep is determined from the SNR statistics of normal training samples, enabling one-step diffusion inference without iterative denoising or inversion.
- We introduce an effective IRF estimation strategy based on an aggregated score function evaluated over samples drawn from the posterior Gaussian distribution of the noisy state. The probing timestep for score evaluation is adaptively determined by SNR statistics of normal training data, eliminating the need for empirical timestep selection.

- Extensive experiments on three widely used uIAD benchmarks demonstrate that our method achieves competitive anomaly detection and localization performance while substantially reducing inference costs compared with prior diffusion-based methods.

2 PRELIMINARIES

2.1 DIFFUSION MODEL

DDPM (Ho et al., 2020) is a prominent diffusion model that comprises a forward diffusion process and a reverse denoising process. In the forward process, Gaussian noise is incrementally introduced, ultimately converting the data variable $\boldsymbol{x}_0$ into Gaussian noise $\boldsymbol{x}_T$ over T timesteps:

$$q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0) = \prod_{t=1}^{T} q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}), \qquad q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) = \mathcal{N}\left(\sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t I\right) \tag{1}$$

where β_t denotes the noise level at timestep t. Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$. The relationship between $\boldsymbol{x}_t$ and $\boldsymbol{x}_0$ can be expressed by $q(\boldsymbol{x}_t \mid \boldsymbol{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)I)$. Using a Markov chain with Gaussian transitions, the reverse denoising process from $\boldsymbol{x}_T$ back to $\boldsymbol{x}_0$ is formulated as:

$$p_\theta(\boldsymbol{x}_{0:T}) = p(\boldsymbol{x}_T) \prod_{t=1}^{T} p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t), \qquad p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t) = \mathcal{N}\left(\mu_\theta(\boldsymbol{x}_t, t), \sigma_t^2 I\right) \tag{2}$$

where $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)$ is parameterized by the noise prediction network $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$, and σ_t^2 is a fixed time-dependent variance. The forward process provides ground-truth noise signals for training $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\cdot)$ in the reverse denoising process. Following the standard DDPM objective, $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\cdot)$ is trained by minimizing:

$$\mathcal{L}_{\text{DDPM}} = \mathbb{E}_{t, \boldsymbol{x}_0, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) \right\|_2^2 \right], \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_t^2 \mathbf{I}) \tag{3}$$

2.2 SCORE FUNCTION

Song et al. (2021b) establishes a connection between diffusion models and score-based generative modeling through the score function of the perturbed data distribution. At timestep t, an ϵ-prediction DDPM provides an estimate of the marginal score as:

$$s_\theta(\boldsymbol{x}_t, t) := -\frac{1}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\boldsymbol{x}_t, t) \approx \nabla_{\boldsymbol{x}_t} \log q_t(\boldsymbol{x}_t), \tag{4}$$

where $q_t(\boldsymbol{x}_t)$ denotes the marginal distribution of the perturbed data at timestep t. Thus, the predicted noise is proportional to an estimate of the score of the perturbed data distribution. In uIAD, since the DDPM is trained exclusively on normal data, the resulting score estimator should characterize the perturbed distribution of normal data.

2.3 DIFFUSION-BASED UIAD METHODS

As illustrated in Fig. 1(a) and (b), diffusion-based uIAD methods mainly follow two paradigms: reconstruction-based methods and density estimation-based methods. We first state the assumptions underlying these two paradigms and then briefly introduce their inference pipelines.

Reconstruction-based methods (Wyatt et al., 2022; He et al., 2024) assume that normal images lie on a normal manifold and can be faithfully reconstructed, whereas abnormal patterns deviate from this manifold and thus cannot be accurately reconstructed. The anomaly score is therefore derived from the discrepancy between the input image (or feature) and its reconstructed counterpart. In practice, such methods first perturb an input x_0 to an intermediate diffusion state x_t using $x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$, and then denoise x_t back to $\hat{x}_0$ over multiple timesteps using the pretrained noise prediction function $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$. Conditioning information $\boldsymbol{c}$ can be incorporated by $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}([\boldsymbol{x}_t, \boldsymbol{c}], t)$ to further improve reconstruction quality (Mousakhan et al., 2024; Feng et al., 2025).

Density estimation-based methods (Sakai et al., 2025) instead assume the existence of a tractable distribution over certain diffusion-related latent variables, where normal and abnormal patterns

are probabilistically distinguishable. For example, normal patterns are expected to lie in high-probability regions, whereas abnormal patterns tend to fall into low-probability regions. In practice, such a model uses a pre-trained parameterized noise function $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ from Eq. 3 to perturb an input $\boldsymbol{x}_0$ to a latent state $\boldsymbol{x}_T$ over multiple timesteps using jump perturbation for speedup, where $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The anomaly score is then determined based on the typicality of $\boldsymbol{x}_T$ under the Gaussian reference distribution. Despite their effectiveness, these diffusion-based methods still suffer from either complicated training or slow inference.

3 METHOD

We first introduce the inverse residual field (IRF). We then develop a robust IRF estimator based on the aggregated score function, followed by an adaptive probing timestep selection mechanism guided by signal-to-noise ratio (SNR) statistics of normal training data. Finally, we derive the anomaly score from the log-likelihood of the estimated IRF under a standard Gaussian distribution.

3.1 INVERSE RESIDUAL FIELDS

As illustrated in Fig. 1(c), rather than assessing the typicality of $\boldsymbol{x}_T$ against a Gaussian distribution at the terminal timestep, as in InvAD (Sakai et al., 2025), we find that the residual $\Delta\boldsymbol{x}_t$ provides a more discriminative signal for anomaly detection. We term this residual the inverse residual field (IRF), and further validate its discriminative behavior through a controlled toy-data analysis in Sec. 3.3. At an appropriately selected probing timestep, the IRF yields a clearer separation between normal and anomalous samples than $\boldsymbol{x}_T$, while avoiding the iterative inversion procedure required by InvAD.

Formally, we define the IRF at timestep t as the residual between $\boldsymbol{x}_t$ and $\boldsymbol{x}_0$:

$$\Delta\boldsymbol{x}_t = \frac{\boldsymbol{x}_t - \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0}{\sqrt{1-\bar{\alpha}_t}}, \tag{5}$$

However, the IRF defined above cannot be directly used to generate samples for subsequent anomaly detection. Fortunately, according to Eq. 3, the IRF approximately follows the same distribution as $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\cdot, t)$. This observation allows us to reformulate $\Delta\boldsymbol{x}_t$ in the IRF as:

$$\Delta\boldsymbol{x}_t = \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t), \quad \boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon} \tag{6}$$

Remark: The IRF in Eq. 6 is proportional to the score function in Eq. 4, up to the scaling factor $-\frac{1}{\sqrt{1-\bar{\alpha}_t}}$. In practice, applying Eq. 6 yields unstable estimates, limiting its reliability for anomaly detection. We attribute this instability to the noisy state $\boldsymbol{x}_t$, which is sensitive to the perturbation $\boldsymbol{\epsilon}$.

3.2 STABLE IRF ESTIMATION

To mitigate this sensitivity, we develop a robust IRF estimator by aggregating over stochastic perturbations at both the input and output levels of the IRF.

3.2.1 INPUT AGGREGATION

Following the reparameterization trick in (Kingma & Welling, 2013), we replace the stochastic noisy state $\boldsymbol{x}_t$ in Eq. 6 with its mean. This yields the input-aggregated IRF $\Delta\boldsymbol{x}_t^{in}$ as:

$$\Delta\boldsymbol{x}_t^{in} = \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mu(\boldsymbol{x}_t), t), \quad \mu(\boldsymbol{x}_t) = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 \tag{7}$$

By eliminating the influence of the random perturbation term $\boldsymbol{\epsilon}$, the input-aggregated IRF provides more robust and stable estimates for uIAD than the original IRF defined in Eq. 6.

3.2.2 OUTPUT AGGREGATION

We first generate K IRF samples according to Eq. 6, and then aggregate them via average pooling to obtain the output-aggregated IRF $\Delta\boldsymbol{x}_t^{out}$ for $k = 1, ..., K$ as:

$$\Delta\boldsymbol{x}_t^{out} = \frac{1}{K}\sum_{k=1}^{K}\Delta\boldsymbol{x}_t^k, \quad \Delta\boldsymbol{x}_t^k = \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t^k, t), \quad \boldsymbol{x}_t^k = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}^k \tag{8}$$

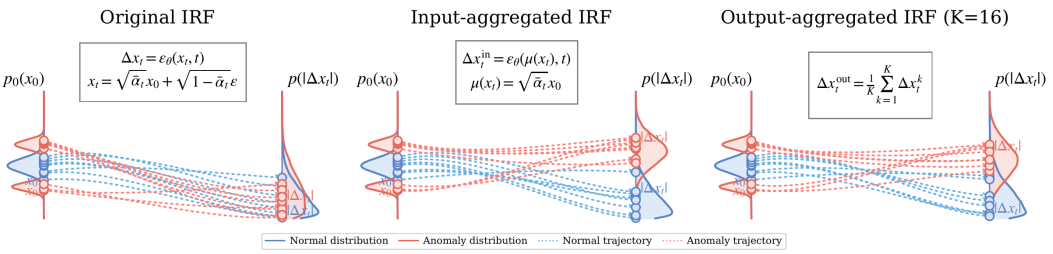


Figure 2: From left to right: trajectories from the input space $\boldsymbol{x}_0$ to the corresponding IRF spaces defined by $\Delta \boldsymbol{x}_t$, $\Delta \boldsymbol{x}_t^{\text{in}}$, and $\Delta \boldsymbol{x}_t^{\text{out}}$, respectively.

Similar to its input-aggregated counterpart, the output-aggregated IRF provides more robust and stable estimates for anomaly detection by reducing the influence of random perturbations. However, its effectiveness depends on the number of sampled IRFs. When $K = 1$, it degenerates to the original IRF in Eq. 6. Consequently, compared with the input-aggregated IRF, $\Delta \boldsymbol{x}_t^{\text{out}}$ incurs higher computational cost due to repeated sampling. As K increases, $\Delta \boldsymbol{x}_t^{\text{out}}$ is expected to progressively approach a stable estimate comparable to $\Delta \boldsymbol{x}_t^{\text{in}}$, a trend that is empirically validated in our experiments.

3.3 A Toy Data Analysis on OSD-IRF

We construct a one-dimensional toy example to provide a controlled illustration of different IRF estimation strategies. Normal data is generated from a Gaussian distribution, while anomalous data is generated from a Gaussian mixture distribution:

$$p_{\text{normal}}(x) = \mathcal{N}(2.5, 0.35^2), \quad p_{\text{anomalous}}(x) = 0.5\mathcal{N}(1.5, 0.2^2) + 0.5\mathcal{N}(3.5, 0.22^2) \tag{9}$$

We draw 10,000 samples from $p_{\text{normal}}(x)$ to train a standard DDPM, without exposing the model to any anomalous samples. At inference time, we fix the probing timestep to $t = 250$ and independently draw 6,000 normal and 6,000 anomalous samples. For each sample, we compute the IRF response using three estimation strategies: the original IRF, the input-aggregated IRF, and the output-aggregated IRF with $K = 16$, corresponding to the left, middle, and right panels of Fig. 2, respectively. As shown in Fig. 2, the original IRF exhibits substantial overlap between the responses of normal and anomalous samples, making the two distributions difficult to distinguish. In contrast, both the input-aggregated and output-aggregated IRFs produce substantially clearer separation between normal and anomalous responses. These results demonstrate that aggregation effectively suppresses the variability induced by stochastic perturbations and yields more discriminative and stable IRF estimates for anomaly detection.

3.4 SNR-Guided Adaptive Timestep Selection

As indicated by Eqs. 7 and 8, the effectiveness of both the input- and output-aggregated IRFs depends critically on the choice of probing timestep t, since the discriminability of the IRF varies substantially across timesteps. As shown in Fig. 3 (left), the averaged anomaly detection performance (mAD, %) improves over an appropriate range of timesteps, but degrades markedly when the probing timestep is either too small or too large. We attribute this behavior to the timestep-dependent discrepancy between the normal and anomalous feature distributions. As shown in Fig. 3 (middle), the distributional discrepancy, measured by the optimal transport (OT) distance (Cuturi, 2013), exhibits a trend consistent with the mAD performance, a larger OT distance indicates better separation between normal and anomalous samples, whereas a smaller OT distance corresponds to reduced discriminability. However, selecting the probing timestep $t^\star$ based on performance is impractical in the unsupervised setting, as anomalous samples are unavailable for model selection. Fortunately, prior studies on deep generative models have shown that the signal intensity of generated data affects the effective signal-to-noise ratio (SNR) of the noisy state $\boldsymbol{x}_t$ (Yao et al., 2024). This observation motivates us to develop an SNR-guided adaptive timestep selection strategy for uIAD.

For a diffusion model, the schedule-dependent signal-to-noise ratio (SNR) at timestep t can be expressed as $\text{SNR}_{\text{sch}}(t) = \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}$ However, this quantity is determined solely by the predefined

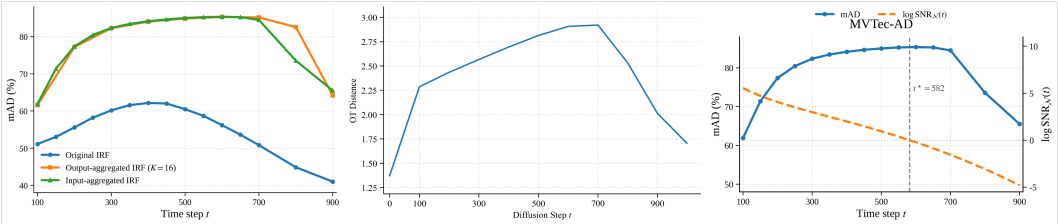


Figure 3: From left to right: mAD performance, OT distance, and mAD/$\log \mathrm{SNR}$ as functions of the probing timestep. All results are reported on the MVTec-AD dataset.

diffusion schedule and does not account for the scale of the feature representations. We therefore estimate the average signal power from the normal training data $\mathcal{D}_{\mathcal{N}}$ as $P_{\mathcal{N}} = \mathbb{E}_{x_0 \sim \mathcal{D}_{\mathcal{N}}}\left[\frac{\|\boldsymbol{x}_0\|_2^2}{d}\right]$, where d is the feature dimension of $\boldsymbol{x}_0$.

Accordingly, we define the calibrated SNR as $\mathrm{SNR}_{\mathcal{N}}(t) = P_{\mathcal{N}} \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}$. Previous studies identify the neighborhood of $\log \mathrm{SNR} = 0$ as a transition regime between signal-dominated and noise-dominated diffusion states (Yao et al., 2024; Hang et al., 2025). Motivated by this observation, we select the probing timestep $t^\star$ such that its $\log \mathrm{SNR}$ is closest to this transition point:

$$t^\star = \arg\min_t \left|\log \mathrm{SNR}_{\mathcal{N}}(t)\right|. \tag{10}$$

Importantly, Eq. equation 10 relies solely on normal training features and the predefined diffusion schedule, without requiring anomalous samples or anomaly detection performance for timestep selection. In Fig. 3 (right), we report the mAD performance and $\log \mathrm{SNR}$ across different timesteps on the MVTec-AD dataset. The close correspondence between their trends empirically supports our analysis and validates the effectiveness of the proposed SNR-guided timestep selection strategy.

3.5 Anomaly Detection and Localization

During inference, given the adaptively selected timestep $t^\star$, we compute the input/output-aggregated IRF response (taking the input-aggregated IRF for example) of the input image features x_0 as $\boldsymbol{R} := \Delta \boldsymbol{x}_t^{in} = \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\sqrt{\bar{\alpha}_{t^\star}}\boldsymbol{x}_0, t^\star)$ For anomaly localization, we aggregate the IRF along the channel dimension to obtain a feature-scale anomaly map $\hat{\boldsymbol{A}} \in \mathbb{R}^{h \times w}$ and then the image-scale anomaly map $\boldsymbol{A} \in \mathbb{R}^{H \times W}$ is obtained by bilinear interpolation $\mathrm{Upsample}(\cdot)$:

$$\hat{\boldsymbol{A}}[i,j] = \sqrt{\sum_{r=1}^{c} \boldsymbol{R}^2[r,i,j]}, \quad \boldsymbol{A} = \mathrm{Upsample}(\hat{\boldsymbol{A}}) \tag{11}$$

For anomaly detection, the anomaly score is computed as:

$$s = s_{\mathrm{diff}} + s_{\mathrm{nll}}, \quad s_{\mathrm{diff}} = \max_{(i,j)} \hat{\boldsymbol{A}}[i,j] - \min_{(i,j)} \hat{\boldsymbol{A}}[i,j], \quad s_{\mathrm{nll}} = \frac{1}{2}\sum_{r,i,j} \boldsymbol{R}^2[r,i,j] \tag{12}$$

where s_{diff} is the difference between the maximum and minimum values of the feature-scale score map $\hat{\boldsymbol{A}}$, and s_{nll} is the negative log-likelihood of the IRF under a standard Gaussian distribution.

4 Experiments

4.1 Datasets

We evaluate the effectiveness and efficiency of our model on three widely used uIAD benchmarks including MVTec-AD Bergmann et al. (2019), ViSA Zou et al. (2022), and MPDD Jezek et al. (2021). MVTec-AD (15 classes) has 2,629 normal training images, 467 normal and 1,258 abnormal test images at 1024×1024 resolution. ViSA (12 classes) has 8,659 normal training images, 962 normal and 1,200 abnormal test images at 256×256 resolution. MPDD (6 classes) has 888 normal training images, 176 normal and 282 abnormal test images at 1024×1024 resolution.

4.2 Evaluation Metrics

Following (Feng et al., 2025), we evaluate image-level and pixel-level AD performance using standard metrics (%): AU-ROC, AP, and $\mathrm{F1}_{\max}$ for image-level detection, and AU-ROC, AP, $\mathrm{F1}_{\max}$, and AU-PRO for pixel-level localization. We additionally report mAD, defined as the average of these seven metrics, to summarize the overall detection and localization performance. We employ frames per second (FPS) to report the inference speed.

4.3 Baseline AD Methods

We compare our model with most recently proposed uIAD methods, including UniAD You et al. (2022) and HVQTrans Lu et al. (2023b), SimpleNet Liu et al. (2023), DeSTSeg Zhang et al. (2023), RD4AD Deng & Li (2022), DiAD He et al. (2024), OmiAD Feng et al. (2025), and InvAD Sakai et al. (2025). UniAD and HVQTrans are transformer-based reconstruction methods. SimpleNet, DeSTSeg, and RD4AD are CNN-based reconstruction methods. DiAD and OmiAD are diffusion-based reconstruction methods. InvAD is diffusion-based density estimation method.

4.4 Implementation Details

We resize all images to 256×256 without data augmentation. Image features are extracted using an EfficientNet-B4 (Tan & Le, 2019) pretrained on ImageNet-1K (Deng et al., 2009), and a DiT-giant (Peebles & Xie, 2023) with 16 Transformer blocks, a hidden dimension of 2048, and 768 latent channels is trained from scratch on the normal features. Training follows the standard DDPM setting (Ho et al., 2020) with a linear noise schedule and $T = 1000$ timesteps. We use AdamW (Loshchilov & Hutter, 2019) and train for 300 epochs with a batch size of 32 for all datasets.

During inference, we use the input-aggregated IRF. The probing timestep is selected from normal training features according to Eq. 10, yielding $t^\star = 582$, 587, and 588 for MVTec-AD, VisA, and MPDD, respectively. All experiments are conducted on a single NVIDIA RTX 5880 Ada GPU with 48 GB memory.

4.5 Comparison with State-of-the-Art Methods

Table 1 compares OSD-IRF with representative uIAD methods on the three benchmarks. OSD-IRF achieves mAD scores of 86.0, 81.5, and 83.5 on MVTec-AD, VisA, and MPDD, respectively, outperforming the compared methods on all three benchmarks. It also demonstrates strong pixel-level localization performance, achieving the best P-AP, P-$\mathrm{F1}_{\max}$, and AU-PRO on all three datasets. Notably, on MPDD, OSD-IRF achieves the best results across all seven image- and pixel-level metrics. Unlike diffusion-based reconstruction or inversion methods that require iterative inference, OSD-IRF directly evaluates the IRF at a single probing timestep. The input-aggregated IRF requires only one diffusion-network evaluation, without iterative denoising, multi-step inversion, or model distillation, achieving 133.0, 105.9, and 212.4 FPS on MVTec-AD, VisA, and MPDD, respectively.

4.6 Qualitative Results

Figure 4 presents representative anomaly localization results on MVTec-AD, VisA, and MPDD. OSD-IRF produces well-localized anomaly responses across defects with different spatial scales and structures, including small localized defects, texture irregularities, and larger structural anomalies. The predicted anomaly maps generally focus on the defective regions while suppressing responses in normal background areas.

Despite the considerable variation in object appearance and defect patterns across the three benchmarks, OSD-IRF exhibits consistent localization behavior and captures both fine-grained and spatially extended anomalies. These qualitative observations are consistent with the strong pixel-level performance reported in Table 1, further demonstrating that the IRF response provides informative spatial cues for anomaly localization.

Table 1: Quantitative comparison with state-of-the-art uIAD methods on MVTec-AD, VisA, and MPDD under the multi-class setting. The best and second-best results are highlighted in bold and underlined, respectively.

Dataset	Method	Image-level			Pixel-level				mAD	FPS
		AU-ROC	AP	$F1_{max}$	AU-ROC	AP	$F1_{max}$	AU-PRO		
MVTec-AD	RD4AD	94.6	96.5	95.2	96.1	48.6	53.8	91.1	82.3	4.8
	UniAD	96.5	98.8	96.2	96.8	43.4	49.5	90.7	81.7	5.3
	SimpleNet	95.3	98.4	95.8	96.9	45.9	49.7	86.5	81.2	N/A
	DeSTSeg	89.2	95.5	91.6	93.1	54.3	50.9	64.8	77.1	N/A
	DiAD	97.2	99.0	96.5	96.8	52.6	55.5	90.7	84.0	0.1
	HVQ-Trans	98.0	99.5	97.5	97.3	48.2	53.3	91.4	83.6	5.6
	OmiAD	98.8	**99.7**	**98.5**	97.7	52.6	56.7	93.2	85.3	39.4
	InvAD	**99.0**	99.6	**98.5**	97.5	46.5	52.3	92.7	83.7	88.1
	OSD-IRF (Ours)	98.9	99.5	98.3	**97.8**	**55.7**	**58.1**	**93.7**	**86.0**	**133.0**
VisA	RD4AD	92.4	92.4	89.6	98.1	38.0	42.6	91.8	77.8	4.9
	UniAD	88.8	90.8	85.8	98.3	33.7	39.0	85.5	74.6	4.6
	SimpleNet	87.2	87.0	81.8	96.8	34.7	37.8	81.4	72.4	N/A
	DeSTSeg	88.9	89.0	85.2	96.1	39.6	43.4	67.4	72.8	N/A
	DiAD	86.8	88.3	85.1	96.0	26.1	33.0	75.2	70.1	0.1
	HVQ-Trans	93.2	92.8	87.6	98.7	35.0	39.6	86.3	76.2	5.0
	OmiAD	95.3	96.0	91.2	98.9	40.4	44.1	89.2	79.3	35.3
	InvAD	**96.9**	97.2	93.7	**99.1**	39.2	43.1	92.9	80.3	74.1
	OSD-IRF (Ours)	**96.9**	**97.4**	**93.9**	99.0	**43.7**	**46.1**	**93.3**	**81.5**	**105.9**
MPDD	RD4AD	84.1	83.2	84.1	98.1	35.2	38.7	93.4	73.8	4.7
	UniAD	82.2	87.1	85.1	95.1	18.9	25.0	81.9	67.9	5.8
	SimpleNet	90.6	94.1	89.7	97.1	33.6	35.7	90.0	75.8	N/A
	DeSTSeg	93.0	95.1	90.6	94.1	33.2	37.6	59.8	71.9	N/A
	DiAD	74.6	82.1	82.5	93.0	15.9	21.2	78.4	64.0	0.1
	HVQ-Trans	86.5	88.1	85.8	96.7	27.6	31.4	86.9	71.9	6.2
	OmiAD	93.7	95.5	90.9	98.6	37.6	42.3	94.0	78.9	49.8
	InvAD	96.5	96.5	94.4	98.3	40.1	43.9	91.2	80.1	120.0
	OSD-IRF (Ours)	**97.7**	**98.2**	**95.7**	**98.8**	**48.5**	**50.6**	**95.1**	**83.5**	**212.4**

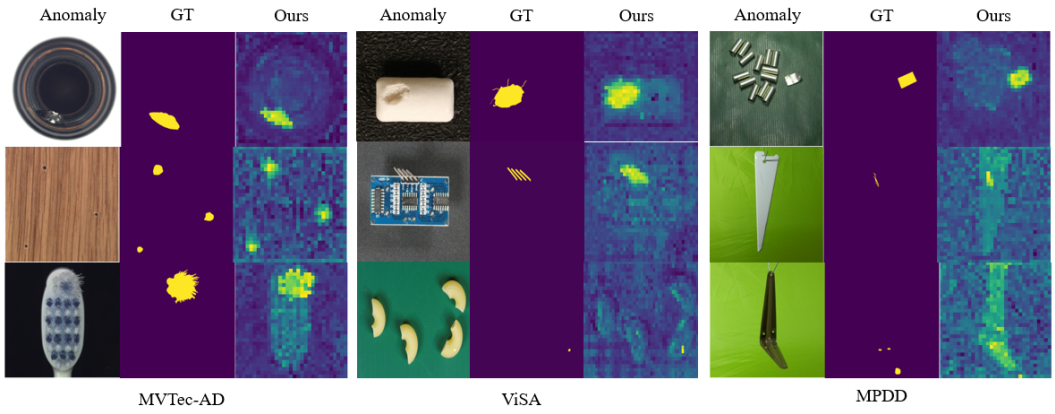
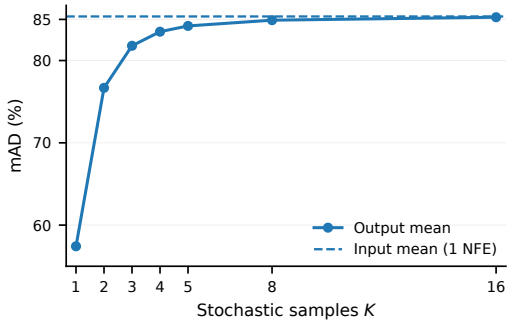


Figure 4: Qualitative anomaly localization results of OSD-IRF on MVTec-AD, VisA, and MPDD. Each example shows the anomalous input image, ground-truth anomaly mask, and predicted anomaly map.

4.7 Analysis of IRF Estimation Strategies

We analyze the three IRF estimation strategies introduced in our method: Original IRF, Output-aggregated IRF, and Input-aggregated IRF. Unless otherwise specified, this analysis is conducted on MVTec-AD at the SNR-guided timestep $t^\star = 582$. Here, NFE denotes the number of diffusion-network evaluations required for one test sample.

Figure 5(a) shows the effect of the number of stochastic samples K on Output-aggregated IRF. Increasing K consistently improves performance, with mAD rising from 57.4 at $K = 1$ to 76.7, 84.2, and 85.3 at $K = 2$, 5, and 16, respectively, while the improvement gradually saturates. Figure 5(b) further compares the three IRF estimation strategies. Input-aggregated IRF achieves 86.0 mAD with



(a) Output-aggregated IRF vs. K

Metric	Original	Output-Agg.	Input-Agg.
NFE	1	16	1
I-AUROC	75.7	98.7	98.9
I-AP	88.6	99.5	99.5
I-F1$_{\max}$	86.1	98.1	98.3
P-AUROC	71.1	97.6	97.8
P-AP	14.7	53.1	55.7
P-F1$_{\max}$	21.3	56.6	58.1
AU-PRO	44.2	93.3	93.7
mAD	57.4	85.3	86.0

(b) IRF estimator comparison

Figure 5: Comparison of IRF estimation strategies on MVTec-AD at $t^\star = 582$. (a) Performance of Output-aggregated IRF as the number of stochastic samples K increases; the dashed line denotes Input-aggregated IRF, which requires one NFE. (b) Direct comparison of Original IRF, Output-aggregated IRF ($K = 16$), and Input-aggregated IRF.

only one NFE, slightly outperforming the 85.3 mAD of Output-aggregated IRF with $K = 16$, which requires 16 NFEs. These results show that aggregating multiple stochastic responses progressively mitigates the sensitivity of Original IRF to random perturbations, but incurs additional network evaluations. In contrast, Input-aggregated IRF provides a deterministic response and achieves comparable or better performance with a single network evaluation. We therefore adopt Input-aggregated IRF as the default estimator.

4.8 Ablation of Anomaly Scoring

We evaluate the two components of the image-level anomaly score on MVTec-AD at $t^\star = 582$. Specifically, s_{diff} measures the anomaly response derived from the IRF representation, while s_{nll} evaluates the likelihood of the corresponding response under the normal-data distribution. Since this ablation only changes the image-level scoring function, the pixel-level anomaly map remains unchanged.

Table 2: Ablation of image-level anomaly scoring on MVTec-AD at $t^\star = 582$.

Score	I-AUROC	I-AP	I-F1$_{\max}$
s_{diff}	97.8	98.8	97.3
s_{nll}	94.7	97.8	95.0
$s_{\text{diff}} + s_{\text{nll}}$	**98.9**	**99.5**	**98.3**

As shown in Table 2, s_{diff} provides stronger image-level discrimination than s_{nll} when used independently, while their combination further improves all three image-level metrics. This suggests that the two terms capture complementary anomaly information: s_{diff} reflects the magnitude of the IRF response, whereas s_{nll} provides a distribution-aware measure of deviation from normality. We therefore use $s_{\text{diff}} + s_{\text{nll}}$ as the final image-level anomaly score.

5 Conclusion

We investigate whether the noise-prediction function learned by a DDPM on normal data can support fast unsupervised anomaly detection without iterative denoising or inversion. Single noisy-state IRF responses are sensitive to stochastic perturbations, whereas input-aggregated IRF achieves performance comparable to multi-sample averaging with only one network evaluation. Normal-data SNR statistics further provide a practical inference-timestep reference without anomaly-dependent tuning. Overall, OSD-IRF enables direct and efficient anomaly detection from IRF responses.

AI USE STATEMENT

Generative AI tools were used solely for translation, language editing, and improving the clarity and presentation of the manuscript. All research ideas, methodological decisions, experiments, analyses, and conclusions were developed, conducted, and verified by the authors. The authors take full responsibility for the accuracy and integrity of the final manuscript.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the proposed method, model configuration, training settings, and evaluation protocols in the main paper to facilitate reproducibility. All reported results were obtained following the procedures described in the manuscript.

REFERENCES

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. 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.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Hanqiu Deng and Xingyu Li. 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.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Yaoxuan Feng, Wenchao Chen, Yuxin Li, Bo Chen, Yubiao Wang, Zixuan Zhao, Hongwei Liu, and Mingyuan Zhou. OmiAD: One-step adaptive masked diffusion model for multi-class anomaly detection via adversarial distillation. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 16604–16633. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/feng25b.html.

Tiankai Hang, Shuyang Gu, Jianmin Bao, Fangyun Wei, Dong Chen, Xin Geng, and Baining Guo. Improved noise schedule for diffusion training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4796–4806, 2025.

Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 8472–8480, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In *2021 13th International congress on ultra modern telecommunications and control systems and workshops (ICUMT)*, pp. 66–71. IEEE, 2021.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. 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.

Victor Livernoche, Vineet Jain, Yashar Hezaveh, and Siamak Ravanbakhsh. On diffusion modeling for anomaly detection. *ICLR*, 2024.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Fanbin Lu, Xufeng Yao, Chi-Wing Fu, and Jiaya Jia. Removing anomalies as noises for industrial defect localization. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 16166–16175, 2023a.

Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. *Advances in Neural Information Processing Systems*, 36:8487–8500, 2023b.

Arian Mousakhan, Thomas Brox, and Jawad Tayyub. Anomaly detection with conditioned denoising diffusion models. In *DAGM German Conference on Pattern Recognition*, pp. 181–195. Springer, 2024.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023.

Shunsuke Sakai, Xiangteng He, Chunzhi Gu, Leonid Sigal, and Tatsuhito Hasegawa. Invad: Inversion-based reconstruction-free anomaly detection with diffusion models. *arXiv preprint arXiv:2504.05662*, 2025.

Woosang Shin, Jonghyeon Lee, Taehan Lee, Sangmoon Lee, and Jong Pil Yun. Anomaly detection using score-based perturbation resilience. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23372–23382, 2023.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *ICLR*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *ICLR*, 2021b.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114. PMLR, 2019.

Julian Wyatt, Adam Leach, Sebastian M Schmon, and Chris G Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 650–656, 2022.

Jingfeng Yao, Cheng Wang, Wenyu Liu, and Xinggang Wang. FasterDiT: Towards faster diffusion transformers training without architecture modification. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022.

Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. 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.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European conference on computer vision*, pp. 392–408. Springer, 2022.

A APPENDIX

A.1 MORE QUANTITATIVE RESULTS FOR EACH CATEGORY

To provide a more comprehensive evaluation, we present additional quantitative results for each category across three benchmark datasets: MVTec-AD, VisA, and MPDD. The reported evaluation metrics include the Area Under the Receiver Operating Characteristic Curve (AUROC), Average Precision (AP), and maximum F1 score (F1-max) for image-level anomaly detection. Additionally, for pixel-level anomaly localization, we report pixel-level AUROC, AP, F1-max, and the Area Under the Per-Region-Overlap (AU-PRO).

These detailed per-category results provide a deeper understanding of the proposed method across different inspection categories. By analyzing the results on each dataset, we can better observe the robustness and adaptability of the proposed input-aggregated IRF formulation in diverse industrial inspection scenarios, demonstrating its effectiveness in both anomaly detection and anomaly localization tasks.

Table 3: Comparison with SOTA methods on MVTec-AD dataset for multi-class anomaly detection with AU-ROC/AP/F1$_{\max}$ metrics. OSD-IRF results use the SNR-guided timestep $t^\star = 582$.

	Category	RD4AD	SimpleNet	DeSTSeg	DiAD	HVQ-Trans	OmiAD	Ours
Object	Bottle	99.6/99.9/98.4	**100.0/100.0/100.0**	98.7/99.6/96.8	99.7/96.5/91.8	**100.0/100.0/100.0**	**100.0/100.0/100.0**	**100.0/100.0/100.0**
	Cable	84.1/89.5/82.5	97.5/98.5/94.7	89.5/94.6/85.9	94.8/98.8/95.2	**99.0**/98.8/95.1	98.4/**99.4/95.6**	98.7/99.2/**95.6**
	Capsule	94.1/96.9/96.9	90.7/97.9/93.5	82.8/95.9/92.6	89.0/97.5/95.5	95.4/99.2/96.3	94.7/99.3/96.8	**98.6/99.7/98.2**
	Hazelnut	60.8/69.8/86.4	99.9/99.9/99.3	98.8/99.2/98.6	99.5/99.7/97.3	**100.0/100.0**/99.3	**100.0/100.0/100.0**	**100.0/100.0**/99.3
	MetalNut	**100.0/100.0/99.5**	96.9/99.3/96.1	92.9/98.4/92.2	99.1/96.0/91.6	99.9/99.9/98.9	99.4/99.9/98.9	98.9/99.7/98.9
	Pill	97.5/99.6/96.8	88.2/97.7/92.5	77.1/94.4/91.7	95.7/98.5/94.5	95.8/99.2/94.9	94.2/99.2/95.4	**98.9/99.8/98.9**
	Screw	97.7/99.3/95.8	76.7/90.6/87.7	69.9/88.4/85.4	90.7/**99.7/97.9**	95.6/97.9/92.1	96.9/98.8/96.3	**98.3**/99.4/95.7
	Toothbrush	97.2/99.0/94.7	89.7/95.7/92.3	71.7/89.3/84.5	**99.7**/99.9/99.2	93.6/99.9/98.4	**99.7/100.0/100.0**	91.7/96.1/95.2
	Transistor	94.2/95.2/90.0	99.2/98.7/97.6	78.2/79.5/68.8	99.8/99.6/97.4	99.7/99.5/96.4	**99.9/99.9/98.8**	99.5/99.4/97.4
	Zipper	99.5/99.9/99.2	99.0/99.7/98.3	88.4/96.3/93.1	95.1/99.1/94.4	97.9/99.6/98.3	99.8/**100.0**/99.6	**100.0/100.0/100.0**
Texture	Carpet	98.5/99.6/97.2	95.7/98.7/93.2	95.9/98.8/94.9	99.4/99.9/98.3	**99.9/100.0/100.0**	99.6/**100.0**/99.4	99.2/99.8/98.3
	Grid	98.0/99.4/96.5	97.6/99.2/96.4	97.9/99.2/96.6	98.5/99.8/97.7	97.0/99.5/97.3	99.8/99.9/99.1	**100.0/100.0/100.0**
	Leather	**100.0/100.0/100.0**	**100.0/100.0/100.0**	99.2/99.8/98.9	99.8/99.7/97.6	**100.0/100.0/100.0**	**100.0/100.0/100.0**	**100.0/100.0/100.0**
	Tile	98.3/99.3/96.4	99.3/99.8/98.8	97.0/98.9/95.3	96.8/99.9/98.4	99.2/99.8/98.2	**100.0**/99.9/98.8	**100.0/100.0/100.0**
	Wood	99.2/99.8/98.3	98.4/99.5/96.7	**99.9/100.0**/99.2	99.7/**100.0/100.0**	97.2/99.6/97.4	99.0/99.8/98.3	99.0/99.7/97.5
	Mean	94.6/96.5/95.2	95.3/98.4/95.8	89.2/95.5/91.6	97.2/99.0/96.5	98.0/99.5/97.5	98.8/**99.7/98.5**	**98.9**/99.5/98.3

Table 4: Comparison with SOTA methods on MVTec-AD dataset for multi-class anomaly localization with AU-ROC/AP/F1$_{\max}$/AU-PRO metrics. OSD-IRF results use the SNR-guided timestep $t^\star = 582$.

	Category	RD4AD	SimpleNet	DeSTSeg	DiAD	HVQ-Trans	OmiAD	Ours
Object	Bottle	97.8/68.2/67.6/94.0	97.2/53.8/62.4/89.0	93.3/61.7/56.0/67.5	98.4/52.2/54.8/86.6	98.3/71.8/70.2/94.6	**98.6/74.9**/73.8/**95.6**	98.5/71.3/**74.9**/95.0
	Cable	85.1/26.3/33.6/75.1	96.7/42.4/51.2/85.4	89.3/37.5/40.5/49.4	96.8/50.1/57.8/80.5	98.1/52.4/59.0/87.5	**98.3/60.5/62.9/91.8**	97.5/52.2/56.8/90.0
	Capsule	98.8/43.4/50.0/94.8	98.5/5.4/44.3/84.5	95.8/47.9/48.9/62.1	97.1/42.0/45.3/87.2	98.8/45.3/49.7/90.7	98.9/48.1/52.0/92.5	**99.0/50.8/54.0/95.9**
	Hazelnut	97.9/36.2/51.6/92.7	98.4/44.6/51.4/87.4	98.2/65.8/61.6/84.5	98.3/**79.2/80.4**/91.5	**98.8**/62.7/63.2/92.5	98.6/59.6/59.9/94.3	98.4/61.7/59.4/**95.2**
	MetalNut	94.8/55.5/66.4/91.9	**98.0/83.1**/79.4/85.2	84.2/42.0/22.8/53.0	97.3/30.0/38.3/90.6	96.3/67.1/75.5/90.9	96.5/66.6/75.6/90.3	97.7/77.3/**81.2/94.2**
	Pill	97.5/63.4/65.2/95.8	96.5/**72.4/67.7**/81.9	96.2/61.7/41.8/27.9	95.7/46.0/51.4/89.0	97.1/50.1/57.6/94.9	96.6/56.8/60.7/**95.9**	**97.6**/63.7/67.0/95.8
	Screw	99.4/40.2/44.6/96.8	96.5/15.9/23.2/84.0	93.8/19.9/25.3/47.3	97.9/**60.6/59.6**/95.0	98.9/28.8/36.2/94.3	**99.5**/38.7/43.5/97.2	**99.5**/54.7/54.4/**97.5**
	Toothbrush	**99.0**/53.6/58.8/92.0	98.4/46.9/52.5/87.4	96.2/52.9/58.8/30.9	**99.0/78.7/72.8/95.0**	98.6/40.8/51.4/89.2	98.7/40.5/56.2/91.1	98.8/53.0/58.4/92.8
	Transistor	85.9/42.3/45.2/74.7	95.8/58.2/56.0/83.2	73.6/38.4/39.2/43.9	95.1/15.6/31.7/90.0	97.9/71.2/67.2/95.4	**98.4/73.4/72.5/96.1**	96.1/60.3/58.3/89.3
	Zipper	98.5/53.9/60.3/94.1	97.9/53.4/54.6/90.7	97.3/64.7/59.2/66.9	96.2/60.7/60.0/91.6	97.5/38.7/48.8/91.7	98.6/52.7/59.3/95.6	**99.1/68.1/66.9/97.1**
Texture	Carpet	**99.0**/58.5/**60.4/95.1**	97.4/38.7/43.2/90.6	93.6/**59.9**/58.9/89.3	98.6/42.2/46.4/90.6	98.7/57.5/57.7/94.7	98.5/52.9/54.8/94.6	98.5/52.0/53.3/95.0
	Grid	96.5/23.0/28.4/**97.0**	96.8/20.5/27.6/88.6	97.0/42.1/46.9/86.8	96.6/**66.0/64.1**/94.0	97.0/24.5/30.5/89.5	98.5/35.4/37.1/95.5	**98.7**/39.1/43.2/96.0
	Leather	99.3/38.0/45.1/97.4	98.7/28.5/32.9/92.7	**99.5/71.5/66.5**/91.1	98.8/56.1/62.3/91.3	98.8/33.7/36.6/**97.6**	98.9/36.3/39.4/96.9	98.8/39.2/39.4/97.0
	Tile	95.3/48.5/60.5/85.8	**95.7**/60.5/59.9/90.6	93.0/**71.0/66.2**/87.1	92.4/65.7/64.1/**90.7**	92.2/41.6/52.9/81.2	92.7/47.5/54.7/82.2	94.3/47.0/57.7/86.1
	Wood	95.3/47.8/50.0/90.0	91.4/34.8/39.7/76.3	**95.9/77.3/71.3**/83.4	93.3/43.3/43.5/**97.5**	92.4/37.2/42.6/86.6	94.2/44.5/48.0/88.5	94.1/44.4/46.4/88.5
	Mean	96.1/48.6/53.8/91.1	96.9/45.9/49.7/86.5	93.1/54.3/50.9/64.8	96.8/52.6/55.5/90.7	97.3/48.2/53.3/91.4	97.7/52.6/56.7/93.2	**97.8/55.7/58.1/93.7**

Table 5: Comparison with SOTA methods on MPDD dataset for multi-class anomaly detection with AU-ROC/AP/F1$_{\max}$ metrics. OSD-IRF results use the SNR-guided timestep $t^\star = 588$.

Category	RD4AD	SimpleNet	DeSTSeg	DiAD	HVQ-Trans	OmiAD	Ours
BracketBlack	81.1/86.5/83.6	78.7/85.3/80.4	82.3/87.9/81.2	81.7/87.0/84.6	90.2/94.5/89.4	92.8/95.4/90.9	**99.4/99.6/96.8**
BracketBrown	84.0/85.3/91.1	92.4/95.3/93.5	96.9/97.2/98.1	76.4/87.3/83.2	89.7/88.5/89.3	94.9/94.2/90.3	**100.0/100.0/100.0**
BracketWhite	81.1/79.8/77.1	88.4/91.4/83.0	94.3/95.8/91.5	72.6/74.7/75.0	78.0/78.4/75.0	87.4/92.7/85.2	**97.0/97.8/94.9**
Connector	59.8/47.8/54.9	**99.3/98.5/96.6**	97.1/94.5/87.5	94.3/87.1/84.8	89.5/80.1/80.0	95.7/96.6/90.3	98.1/95.3/93.3
MetalPlate	**100.0/100.0/100.0**	**100.0/100.0/100.0**	**100.0/100.0/100.0**	65.3/83.6/85.5	97.3/99.9/98.6	99.7/**100.0/100.0**	**100.0/100.0/100.0**
Tubes	**98.6/99.5/97.8**	84.6/94.1/84.6	87.4/94.9/85.0	57.5/72.6/81.7	74.3/87.3/82.6	91.4/94.1/88.7	91.4/96.5/89.4
Mean	84.1/83.2/84.1	90.6/94.1/89.7	93.0/95.1/90.6	74.6/82.1/82.5	86.5/88.1/85.8	93.7/95.5/90.9	**97.7/98.2/95.7**

Table 6: Comparison with SOTA methods on MPDD dataset for multi-class anomaly localization with AU-ROC/AP/F1$_{\max}$/AU-PRO metrics. OSD-IRF results use the SNR-guided timestep $t^\star = 588$.

Category	RD4AD	SimpleNet	DeSTSeg	DiAD	HVQ-Trans	OmiAD	Ours
BracketBlack	97.4/9.5/19.0/92.1	94.9/3.6/9.1/89.3	95.1/9.5/**21.9**/49.0	91.5/1.1/2.9/74.7	96.1/1.1/2.7/87.3	98.1/3.7/8.4/94.8	**98.8/10.5**/19.4/**96.2**
BracketBrown	97.3/13.0/23.4/**94.8**	94.9/8.8/17.0/87.4	83.9/5.4/12.5/24.9	95.6/8.1/16.7/84.6	98.2/32.3/36.4/86.5	**98.9/42.8**/44.9/93.8	98.8/40.5/**45.5**/93.8
BracketWhite	**98.3**/2.6/6.4/**92.2**	97.8/2.2/5.6/86.4	95.8/4.6/12.3/63.1	90.8/0.5/1.6/77.4	94.1/0.6/2.6/79.8	**98.3**/7.3/18.8/90.8	97.8/**15.6/25.4**/90.7
Connector	97.4/12.5/20.6/90.6	98.9/56.0/55.2/96.5	96.2/30.0/35.5/61.7	98.1/25.1/30.7/93.5	97.9/18.8/29.7/92.7	98.9/32.3/41.2/96.3	**99.5/65.5/58.4/98.3**
MetalPlate	**99.3/95.9/89.7/94.5**	98.2/88.9/81.3/88.1	98.7/94.8/88.8/74.9	93.8/54.9/65.7/79.6	96.6/73.7/74.4/86.3	98.5/87.4/83.8/93.4	98.8/90.5/85.8/93.9
Tubes	99.1/**77.8/73.0**/96.6	97.9/42.4/46.1/92.6	94.8/55.2/54.7/85.4	88.3/5.9/9.5/60.4	97.0/38.8/42.8/88.9	98.7/52.3/56.6/94.8	**99.4**/68.7/69.1/**97.5**
Mean	98.1/35.2/38.7/93.4	97.1/33.6/35.7/90.0	94.1/33.2/37.6/59.8	93.0/15.9/21.2/78.4	96.7/27.6/31.4/86.9	98.6/37.6/42.3/94.0	**98.8/48.5/50.6/95.1**

Table 7: Comparison with SOTA methods on VisA dataset for multi-class anomaly detection with AU-ROC/AP/F1$_{\max}$ metrics. OSD-IRF results use the SNR-guided timestep $t^\star = 587$.

	Category	RD4AD	SimpleNet	DeSTSeg	DiAD	HVQ-Trans	OmiAD	Ours
Complex structure	PCB1	96.2/95.5/91.9	91.6/91.9/86.0	87.6/83.1/83.7	88.1/88.7/80.7	96.7/93.2/87.7	97.8/97.4/95.1	**99.5/99.5/97.1**
	PCB2	97.8/97.8/94.2	92.4/93.3/84.5	86.5/85.8/82.6	91.4/91.4/84.7	93.4/94.8/88.0	97.8/98.5/94.1	**99.6/99.6/97.5**
	PCB3	96.4/96.2/91.0	89.1/91.1/82.6	93.7/95.1/87.0	86.2/87.6/77.6	92.0/87.1/79.5	96.7/95.1/87.6	**98.6/98.7/92.9**
	PCB4	99.9/99.9/**99.0**	97.0/97.0/93.5	97.8/97.8/92.7	99.6/99.5/97.0	99.5/99.0/97.0	**100.0/100.0/99.0**	99.9/99.9/98.0
Multiple instances	Macaroni1	75.9/61.5/76.8	85.9/82.5/73.1	76.6/69.0/71.0	85.7/85.2/78.8	93.1/84.1/79.8	97.3/97.5/92.8	**98.3/98.3/93.2**
	Macaroni2	**88.3/84.5/83.8**	68.3/54.3/59.7	68.9/62.1/67.7	62.5/57.4/69.6	86.2/84.1/81.5	85.1/83.3/79.5	85.3/84.0/79.4
	Capsules	82.2/90.4/81.3	74.1/82.8/74.6	87.1/93.0/84.2	58.2/69.0/78.5	77.1/83.4/77.5	85.7/89.0/78.8	**92.1/94.3/90.2**
	Candles	92.3/92.9/86.0	84.1/73.3/76.6	94.9/94.8/89.2	92.8/92.0/87.6	96.8/98.0/93.1	97.4/**98.6**/93.4	**98.3**/98.4/**93.7**
Single instance	Cashew	92.0/95.8/90.7	88.0/91.3/84.7	92.0/96.1/88.1	91.5/95.7/89.7	94.9/96.8/90.4	93.3/96.4/90.9	**95.6/97.8/94.6**
	Chewinggum	94.9/97.5/92.1	96.4/98.2/93.8	95.8/98.3/94.7	99.1/99.5/95.9	99.4/99.6/97.5	99.2/99.8/97.5	**99.8/99.9/99.5**
	Fryum	95.3/97.9/91.5	88.4/93.0/83.3	92.1/96.1/89.5	89.8/95.0/87.2	90.4/94.5/84.9	94.0/96.5/88.5	**96.9/98.4/93.8**
	Pipefryum	97.9/98.9/96.5	90.8/95.5/88.6	94.1/97.1/91.9	96.2/98.1/93.7	98.5/98.4/94.0	98.9/99.4/**97.0**	**99.0/99.5/97.0**
	Mean	92.4/92.4/89.6	87.2/87.0/81.8	88.9/89.0/85.2	86.8/88.3/85.1	93.2/92.8/87.6	95.3/96.0/91.2	**96.9/97.4/93.9**

Table 8: Comparison with SOTA methods on VisA dataset for multi-class anomaly localization with AU-ROC/AP/F1$_{\max}$/AU-PRO metrics. OSD-IRF results use the SNR-guided timestep $t^\star = 587$.

	Category	RD4AD	SimpleNet	DeSTSeg	DiAD	HVQ-Trans	OmiAD	Ours
Complex structure	PCB1	99.4/66.2/62.4/95.8	99.2/**86.1/78.8**/83.6	95.8/46.4/49.0/83.2	98.7/49.6/52.8/80.2	99.4/63.1/58.8/87.4	**99.7**/70.8/66.2/93.0	**99.7**/74.6/68.3/**96.4**
	PCB2	98.0/22.3/30.0/90.8	96.6/8.9/18.6/85.7	97.3/14.6/28.2/79.9	95.2/7.5/16.7/67.0	98.0/10.1/18.2/82.0	**98.9**/16.6/22.2/87.1	**98.9/31.7/33.7/92.4**
	PCB3	97.9/26.2/35.2/**93.9**	97.2/**31.0/36.1**/85.1	97.7/28.1/33.4/62.4	96.7/8.0/18.8/68.9	98.3/21.1/23.7/80.5	99.1/29.7/30.4/87.2	**99.2**/28.4/30.8/**93.9**
	PCB4	97.8/31.4/37.0/88.7	93.9/23.9/32.9/61.1	95.8/**53.0/53.2**/76.9	97.0/17.6/27.2/85.0	97.7/21.1/29.8/85.9	98.1/42.0/44.1/86.4	**98.3**/39.9/43.2/**91.7**
Multiple instances	Macaroni1	99.4/2.9/6.9/95.3	98.9/3.5/8.4/92.0	99.1/5.8/13.4/62.4	94.1/10.2/16.7/68.5	99.4/9.9/19.3/91.2	**99.7**/20.1/**29.8**/96.3	**99.7/23.0**/29.6/**98.2**
	Macaroni2	**99.7/13.2**/21.8/**97.4**	93.2/0.6/3.9/77.8	98.5/6.3/14.4/70.0	93.6/0.9/2.8/73.1	98.5/5.5/13.8/91.3	99.4/8.5/15.4/94.0	99.3/13.1/**23.1**/95.8
	Capsules	99.4/60.4/60.8/93.1	97.1/52.9/53.3/73.7	96.9/33.2/9.1/76.7	97.3/10.0/21.0/77.9	99.0/51.9/55.2/76.3	99.4/62.7/59.6/90.8	**99.7/68.7/66.4/93.3**
	Candles	99.1/25.3/35.8/94.9	97.6/8.4/16.5/87.6	98.7/**39.9/45.8**/69.0	97.3/12.8/22.8/89.4	99.2/20.4/30.6/92.7	**99.4**/25.7/35.1/**97.1**	**99.4**/26.5/35.1/95.8
Single instance	Cashew	91.7/44.2/49.7/86.2	98.9/**68.9/66.0**/84.1	87.9/47.6/52.1/66.3	90.9/53.1/60.9/61.8	**99.2**/58.3/60.9/89.3	97.2/46.3/55.3/83.8	97.9/52.1/53.8/**90.5**
	Chewinggum	98.7/59.9/61.7/76.9	97.9/26.8/29.8/78.3	98.8/**86.9/81.0**/68.3	94.7/11.9/25.8/59.5	98.8/43.7/44.4/81.7	98.8/58.0/55.6/73.1	**99.1**/55.8/54.7/**86.9**
	Fryum	97.0/47.6/51.5/**93.4**	93.0/39.1/45.4/85.1	88.1/35.2/38.5/47.7	97.6/**58.6/60.1**/81.3	**97.7**/50.8/55.3/83.6	**97.7**/47.7/55.2/87.2	97.3/48.5/53.4/89.0
	Pipefryum	99.1/56.8/58.8/95.4	98.5/65.6/63.4/83.0	98.9/**78.8/72.7**/45.9	**99.4**/72.7/69.9/89.9	**99.4**/64.6/65.8/93.4	99.2/56.2/60.1/94.5	99.1/62.4/60.6/**95.7**
	Mean	98.1/38.0/42.6/91.8	96.8/34.7/37.8/81.4	96.1/39.6/43.4/67.4	96.0/26.1/33.0/75.2	98.7/35.0/39.6/86.3	98.9/40.4/44.1/89.2	**99.0/43.7/46.1/93.3**

A.2 More Qualitative Results

To further provide a visual assessment of the proposed method, we present qualitative anomaly localization results on MVTec-AD, VisA, and MPDD. As shown in Fig. 6, Fig. 7, and Fig. 8, each row corresponds to one category and contains three samples from that category. For each sample, the original image, ground-truth mask, and predicted anomaly map are displayed from left to right. The visual results show that our method can accurately highlight anomalous regions while suppressing responses in normal areas. Across different datasets, categories, texture patterns, and defect types, the predicted anomaly maps remain well aligned with the ground-truth masks, demonstrating the robustness of the proposed mean-path IRF formulation for anomaly localization.

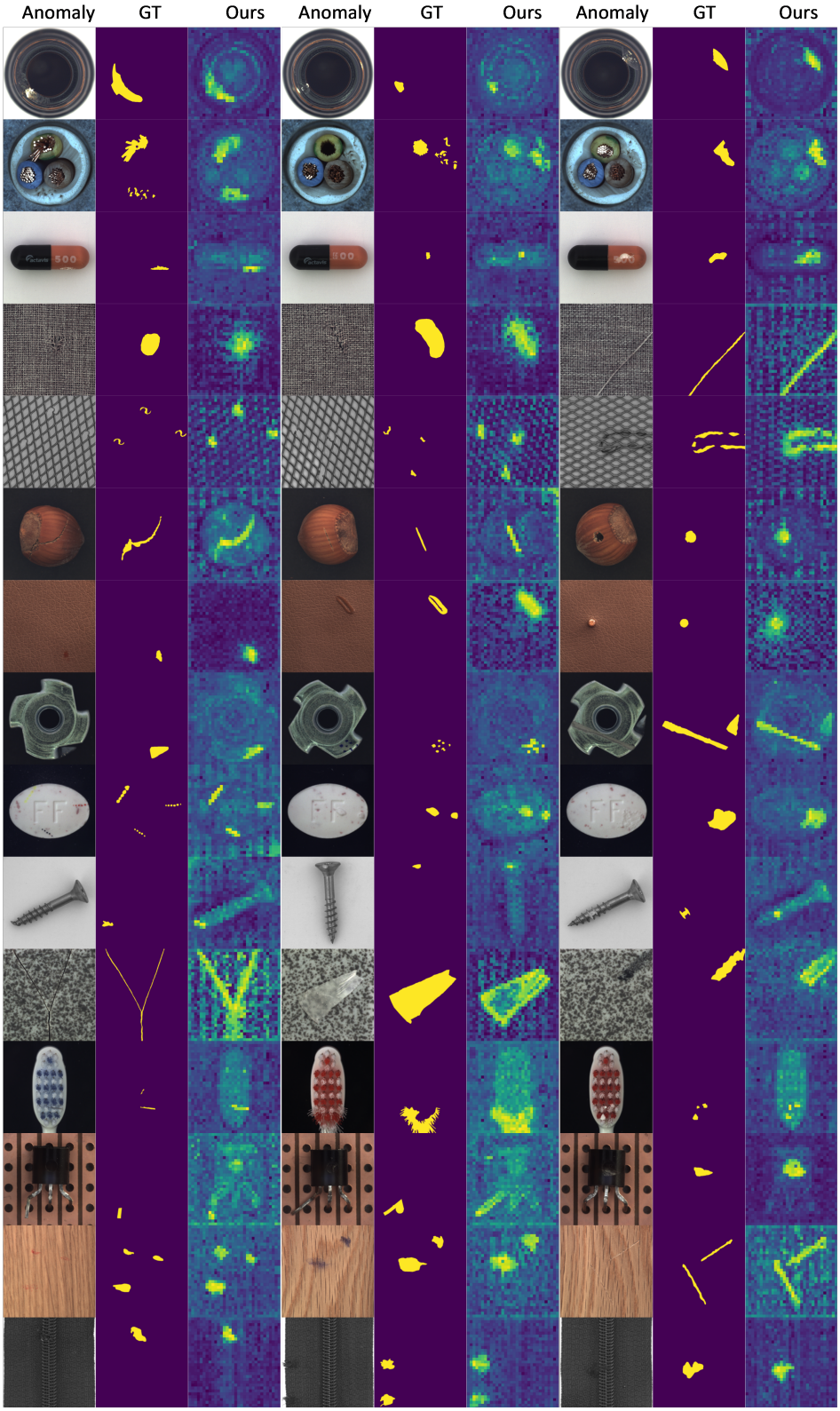


Figure 6: Qualitative anomaly localization results of our method on MVTec-AD. Each row corresponds to one category and contains three samples from that category. For each sample, the original image, ground-truth mask, and predicted anomaly map are displayed from left to right.

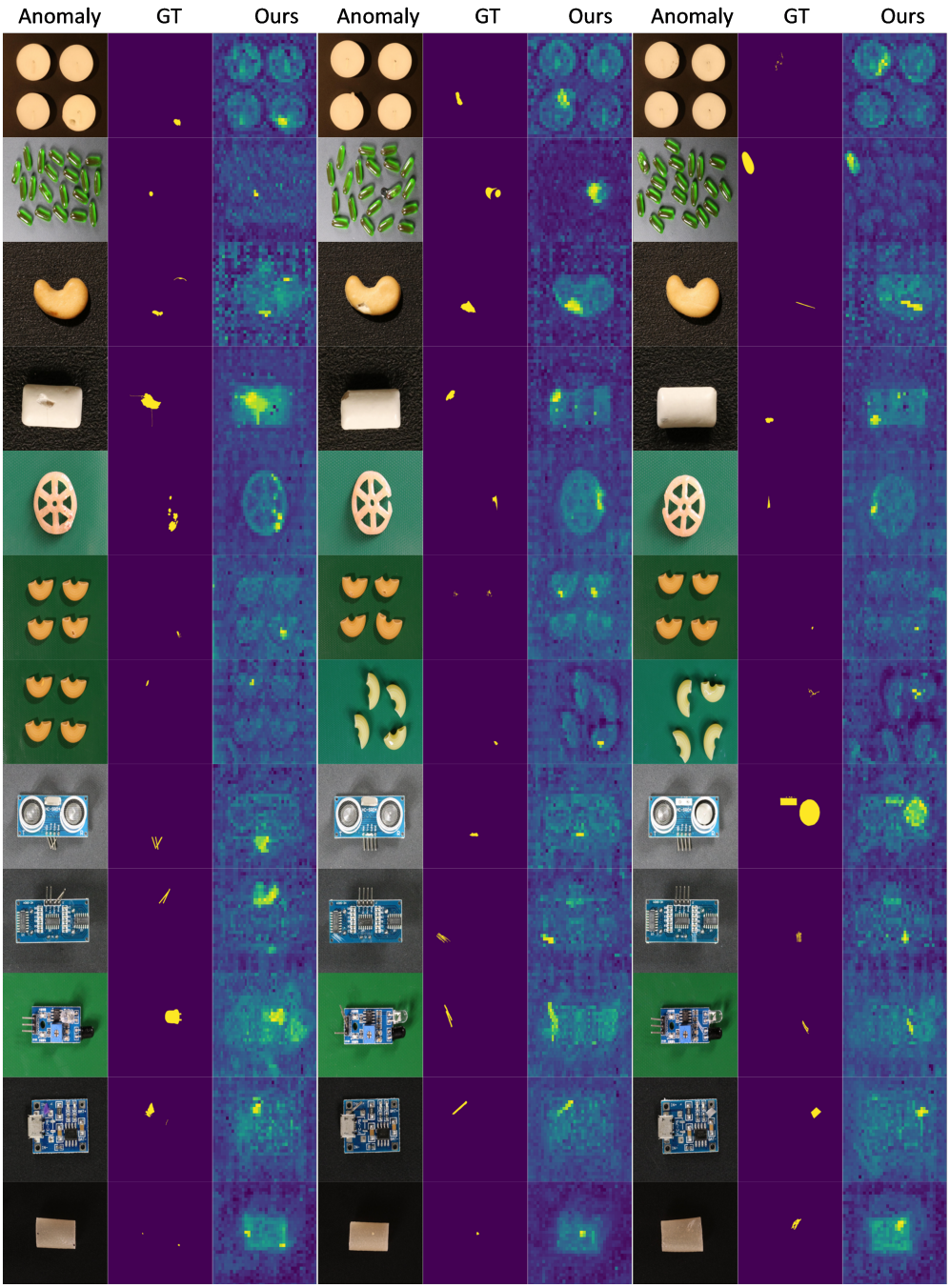


Figure 7: Qualitative anomaly localization results of our method on VisA. Each row corresponds to one category and contains three samples from that category. For each sample, the original image, ground-truth mask, and predicted anomaly map are displayed from left to right.

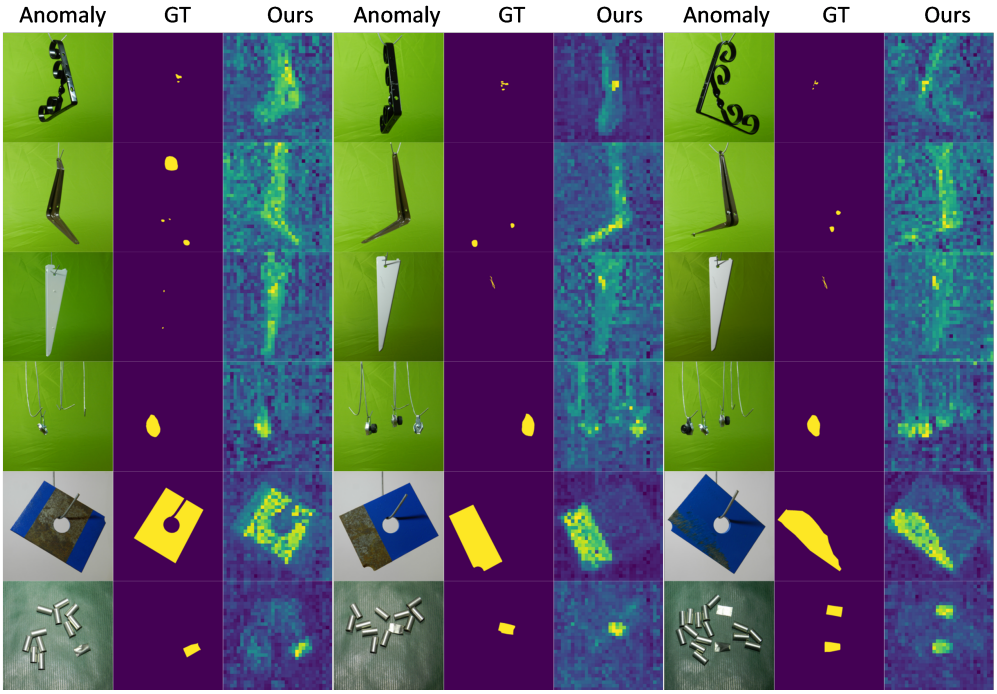


Figure 8: Qualitative anomaly localization results of our method on MPDD. Each row corresponds to one category and contains three samples from that category. For each sample, the original image, ground-truth mask, and predicted anomaly map are displayed from left to right.