This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

A Unified Latent Schrödinger Bridge Diffusion Model for Unsupervised Anomaly Detection and Localization

Shilhora Akshay[1,*] Niveditha Lakshmi Narasimhan[2] Jacob George[2]
Vineeth N Balasubramanian[1]
[1]Indian Institute of Technology, Hyderabad [2]KLA Corporation
[*]shilhora.akshay333@gmail.com

Abstract

*Anomaly detection and localization remain pivotal challenges in computer vision, with applications ranging from industrial inspection to medical diagnostics. While current supervised methods offer high precision, they are often impractical due to the scarcity of annotated data and the infrequent occurrence of anomalies. Recent advancements in unsupervised approaches, particularly reconstruction-based methods, have addressed these issues by training models exclusively on normal data, enabling them to identify anomalies during inference. However, these methods frequently rely on auxiliary networks or specialized adaptations, which can limit their robustness and practicality. This work introduces the **L**atent **A**nomaly **S**chrödinger **B**ridge (LASB), a unified unsupervised anomaly detection model that operates entirely in the latent space without requiring additional networks or custom modifications. LASB transforms anomaly images into normal images by preserving structural integrity across varying anomaly classes, lighting, and pose conditions, making it highly robust and versatile. Unlike previous methods, LASB does not focus solely on reconstructing anomaly features, but emphasizes anomaly transformation, achieving smooth anomaly-to-normal image conversions. Our method achieves state-of-the-art performance on both the MVTec-AD and VisA datasets, excelling in detection and localization tasks. Our code is available at https://github.com/ShilhoraAkshayPatel/LASB.*

1. Introduction

Anomaly detection and localization is a critical task in computer vision with several applications in medical [4, 18, 52, 53] and industrial [21, 31, 43, 60], which has attracted significant attention from the research community. The main goal of anomaly detection is to identify and localize abnormal patterns that are unusual from those seen in normal instances. In the past few years the research community has proposed various supervised anomaly detection methods [15, 24, 30, 44, 50, 59]. The extensive need for annotations is expensive and the infrequent presence of anomalous samples makes these methods unsuitable for practical applications due to their limitations in addressing real-world scenarios.

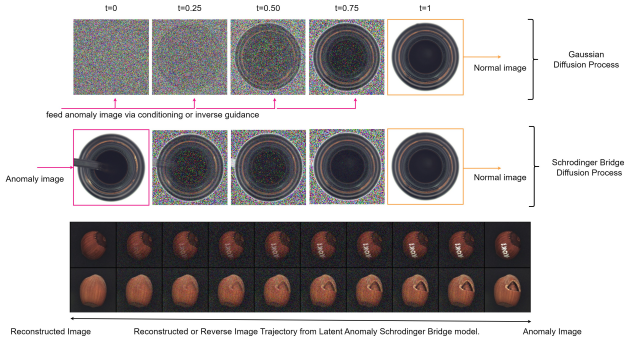


Figure 1. Illustration of the Gaussian and Bridge diffusion processes, with reverse image trajectory from LASB. LASB learns a direct diffusion bridge between anomaly and normal distributions, enhancing interpretability and anomaly-free transformation.

Recent research has sought to overcome the limitations of supervised anomaly detection methods by relying solely on normal images during training and localizing abnormal patterns at inference time. Among the various categories within the unsupervised paradigm, reconstruction-based methods have garnered significant attention due to their promising results and strong performance in real-world scenarios. The core concept behind these methods is that the model is trained exclusively on normal images and during inference, it reconstructs abnormal samples as normal samples. Thus, doing so benefits in two directions, one can interpret how the model's ability to detect and localize anomalies. Second, it aids the model to perform better while reconstructing the anomalies. Recent advancements have introduced novel approaches lever-

aging AutoEncoder-based (AE) [3, 6, 56–58], Generative Adversarial Networks-based (GANs) [1, 26, 45, 62], and diffusion-based approaches [4, 18, 21, 31, 43, 52, 60]. Our research specifically focuses on various diffusion models tailored for anomaly detection tasks.

Existing diffusion-based methods primarily focus on developing novel noise-conditioning techniques, often incorporating an additional discriminative sub-network [60]. Another approach utilizes a score-based diffusion model to identify anomalies by evaluating how effectively samples can return to the normal data distribution after perturbation, though it fails to deliver competitive performance [43]. Additionally, DiAD [21] proposes novel semantic guidance to specifically understand the anomalies and showcase its robustness. However, earlier diffusion-based methods generally depend on auxiliary networks or tailored diffusion processes to extract anomaly features. Moreover, most of these methods prioritize the extraction and reconstruction of anomaly features.

In this work, we introduce an intuitive mechanism that neither requires model-specific adaptations to extract anomaly features nor relies on an additional feature extractor for training or inference, to extract discriminative features. By leveraging the capabilities of Schrödinger Bridge [8, 11, 14, 42, 51], we propose the Latent Anomaly Schrödinger Bridge (LASB), which operates entirely in the latent space, transforming anomalous images into normal ones, regardless of anomaly class, and demonstrating robustness to variations in lighting and pose. In the following, we outline the key contributions of our study.

- We propose LASB, a unsupervised bridge-based anomaly detection model that operates in latent space. LASB transforms anomaly images into normal images, regardless of the anomaly class.
- **Unified Framework:** LASB offers a comprehensive framework for both anomaly detection and localization without the need for auxiliary networks. Unlike conventional methods that rely on Gaussian diffusion, LASB employs a bridge-based diffusion process that inherently preserves structural integrity, marking the first application of such a process in latent diffusion model for anomaly detection.
- **Efficient and Scalable:** By utilizing the Linear Schrödinger Bridge in latent space, LASB significantly reduces training time, memory consumption, and sampling speed, enhancing overall efficiency and scalability.
- Our method achieves state-of-the-art performance on the MVTec dataset, with a image-level AUROC$_{cls}$/AP$_{cls}$ of **99.2%/99.3%** and an pixel-level AUROC$_{seg}$/AP$_{seg}$ of **98.6%/78.2%** on the test set, establishing a new benchmark for unsupervised anomaly detection.

2. Related Work

Reconstruction-based Methods: Reconstruction-based anomaly detection models operate on the premise that networks trained exclusively on normal images will fail to accurately reconstruct anomalous ones due to their unfamiliarity with the abnormal distribution. Notably, autoencoder [6, 10, 28] and Generative Adversarial Network (GAN) [1, 41] frameworks have been widely utilized for this task, where anomaly scores are computed based on the reconstruction error between the input and its generated counterpart. Nevertheless, such approaches frequently struggle with direct copy issues and elevated false positive rates due to their limited generalization capacity over anomalous regions or due to the overgeneralization capacity of reconstructive methods. To mitigate these generalization challenges, recent innovations have pivoted towards inpainting strategies [35, 57] or integrating supplementary memory modules [17, 47]. Furthermore, the DRAEM [56] model enhances performance by leveraging pseudo-anomalies and coupling autoencoders with a segmentation network, though its effectiveness diminishes when faced with substantial deviations between real and pseudo anomalies.

Diffusion-based Methods: Diffusion models, known for their high-fidelity image synthesis, are increasingly applied to anomaly detection in both medical and industrial domains [18, 21, 43, 52, 60]. In the medical field, SANO [18] leverages score-based diffusion models to localize skin conditions such as eczema, while AnoDDPM [52] employs the Denoising Diffusion Probabilistic Model (DDPM) [23] to segment tumors, initiating with simplex noise. These methods primarily utilize standard DDPM [23] or Score-based Generative Models (SGMs) [46], yet they often struggle to capture structural features because they rely on starting from pure Gaussian or simplex noise and frequently require external guidance for complex feature integration. In industrial applications, DiffAD [60] introduces a noise interpolation technique coupled with a discriminative sub-network to enhance detection capabilities. AD-SPR [43] emphasizes the significance of perturbed resilience using SGMs [46] to identify anomalies. Additionally, DiAD [21] presents a novel semantic guidance network that directs a stable diffusion model [37] to effectively detect and localize anomalies in both industrial and selected medical imaging applications. However, these models often require class-specific training or additional networks to guide the diffusion process or to perform further segmentation via a discriminative sub-network.

Unified/Multi-class Methods: Early approaches in the literature developed models on a per-class basis, which limited their generalizability. More recent research has shifted towards unified or multi-class models that are trained on

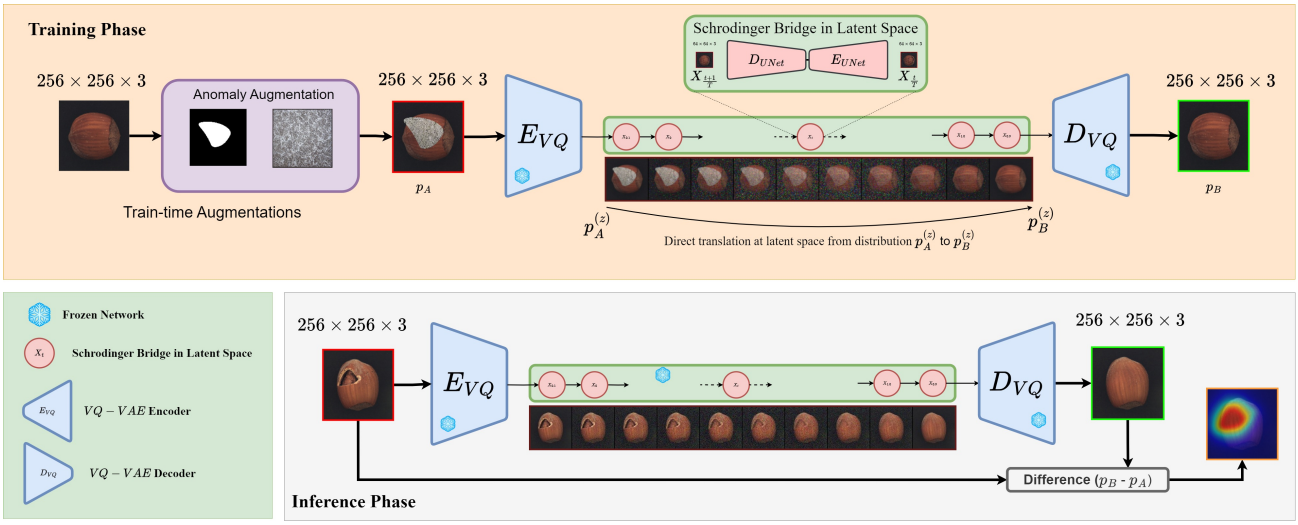


Figure 2. The LASB model framework consists of two key stages: *training* and *inference*. During training, anomaly augmentations are applied to images, introducing distortions that the LASB model learns to remove, ultimately reconstructing a normal image. This iterative process continues until the model effectively filters out the anomalies. In the inference stage, the model processes real anomalous images, reconstructing normal versions. The anomaly detection is achieved by computing the difference ($p_B - p_A$) between the original and reconstructed images, with anomalies visualized via a heatmap.

entire datasets, demonstrating improved robustness [21, 22, 32, 55]. For instance, UniAD [55] utilizes layer-wise query decoding, neighbor-masked attention, and feature jittering techniques to enhance multi-class anomaly detection. In contrast, HVQ-Trans [32] employs a hierarchical vector quantized prototype-oriented Transformer to improve discriminative capacity across multiple classes. Both methods focus on tackling the *"identical shortcut"* issue. On the other hand, RLR [22] introduces a novel framework using learnable reference representations with locality constraints to explicitly learn normal patterns and circumvent *"learning shortcuts"*, achieving superior results on standard datasets (MVTec-AD and VisA). Additionally, the DiAD [21] model, designed for multi-class settings and based on diffusion-based reconstruction, outperforms UniAD [55], RLR [22], and HVQ-Trans [32] in terms of localization and detection performance.

3. Proposed Methodology

3.1. Preliminaries

Before introducing the full methodology, we begin with a preliminary overview of score-based generative models (SGMs). Next, we explain the Schrödinger Bridge concept and discuss its connection to SGMs. Finally, we present the core methodology, detailing how and why the latent Schrödinger Bridge effectively addresses anomaly detection and localization.

Notations. Here we introduce some notations and use them throughout the work. Let $X_t (\in \mathbb{R}^d)$ denote a stochastic process, where $t \in [0, 1]$ is a continuous time step. The intermediate steps are uniformly distributed in the interval ($\mathcal{U}(t \sim [0, 1])$). The initial distribution is the corrupted data distribution is denoted as p_A and the terminal distribution is clean data distribution and is denoted as p_B. The Wiener process and its reversed counterpart, adopted from Anderson *et al.* [2], are represented as W_t and $\overline{W}_t$, respectively, both in $\mathbb{R}^d$. $\mathbb{I} \in \mathbb{R}^{d\times d}$ is the identity matrix.

Score-based Generative Models. Score-based Generative Models (SGMs) perturb the data stochastically across continuous noise scales and then use reverse-time stochastic differential equations (SDEs) to learn the reverse-time diffusion process. This reverse-time SDE relies on a score function enabling the reconstruction of any distribution starting from a Gaussian [46]. Given data $X_0 \sim p_A$, forward and backward SDEs are formulated as:

$$dX_t = f_t(X_t)\, dt + \sqrt{\beta_t}\, dW_t, \quad X_0 \sim p_A, \tag{1}$$

$$dX_t = [f_t - \beta_t \nabla \log p(X_t, t)]\, dt + \sqrt{\beta_t}\, d\overline{W}_t, \tag{2}$$

where, $f(\cdot, t) : \mathbb{R}^n \to \mathbb{R}^n$ and the terminal distributions (i.e., at $t = 1$) approach Gaussian ($X_1 \sim \mathcal{N}(0, I)$). To achieve this, diffusion coefficient $\beta_t \in \mathbb{R}$ is carefully tuned and ensuring that the base drift f_t is linear in X_t. Here, p is the marginal density of equation (2) at time t, and $\nabla \log p$ denotes its *score* [46].

SGMs as Schrödinger Bridge. The absence of flexibility in SGM's to transport data to desired distribution, demands a versatile strategy. The Schrödinger Bridge (SB) is a strategy often applied in optimal transport to find the optimal path measure between two marginal densities and it is expressed as,

$$\min_{Q \in P(p_A, p_B)} \mathrm{KL}(\mathbb{Q} \,||\, \mathbb{P}), \tag{3}$$

Here, $\mathbb{Q}$ represents a path measure within $\mathbb{P}(p_A, p_B)$, characterized by having marginal densities p_A and p_B at times $t = 0$ and 1, respectively. We consider $\mathbb{P}$ as the reference measure, specifically chosen as the path measure in equation (1). We will further elaborate on the conditions that define the optimality of the SB (as shown in equation (3)) with specified boundary conditions. The optimality condition for SB is characterized by solving PDEs [9][8]. Let $\Psi(t, x)$ and $\widehat{\Psi}(t, x)$ be the solutions to the following PDEs:

$$\frac{\partial \Psi(z,t)}{\partial t} = -\nabla \Psi^T f - \frac{1}{2}\beta \Delta \Psi \tag{4}$$

$$\frac{\partial \widehat{\Psi}(z,t)}{\partial t} = -\nabla \cdot (\widehat{\Psi} f) + \frac{1}{2}\beta \Delta \widehat{\Psi} \tag{5}$$

subject to the conditions,

$$\Psi(z,0)\widehat{\Psi}(z,0) = p_A(z), \quad \Psi(z,1)\widehat{\Psi}(z,1) = p_B(z).$$

Then, the solution to optimization (3) can be expressed by the path measure of the following forward equation (6), or equivalently backward equation (7), SDE:

$$dX_t = [f_t + \beta_t \nabla \log \Psi(X_t, t)]\, dt + \sqrt{\beta_t}\, dW_t, \tag{6}$$

$$dX_t = [f_t - \beta_t \nabla \log \widehat{\Psi}(X_t, t)]\, dt + \sqrt{\beta_t}\, d\overline{W}_t, \tag{7}$$

where $\nabla \log \Psi$ and $\nabla \log \widehat{\Psi}$ represent the non-linear optimal forward and backward drifts for the Schrödinger Bridge. This SB has a *non-linear* behavior and it generalizes the SGMs with variation in prior data. Note that, forward and backward equations (6, 7) of Schrödinger Bridge are same as SGMs forward-backward equations (1, 2) except the forward drift term ($\nabla \log \Psi$). Thus, from the Nelson's Duality [34] we obtain $\Psi(x,t)\widehat{\Psi}(x,t) = q^{\text{SB}}(x,t)$. As the backward drift of Schrödinger Bridge does not act as a score function anymore with two indivisible components. Thus, from equation (7) and Nelson's Duality [34] we obtain $dX_t = [f_t - \beta_t(\nabla \log q^{\text{SB}}(X_t, t) - \nabla \log \Psi(X_t, t))]\, dt + \sqrt{\beta_t}\, d\overline{W}_t$.

Linear-SB. The constraints posed above cause the model to be intractable and to overcome this impediment existing literature suggests various lines of strategies such as Iterative Proportional Fitting [11, 25], likelihood-based training of SB [8] etc. These methods help the SGMs to construct non-linear diffusion bridges. But, we can make them tractable by posing some linear conditions. A more detailed examination of SB theory with SGM framework reveals that the nonlinear drifts specified in equations (6, 7) correspond to the score functions described in equation (2). Assuming that $\Psi(\cdot, t)$ and $\widehat{\Psi}(\cdot, t)$ function as probability density functions, we reformulate equations (4, 5) to effectively address the solutions of the Fokker-Planck equation [36].

The Schrödinger Bridges in equation (4, 5) are satisfied, the backward and forward drifts, $\nabla \log \widehat{\Psi}(X_t, t)$ and $\nabla \log \Psi(X_t, t)$ represent the score functions for the following linear SDEs, respectively [36][27]:

$$dX_t = f_t(X_t)\, dt + \sqrt{\beta_t}\, dW_t, \quad X_0 \sim \Psi(\cdot, 0), \tag{8}$$

$$dX_t = f_t(X_t)\, dt + \sqrt{\beta_t}\, d\overline{W}_t, \quad X_1 \sim \widehat{\Psi}(\cdot, 1). \tag{9}$$

The key characteristic of the linear SDEs presented in equations (8, 9) lies in their differing boundary conditions, resulting in distinct distributions compared to nonlinear SDEs. Sampling is facilitated by parameterizing $\nabla \log \widehat{\Psi}$ using a score network and applying established SGM methods. However, the computational complexity arises due to the intractability of the boundary conditions $\Psi(\cdot, 0)$ and $\widehat{\Psi}(\cdot, 1)$. To resolve this, we introduce Dirac delta functions as boundary conditions for these previously intractable terms, thus rendering them manageable. Now to address this we impose Dirac delta boundary conditions [36][27]. Assume $p_A(\cdot) = \delta_\kappa(\cdot)$ is Dirac delta distribution that is centered at $\kappa \in \mathbb{R}^d$. Then, the initial distributions of the linear SDEs (equation (8,9)) are:

$$p_B = \Psi(\cdot, 1)\widehat{\Psi}(\cdot, 1), \quad \widehat{\Psi}(\cdot, 0) = \delta_\kappa(\cdot), \tag{10}$$

The above equation (10) suggests that the optimal backward drift of the process in equation (8) consistently targets the Dirac delta $\delta_\kappa(\cdot)$, achieving convergence to κ independent of p_B. This concept is integrated into the loss function, where the score is recalculated as $\nabla \log p(X_t, t|X_0 = \kappa)$ for each instance κ. This approach enhances computational efficiency and establishes a robust mathematical basis for training $\nabla \log \widehat{\Psi}(\cdot)$.

3.2. Linear Schrödinger Bridge in the Latent Space

We now describe how the above discussion on linear Schrödinger bridges can be used for anomaly detection and localization.

Overall Pipeline. Figure 2 illustrates the entire pipeline of the proposed method. Initially, anomaly augmentation [56] is applied to a normal image, which is then processed by the Latent Anomaly Schrödinger Bridge (LASB) model. The LASB model learns the anomaly statistics and transforms the augmented anomaly image back into a normal image. Crucially, the aim of this step is not to learn the anomalies themselves, but rather to reconstruct the normal image regardless of the anomalies introduced during training. During training, various masks and noise patterns are applied to augment the images. After training, the LASB model is evaluated on anomaly images from the test set, with its weights frozen. During the inference phase, the model is presented with an anomaly image from the test

Algorithm 1 Training Procedure

Input: normal and anomaly latent $p_A^{(z)}(\cdot), p_B^{(z)}(\cdot|z_0)$
repeat
  $z_0 \sim p_A^{(z)}(z_0), z_1 \sim p_B^{(z)}(z_1|z_0)$
  $z_t \sim q(z_t|z_0, z_1)$ according to equation (11)
  calculate gradient $\nabla\, \epsilon(z_t, t; \theta)$ using equation (12)
until convergence

Algorithm 2 Inference/Sampling Procedure

Input: $z_N \sim p_B^{(z)}(z_N)$, trained $\epsilon_\theta(\cdot, \cdot)$
for $n = N$ **to** 1 **do**
  Predict z_0^ϵ using $\epsilon_\theta(z_n, t_n)$
  $z_{n-1} \sim p(z_{n-1}|z_0^\epsilon, z_n)$.
end for
return z_0

set, which it then reconstructs as a normal image. Both the original anomaly image and its reconstructed normal image are subsequently passed through a difference block. In this step, multi-scale features are extracted using a pre-trained ImageNet model, and the discrepancies between these features are computed to pinpoint regions of interest (ROI). These differences are then detected and localized, following a strategy akin to DiAD [21].

Latent Anomaly Schrödinger Bridge. The foundational basis of Latent Anomaly Schrödinger Bridge (LASB) is inspired by linear Schrödinger Bridge's from equation (10) i.e., the optimal backward drift in equation (8) consistently aims for the Dirac delta $\delta_\kappa(\cdot)$, ensuring convergence to κ regardless of p_B. That means, the forward process transforms the noisy anomaly image statistics and learns to reconstruct a normal image irrespective of terminal distribution. Now to elucidate the working mechanism of the LASB diffusion model which will consider a Chang *et al.* [7] design strategy [1].

SGMs and diffusion models operating in pixel space require significant computational resources and memory. To address this challenge, we introduce the LASB model. Before training the Linear-SB, we first train E_{VQ} and D_{VQ}, the encoder and decoder of the VQ-VAE model [49], for perceptual compression, following the approach of stable diffusion [37]. During VQ-VAE training, we input either normal or anomaly-augmented images from the training set into E_{VQ} and aim to reconstruct the input using D_{VQ}. Once the model converges, its weights are frozen, and we proceed to train the Linear-SB in the latent space. Unlike Gaussian diffusion models, where noise is progressively added at each time step, leading to the deterioration of image structure at later stages, we adopt a semi-degradation strategy similar to Saharia *et al.* [40]. This ensures that, during the forward pass, the LASB model performs a smooth transformation while preserving structural information. Figure 1 illustrates the contrast between the Gaussian diffusion process and the Schrödinger Bridge diffusion process when applied to anomaly-free reconstruction. According to Liu *et al.* [27] for this smooth structural transformation the posterior for the SB (equation 6, 7) given a boundary pair condition (X_0, X_1) reveals an analytic. Now as we inject the image statistics via perceptual compression using VQ-VAE into latent space, the analytical form can be formulated as,

$$q(z_t|z_0, z_1) = \mathcal{N}(z_t, \mu_t(z_0, z_1), \Sigma_t) \tag{11}$$

where $\mu_t = \frac{\bar{\sigma}_t^{(z)2}}{\bar{\sigma}_t^{(z)2}+\sigma_t^{(z)2}} z_0 + \frac{\sigma_t^{(z)2}}{\bar{\sigma}_t^{(z)2}+\sigma_t^{(z)2}} z_1$, and $\Sigma_t = \frac{\bar{\sigma}_t^{(z)2}\sigma_t^{(z)2}}{\bar{\sigma}_t^{(z)2}+\sigma_t^{(z)2}}\mathbb{I}$.

The accumulated variances are, $\sigma_t^{(z)2} := \int_0^t \beta_\tau\, d\tau$ and $\bar{\sigma}_t^2 := \int_t^1 \beta_\tau\, d\tau$.

We hence train the proposed LASB using our objective function:

$$\mathcal{L}_{LASB} := \left|\left|\epsilon_\theta(z_t, t) - \left(\frac{z_t - z_0}{\sigma_t^{(z)}}\right)\right|\right| \tag{12}$$

where $\epsilon(\cdot)$ is the network and θ are the parameters associated with it. The training and the inference procedures are summarized in Algorithms 1 and 2.

4. Experiments and Results

Datasets. To validate the efficacy of the LASB method, we utilize two challenging datasets in industrial anomaly detection: the MVTec-AD [5] dataset and the VisA [63] dataset. The MVTec-AD dataset comprises 15 categories, including 10 object types and 5 texture types, with a total of 5,354 high-resolution images. Of these, 3,629 anomaly-free images are used for training, while the remaining 1,725 test images include both normal and anomalous samples. The VisA dataset consists of 12 distinct objects organized into 12 subsets, categorized into complex Structures, Multiple Instances, and Single Instance types. It includes 10,821 high-resolution images, with 9,621 normal images and 1,200 anomalous images exhibiting 78 distinct anomalies, offering a comprehensive benchmark for anomaly detection and localization methods. Both datasets provide pixel-level ground truth annotations to facilitate the evaluation of anomaly localization performance.

Evaluation Metrics. For a rigorous quantitative evaluation of our experimental results in anomaly detection and localization, we employ AUROC (Area Under the Receiver Operating Characteristic Curve), AP (Average Precision),

[1] Chang *et al.* [7] work focuses on explaining diffusion models by signifying the design principles of the training strategy, sampling strategy, and the objective function.

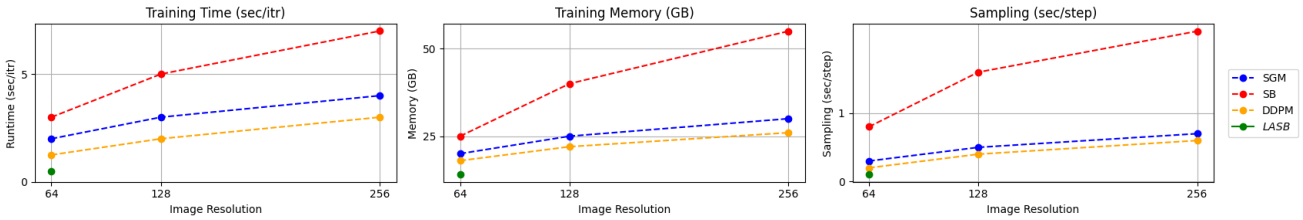


Figure 3. Comparison of training time (left), training memory (middle), and sampling time (right) across different image resolutions for SGM, SB, DDPM, and LASB models. The LASB model (green) demonstrates consistently lower training time, memory usage, and sampling time, particularly in contrast to the SB model (red), which exhibits the highest resource consumption.

and F1max (maximum F1-score) as our primary metrics [21]. Here, *cls* denotes image-level anomaly detection, while *seg* pertains to pixel-level anomaly localization. A detailed class-specific comparison and results for all the aforementioned metrics are provided in Section S2 of supplementary material. In Table 1, we present the average performance of each model across the evaluated dataset.

Method	Venue	Adopted Method	MvTec	ViSA
		Class-based		
SimpleNet [29]	CVPR 2023	Embedding-based	99.6 / 98.1	96.2 / 98.5
PatchCore [38]	CVPR 2022	Embedding-based	99.1 / 98.1	91.0 / 98.1
DSR [58]	ECCV 2022	-	98.2 / 70.2	-
PaDiM [12]	ICPR 2021	Memory Bank	97.5 / 92.1	89.1 / 85.9
CS-Flow [39]	WACV 2022	Normalization Flow	97.2 / 84.5	-
CFLOW-AD [19]	WACV 2022	Normalization Flow	96.2 / 97.1	-
OCR-GAN [26]	TIP 2023	GAN	98.3 / -	97.9 / -
DRAEM [56]	ICCV 2021	AutoEncoder	98.0 / 97.3	88.7/93.5
RD4AD [13]	CVPR 2022	Embedding-based	98.5 / 97.8	96.9 / 98.3
ADSPR [43]	ICCV 2023	Diffusion	97.67 / 97.36	-
DiffAD [60]	ICCV 2023	Diffusion	98.72 / 98.26	89.79 / 68.28
D3AD [48]	CVPR 2024	Diffusion	97.15 / 97.44	95.51 / 94.27
DDAD [33]	Arxiv 2023	Diffusion	99.84 / 98.05	98.9 / 97.58
TransFusion [16]	ECCV 2024	Diffusion	99.24 / 94.33	98.53 / 86.26
LASB (Ours)	-	Diffusion	99.66 / 99.15	98.52 / 99.06
		Unified (multi-class)		
DRAEM [56]	ICCV 2021	AutoEncoder	88.1 / 87.2	79.1 / 91.3
HVQ-Trans [32]	NeurIPS 2023	Non-Diffusion	98.0 / 97.3	93.2 / 98.7
MambaAD [20]	NeurIPS 2024	Non-Diffusion	98.6 / 97.7	94.3 / 98.5
OmniAL [61]	CVPR 2023	Non-Diffusion	97.2 / 98.3	94.2 / 96.0
GLAD [54]	ECCV 2024	Diffusion	97.5 / 97.4	91.8 / 97.8
UniAD [55]	NeurIPS 2022	Transformer	96.52 / 96.8	85.5 / 95.925
DIAD [21]	AAAI 2024	Diffusion	97.15 / 96.76	86.75 / 96.04
LASB (Ours)	-	Diffusion	99.14 / 98.66	94.2 / 98.18

Table 1. Comparison of state-of-the-art (SOTA) anomaly detection methods categorized into class-based and unified (multi-class) approaches. The table highlights the adopted methodologies and their class-wise average performance in terms of AUROC_{cls} / AUROC_{seg} metrics on two datasets: MvTec and ViSA respectively. An exhaustive list of metrics along with class specific performance details is comprehensively illustrated in Section S2 of supplementary material.

Implementation Details. For all our experiments, we resized MVTec-AD images to 256×256 resolution. To fine-tune the VQ-VAE (Auto-encoder) utilizing the KL-based method, we initialized the network weights from the Stable Diffusion [37] model, where the image is encoded into a latent vector with a size of $64 \times 64 \times 3$. We chose this latent vector size due to its preservation of details and low FID generation capabilities. After training the auto-encoder, we froze the weights and trained the latent denoising network. We used the U-Net model as in [23] as our latent denoising network. During inference, we pass only the anomaly test image as input to the DDPM sampler [23]. Further details on batch size and various other network hyperparameters are in Section S4 of the appendix.

Baselines. To benchmark LASB, we provide a detailed comparison with *state-of-the-art* (SOTA) reconstruction methods, as summarized in Table 1. These methods are categorized into two distinct groups: class-based and unified (multi-class) approaches. Class-based methods train models specific to each class, with the number of models increasing linearly with the number of classes, making scalability a challenge. In contrast, unified approaches handle multiple classes within a single model, offering a more scalable solution. By separating the class-based and unified approaches, we ensure a fair and comprehensive comparison.

Class-based Methods: Performance Analysis. Among the class-based methods, PaDiM [12], with its memory bank-based architecture, achieves 97.5% on MVTec-AD and 89.1% on ViSA. While effective for simpler anomalies, its scalability and ability to model inter-class variations remain limited. In contrast, DiffAD [60] demonstrates strong competitiveness, achieving 98.72% on MVTec-AD. However, its performance on ViSA drops to 89.79%, indicating its difficulty in adapting to diverse industrial anomalies. DDAD [33] shows highly competitive results, achieving 99.84% on MVTec-AD and 98.9% on ViSA. Its diffusion-based modeling effectively captures structural patterns, making it one of the closest competitors to LASB. However, LASB surpasses DDAD on ViSA by a margin of **0.38%**↑, demonstrating superior adaptability to multi-class industrial scenarios.

Task	Metrics	Pixel Space			Latent Space		
		DDPM [23]	DDAD [33]	GLAD [54]	LDM [37]	DiAD [21]	LASB (Ours)
Detection	AUROC_{cls}	71.9	**99.8**	97.5	76.6	97.2	99.2 (-0.6%)
	AP_{cls}	81.6	**99.5**	99.1	87.6	99.0	99.3 (-0.2%)
	F1max_{cls}	86.6	97.9	96.6	88.1	96.5	**98.5** (+2.0%)
Localization	AUROC_{seg}	75.6	98.0	97.4	85.1	96.8	**98.6** (+0.60%)
	AP_{seg}	13.3	59.0	60.8	27.6	52.6	**78.2** (+17.4%)
	F1max_{seg}	19.5	59.4	60.7	31.0	55.5	**70.7** (+10.0%)

Table 2. Performance comparison of LASB with diffusion-based models on MVTec-AD using AUROC, AP, and F1max. The best results are in **bold**; the second-best is underline. Improvements are shown as a percentage over the second-best.

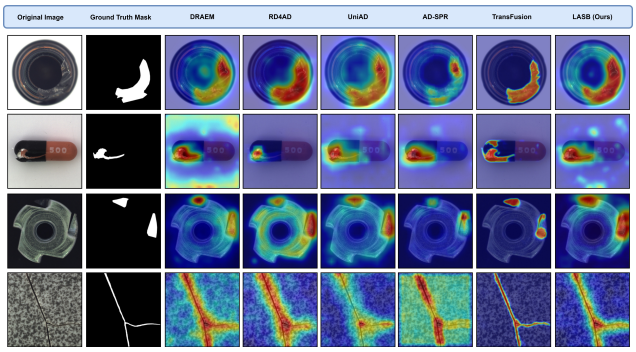


Figure 4. Visual representation of test samples from the MVTec dataset, depicted through heatmaps for various models for different categories.

Unified Models: Performance Analysis. For unified approaches, DRAEM [56] emerges as one of the least competitive models, with AUROC$_{cls}$ scores of 88.1% on MVTec-AD and 79.1% on ViSA. Its reliance on autoencoder-based reconstruction, combined with limited augmentation strategies, hampers its ability to generalize to diverse and complex anomaly scenarios, particularly in the multi-class ViSA dataset. Similarly, UniAD [55] performs less competitively, achieving 96.52% on MVTec-AD and 85.5% on ViSA. Its transformer-based architecture, while robust for certain tasks, struggles to balance efficiency and accuracy in anomaly detection across diverse classes. Similarly, DIAD [21] achieves 97.15% on MVTec-AD and 86.75% on ViSA but falls short in handling the intricate class variations present in the datasets. Among the competitive unified methods, MambaAD [20] achieves 98.6% on MVTec-AD and 94.3% on ViSA, showcasing its strength in handling multi-class scenarios using state-space models. However, LASB outperforms MambaAD with a score of 99.14% on MVTec-AD and 94.2% on ViSA, highlighting its superior capability to balance computational efficiency with detection accuracy. Furthermore, HVQ-Trans [32], with scores of 98.0% on MVTec-AD and 93.2% on ViSA, demonstrates solid performance but remains slightly behind LASB in scalability and robustness. Detailed class-specific results are available in Section S2 of supplementary material, and Figure 4 provides qualitative heat-map visualizations of anomaly regions, further demonstrating LASB's efficacy in both class-based and unified settings.

Task	Metrics	Non-Diffusion Method		Diffusion-based Method		
		DRAEM [23]	UniAD [55]	DiffAD [60]	DiAD [21]	LASB (Ours)
Detection	AUROC$_{cls}$	79.1	85.5	89.5	86.8	**94.2**
	AP$_{cls}$	81.9	85.5	-	88.3	**92.2**
	F1max$_{cls}$	78.9	84.4	-	85.1	**94.5**
Localization	AUROC$_{seg}$	91.3	95.9	71.2	96.0	**98.2**
	AP$_{seg}$	23.5	21.0	-	26.1	**46.4**
	F1max$_{seg}$	29.5	27.0	-	33.0	**52.6**

Table 3. Results for multi-class anomaly detection and localization on VisA dataset. The best results are indicated in **bold**, and the second-best results are denoted with an underline.

5. Ablation Studies and Analysis

In this section, we first differentiate the proposed LASB method's performance in latent spaces and explain the advantages of utilizing Linear-SB in the latent space. We then assess the LASB model's effectiveness when compared with Stable Diffusion [37] and other models for anomaly localization and detection tasks. Finally, we demonstrate the robustness of our proposed method, demonstrating its ability to deliver precise and consistent outcomes during test-time evaluations.

LASB vs Standard Diffusion Models. In the realm of anomaly detection, standard diffusion models such as DDPM [23] and LDM [37] achieve competitive performance in anomaly detection tasks but exhibit significant limitations in localization, as evidenced in Table 2. To mitigate these limitations, DiAD [21] integrates a semantic guidance network to improve LDM's localization by incorporating additional contextual information. Despite these improvements, DiAD [21] still falls short of achieving state-of-the-art (SOTA) results, indicating persistent challenges in fully leveraging diffusion models for comprehensive anomaly detection tasks. In contrast, our LASB employs a novel approach that leverages Linear-SB within the latent space, avoiding the limitations associated with reconstructing from pure Gaussian noise, a common issue in standard diffusion processes when applied to anomaly-free reconstruction. Moreover, LASB semi-degrades latent space to retain structural integrity, facilitating more effective anomaly detection and localization. This method not only enhances the robustness of the detection process but also significantly improves localization accuracy. Consequently, the LASB model surpasses standard diffusion models across multiple metrics, demonstrating superior performance in multi-class anomaly detection, as evidenced in Table 2.

We compare the performance of our proposed Latent Anomaly Schrödinger Bridge (LASB) method with the DDPM [23], DDAD [33], GLAD [54], LDM [37], and DiAD [21] models to demonstrate the effectiveness and efficiency of our approach. First, we evaluate the models on the MVTec-AD dataset for both anomaly detection and localization tasks. As shown in Table 2, LASB significantly outperforms all the other models in localization tasks for all the metrics. Specifically, LASB achieves a 0.6%↑, 17.4%↑, and 10.0%↑ improvement in AUROC$_{seg}$, AP$_{seg}$, F1max$_{seg}$ respectively. Also, it shows 2.0%↑ enhancement in F1max$_{cls}$ for detection, compared to GLAD [54]. These improvements highlight the robustness and effectiveness of the latent space approach specifically for localization. Furthermore, as shown in Figure 3, pixel-space models like DDPM [23], SGM [46], and SB models trained via Iterative Proportional Fitting [11, 25] or likelihood-based methods [8] re-

	NFE-1	NFE-2	NFE-5	NFE-10	NFE-100	NFE-500	NFE-1000
MVTec-AD	97.42 / 96.82	98.16 / 97.84	98.94 / 98.16	99.14 / 98.66	99.14 / 98.66	99.14 / 98.66	99.14 / 98.66
VisA	92.74 / 96.71	93.21 / 97.15	93.96 / 97.90	94.2 / 98.18	94.2 / 98.18	94.2 / 98.18	94.2 / 98.18
Inference Time (secs)	0.12	0.25	0.52	0.74	1.5	7.5	15

Table 4. Performance evaluation across varying numbers of function evaluations (NFEs) on the MVTec-AD and VisA datasets. The tabulated metrics, AUROC$_{cls}$ / AUROC$_{seg}$, provide a comprehensive overview of image-level and pixel-level anomaly detection performance at each NFE. Additionally, the inference time, measured in wall-clock seconds on an NVIDIA V100 GPU, underscores the computational trade-offs associated with increasing NFEs, reflecting the balance between model efficiency and performance.

Method	DRÆM [56]	PatchCore [38]	DiffAD [60]	TransFusion [16]	DIAD [21]	GLAD [54]	ADSPR [43]	DDAD [33]	LASB (ours)
Inference Time (secs)	0.15	0.44	2.6	1.2	2.8	1.8	1.6	1.5	0.74

Table 5. Inference times for various methods where the inference time, measured in wall-clock seconds on an NVIDIA V100 GPU.

quire more training time, memory, and exhibit slower sampling. In contrast, latent-space LASB excels in all areas, achieving up to $5\times$ and $2\times$ faster training, $3\times$ and $2\times$ memory reduction, and $4\times$ and $2\times$ faster sampling compared to SB-based models [8, 11, 25] and DDPM [23], respectively.

Task	**Metrics**	$1\times$	$2\times$	$4\times$	$8\times$
Detection	AUROC$_{cls}$	99.05	99.01	99.01	98.98
	AP$_{cls}$	99.15	99.14	99.14	99.11
	F1max$_{cls}$	98.59	98.33	99.83	98.32
Localization	AUROC$_{seg}$	98.58	98.58	98.59	98.57
	AP$_{seg}$	78.17	78.14	78.14	78.12
	F1max$_{seg}$	70.62	70.59	70.61	70.58

Table 6. Stability analysis of LASB on the MVTec-AD dataset for detection and localization tasks by performing the sampling multiple times. Results reported are mean values across classes and multiple samplings (see Section S1 of appendix for more detailed results).

Generalization and Sampling Stability. In the field of anomaly detection using diffusion models, achieving consistent outcomes during the sampling or inference stage is notably challenging due to the inherent stochastic nature of generative processes. This inconsistency can be particularly problematic in real-world applications where reliable and stable detection is critical. Therefore, assessing the stability of model outputs across multiple inferences is essential. Our approach involves training the model once and then conducting multiple sampling or inference tests to evaluate if the outcomes remain consistent over time. A critical consideration for real-world applicability is achieving fast inference while minimizing computational overhead. The proposed LASB model strikes a balance between performance and efficiency, ensuring its suitability for deployment in practical scenarios.

Our findings, as detailed in Table 6, reveal that the LASB model exhibits remarkable stability across all evaluation metrics, showing negligible variance across numerous inferences. This consistency is attributed to the model's capability to maintain structural integrity and effectively transition from anomalous to normal latent spaces. The LASB model is designed to reconstruct a normal image regardless of the underlying anomalies, compelling it to disregard anomalous features during reconstruction. This process not only ensures that anomalies of various patterns, sizes, and orientations are effectively handled but also enhances the overall reliability of the model.

Inference Complexity. As shown in Table 4, LASB demonstrates exceptional efficiency with its rapid sampling capabilities, achieving near-optimal performance with as few as 10 NFEs (Number of Function Evaluations) and an inference time of only 0.74 seconds on an NVIDIA V100 GPU. This highlights LASB's practicality for real-world applications, where fast and reliable anomaly detection and localization are crucial. To further illustrate its efficiency, we compare LASB's inference time with existing state-of-the-art (SOTA) models in Table 5. LASB is approximately $1.6\times$ faster than TransFusion [16], a leading class-based method, while also outperforming it in both anomaly localization and detection tasks. Compared to DiAD [21], LASB achieves $2\times$ faster inference times, delivering significantly stronger performance in both detection and localization metrics. This demonstrates LASB's clear advantage in balancing computational efficiency and robust anomaly detection.

6. Conclusions and Future Work

The proposed Latent Anomaly Schrödinger Bridge (LASB) model demonstrates robust performance in anomaly detection and localization tasks. Its unified nature delimitates the need for extra guidance or additional network components. LASB also excels in producing stable inferential results, requires less computational memory, and benefits from faster training and sampling rates. Given these advantages, LASB is well-suited for deployment in real-world industrial applications. Looking ahead, future research could focus on enhancing the robust translation mechanisms specific to anomaly detection and localization tasks.

Acknowlegements. We thank the anonymous reviewers for their valuable feedback that improved the presentation of this paper.

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Computer Vision–ACCV 2018*. Springer International Publishing, 2019. 2

[2] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. 3

[3] Alexander Bauer, Shinichi Nakajima, and Klaus-Robert Müller. Self-supervised autoencoders for visual anomaly detection. *Mathematics*, 12(24), 2024. 2

[4] Finn Behrendt et al. Patched diffusion models for unsupervised anomaly detection in brain mri. In *Medical Imaging with Deep Learning*, page PMLR, 2024. 1, 2

[5] 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*, pages 9592–9600, 2019. 5

[6] Paul Bergmann et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011*, 2018. 2

[7] Ziyi Chang, George A Koulieris, and Hubert PH Shum. On the design fundamentals of diffusion models: A survey. *arXiv preprint arXiv:2306.04542*, 2023. 5

[8] Tianrong Chen, Guan-Horng Liu, and Evangelos Theodorou. Likelihood training of schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations*, 2022. 2, 4, 7, 8

[9] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Stochastic control liaisons: Richard sinkhorn meets gaspard monge on a schrodinger bridge. *Siam Review*, 63(2):249–313, 2021. 4

[10] Anne-Sophie Collin and Christophe De Vleeschouwer. Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 7915–7922. IEEE, 2021. 2

[11] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021. 2, 4, 7, 8

[12] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021. 6

[13] 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*, pages 9737–9746, 2022. 6

[14] Wei Deng, Weijian Luo, Yixin Tan, Marin Bilos, Yu Chen, Yuriy Nevmyvaka, and Ricky T. Q. Chen. Variational schrödinger diffusion models. In *International Conference on Machine Learning (ICML)*, 2024. 2

[15] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 1

[16] Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj. Transfusion–a transparency-based diffusion model for anomaly detection. In *European conference on computer vision*, pages 91–108. Springer, 2024. 6, 8

[17] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1705–1714, 2019. 2

[18] Alvaro Gonzalez-Jimenez et al. Sano: Score-based diffusion model for anomaly localization in dermatology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1, 2

[19] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. 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*, pages 98–107, 2022. 6

[20] Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. MambaAD: Exploring state space models for multi-class unsupervised anomaly detection. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 6, 7

[21] 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*, pages 8472–8480, 2024. 1, 2, 3, 5, 6, 7, 8

[22] Liren He, Zhengkai Jiang, Jinlong Peng, Liang Liu, Qiangang Du, Xiaobin Hu, Wenbing Zhu, Mingmin Chi, Yabiao Wang, and Chengjie Wang. Learning unified reference representation for unsupervised multi-class anomaly detection. *arXiv preprint arXiv:2403.11561*, 2024. 3

[23] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 6, 7, 8

[24] Bozhen Hu et al. A lightweight spatial and temporal multi-feature fusion network for defect detection. *IEEE Transactions on Image Processing*, 30:472–486, 2020. 1

[25] Solomon Kullback. Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4): 1236–1243, 1968. 4, 7, 8

[26] Yufei Liang et al. Omni-frequency channel-selection representations for unsupervised anomaly detection. *IEEE Transactions on Image Processing*, 2023. 2, 6

[27] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A. Theodorou, Weili Nie, and Anima Anandkumar. I2sb: Image-to-image schrödinger bridge. In *International Conference on Machine Learning*, 2023. 4, 5

[28] Wenqian Liu, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J Radke, and Octavia Camps. Towards visually explaining variational autoencoders. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8642–8651, 2020. 2

[29] 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*, pages 20402–20411, 2023. 6

[30] Xingming Long et al. Fabric defect detection using tactile information. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021. 1

[31] Fanbin Lu et al. Removing anomalies as noises for industrial defect localization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. 1, 2

[32] 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, 2023. 3, 6, 7

[33] Arian Mousakhan, Thomas Brox, and Jawad Tayyub. Anomaly detection with conditioned denoising diffusion models. *arXiv preprint arXiv:2305.15956*, 2023. 6, 7, 8

[34] Edward Nelson. *Dynamical theories of Brownian motion*. Princeton university press, 2020. 4

[35] Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *International Conference on Image Analysis and Processing*, pages 394–406. Springer, 2022. 2

[36] Hannes Risken and Hannes Risken. *Fokker-planck equation*. Springer, 1996. 4

[37] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 5, 6, 7

[38] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2022. 6, 8

[39] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Fully convolutional cross-scale-flows for image-based defect detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1088–1097, 2022. 6

[40] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022. 5

[41] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019. 2

[42] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[43] 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*, pages 23372–23382, 2023. 1, 2, 6, 8

[44] Kihyuk Sohn et al. Anomaly clustering: Grouping images into coherent clusters of anomaly types. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023. 1

[45] Jouwon Song et al. Anoseg: anomaly segmentation network using self-supervised learning. *arXiv preprint arXiv:2110.03396*, 2021. 2

[46] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. 2, 3, 7

[47] Daniel Stanley Tan, Yi-Chun Chen, Trista Pei-Chun Chen, and Wei-Chao Chen. Trustmae: A noise-resilient defect classification framework using memory-augmented auto-encoders with trust regions. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 276–285, 2021. 2

[48] Justin Tebbe and Jawad Tayyub. Dynamic addition of noise in a diffusion model for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3940–3949, 2024. 6

[49] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. 5

[50] Shashanka Venkataramanan et al. Attention guided anomaly localization in images. In *European Conference on Computer Vision*. Springer International Publishing, 2020. 1

[51] Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via schrödinger bridge. In *International conference on machine learning*, pages 10794–10804. PMLR, 2021. 2

[52] Julian Wyatt et al. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 1, 2

[53] Rui Xu, Yunke Wang, and Bo Du. Maediff: Masked autoencoder-enhanced diffusion models for unsupervised anomaly detection in brain images. *arXiv preprint arXiv:2401.10561*, 2024. 1

[54] Hang Yao, Ming Liu, Zhicun Yin, Zifei Yan, Xiaopeng Hong, and Wangmeng Zuo. Glad: towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection. In *European Conference on Computer Vision*, pages 1–17. Springer, 2024. 6, 7, 8

[55] 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. 3, 6, 7

[56] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draema discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8330–8339, 2021. 2, 4, 6, 7, 8

[57] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021. 2

[58] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr–a dual subspace re-projection network for surface anomaly detection. In *European conference on computer vision*. Springer Nature Switzerland, 2022. 2, 6

[59] Zhaoyang Zeng et al. Reference-based defect detection network. *IEEE Transactions on Image Processing*, 30:6637–6647, 2021. 1

[60] Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, and Shu-Tao Xia. Unsupervised surface anomaly detection with diffusion probabilistic model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6782–6791, 2023. 1, 2, 6, 7, 8

[61] Ying Zhao. Omnial: A unified cnn framework for unsupervised anomaly localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3924–3933, 2023. 6

[62] Zhixuan Zhao et al. A surface defect detection method based on positive samples. In *PRICAI 2018: Trends in Artificial Intelligence*. Springer International Publishing, 2018. 2

[63] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pretraining for anomaly detection and segmentation. In *European Conference on Computer Vision*. Springer, Springer Nature Switzerland, 2022. 5