

tion. The models produce blurred results with relatively poor anomaly localization performance compared with the state-of-the-art.

In this work, we propose a new reconstruction-based approach for anomaly detection, achieving precise anomaly localization and top reconstruction quality; see Fig. 1. Our key idea is to formulate the anomaly detection task as a noise or anomaly removal problem. First, we introduce random noises into the input image and train an autoencoder as a denoising model. The anomalous pixels are considered as noises and will not be excluded from the reconstruction. The previous reconstruction-based methods directly reconstruct the input with noisy images, leading to large reconstruction errors and suboptimal anomaly detection performance. We leverage a diffusion model [15] for denoising and reconstruction. We inspect the intermediate stages of the diffusion model and measure the reconstruction error of each step for accurate anomaly localization. Moreover, we require the model to reconstruct the input features and detect anomalies in both pixel and feature space.

We further propose a gradient denoising process for reconstructing normal images from anomalous ones and provide an interpretable explanation of the anomaly detection results. Our process smoothly transforms an anomalous image into a normal image while preserving the structural appearance and high-frequency details of the normal regions. It is achieved by consistently denoising the gradients from a pre-trained deep feature extractor. Our approach is shown to outperform existing methods in terms of reconstruction quality and anomaly detection accuracy.

## 2. Related Works

**Anomaly Detection** Various methods have been developed to tackle anomaly detection and localization. Support vector data description (SVDD) [31, 26] is proposed for anomaly detection. Teacher-Student [4] proposes to distill the knowledge from a pre-trained teacher network to a student network on the anomaly-free data. The difference in the outputs of teacher and student networks is used as an anomaly score. DRAEM [40] proposes to add artificial defects to the normal images to generate pseudo anomaly samples and labels to train a segmentation network for anomaly segmentation. CutPaste [19] proposed a self-training strategy with a generative one-class classifier.

Reconstruction-based approaches [5, 7] are a widely used branch of anomaly detection. They assume that only the normal image can be well reconstructed. Anomalies can be detected by measuring the difference between original and reconstructed images. Autoencoders [5, 7], variational autoencoders (VAE) [34], and Adversarial generative networks (GAN) [1] are often used to reconstruct an anomalous image to a normal one. However, a limitation of these methods is that anomalies can sometimes be reconstructed,

leading to degraded anomaly detection performance.

Embedding-based methods [6, 8, 25] employ neural networks to extract meaningful features for anomaly detection and localization. Spade [6] first introduced a method for detecting anomalies using ImageNet pre-trained deep networks. This method uses K-NN search to match anomaly features with the K nearest normal features. PaDiM [8] build a multivariate Gaussian distribution and use Mahalanobis distance as the anomaly score. PatchCore [25] proposes a memory bank to save the coreset of the normal features, which improves the time and memory complexity. Recently, UniAD [36] proposed a transformer network for reconstructing features with masked self-attention to avoid the model collapsing into an identity function. This allows a single model to detect anomalies in all categories.

Flow-based methods [11, 16, 13, 38] recently boosted the performance of anomaly detection. Normalized flow models are generative models that learn to map two distributions and estimate the probability density reversibly. CFLOW-AD [13] proposes to use conditional normalized flow with positional embedding on the multi-scale features for anomaly detection. FastFlow [38] proposes to employ a 2D flow model that combines local and global features to estimate the probability density. These methods demonstrate the efficacy of generative models in addressing anomaly detection, which has inspired our work with the diffusion model.

**Diffusion Models** Diffusion models [27, 15] are a powerful generative model that achieves state-of-the-art performance in image generation tasks. Recent methods [15, 23, 10] are proposed to generate a realistic image by gradually denoising random Gaussian noises. The likelihood training makes the diffusion model capable of learning data density. DDIM [28] speeds up the diffusion sampling with a non-Markov reverse sampling. The score-based model [29] is another denoising generative model with a similar diffusion process.

AnoDDPM [35] has introduced diffusion models for medical image anomaly segmentation but only used the diffusion model as a high-quality reconstruction model. Relying on reconstruction error in RGB space for anomaly score leads to noisy predictions and limited performance in many industrial anomaly detection applications.

## 3. Methods

### 3.1. Preliminary

Diffusion models are powerful generative models that can approximate data distribution and create realistic images. Given a data distribution  $p(x)$ , denoising diffusion probabilistic models (DDPM) [15] learn the distribution with a Markov Chain denoising process. During training, it gradually adds random Gaussian noises to a real image  $x_0$ ,