モデルは、最先端技術と比べて異常検出の局所化性能が比較的劣るぼやけた結果を生成します。

本研究では、異常検出のための新しい再構築ベースのアプローチを提案し、正確な異常局所化と最高品質の再構築を実現しています；図1を参照。私たちの主要なアイデアは、異常検出タスクをノイズまたは異常除去問題として定式化することです。まず、入力画像にランダムノイズを導入し、オートエンコーダーをノイズ除去モデルとして訓練します。異常ピクセルはノイズとして扱われ、再構築から除外されません。従来の再構築ベースの手法は、ノイズを含む画像から直接再構築を行うため、大きな再構築誤差と最適な異常検出性能が得られません。私たちは、ノイズ除去と再構築に拡散モデル[15]を活用します。拡散モデルの中間段階を検査し、各ステップの再構築誤差を測定することで、正確な異常局所化を実現します。さらに、モデルが入力特徴を再構築し、ピクセル空間と特徴空間の両方で異常を検出するように要求します。

私たちはさらに、異常画像から正常画像を再構築するための勾配ノイズ除去プロセスを提案し、異常検出結果の解釈可能な説明を提供します。このプロセスは、正常領域の構造的外観と高周波詳細を保持しながら、異常画像を正常画像に滑らかに変換します。これは、事前訓練された深層特徴抽出器から勾配を継続的にノイズ除去することで実現されます。当アプローチは、再構築品質と異常検出精度において既存手法を凌駕することが示されています。

これにより、異常検出の性能が低下します。

異常検出異常検出と局所化に対応するため、さまざまな手法が開発されています。サポートベクターデータ記述（SVDD）[31, 26] は異常検出のために提案されました。ティーチャー・スチューデント [4] は、異常のないデータ上で事前訓練されたティーチャーネットワークから知識をスチューデントネットワークに蒸留する手法を提案しています。教師ネットワークと学生ネットワークの出力の差分が異常スコアとして使用されます。DRAEM [40]は、正常画像に人工的な欠陥を追加して擬似異常サンプルとラベルを生成し、異常セグメンテーション用のセグメンテーションネットワークを訓練する手法を提案しています。CutPaste [19]は、生成型単一クラス分類器を用いた自己訓練戦略を提案しています。
再構築ベースのアプローチ [5, 7]は、異常検出の広く用いられる分野です。これらは、正常な画像のみが適切に再構築可能であると仮定しています。異常は、オリジナル画像と再構築画像の差分を測定することで検出されます。オートエンコーダー[5, 7]、変分オートエンコーダー（VAE）[34]、敵対的生成ネットワーク（GAN）[1]は、異常画像を正常画像に再構築するために頻繁に使用されます。しかし、これらの方法の制限は、異常が時々再構築可能である点です。

フローベースの手法 [11, 16, 13, 38] は最近、異常検出の性能を向上させました。正規化フローモデルは、2つの分布をマッピングし、確率密度を逆方向に推定する生成モデルです。CFLOW-AD [13] は、異常検出のためにマルチスケール特徴量にポジション埋め込みを組み合わせた条件付き正規化フローを使用することを提案しています。FastFlow [38] は、局所的およびグローバルな特徴を組み合わせた2次元フローモデルを用いて確率密度を推定する手法を提案しています。これらの手法は、異常検出における生成モデルの有効性を示しており、これが当研究の拡散モデルを用いたアプローチのきっかけとなりました。

埋め込みベースの手法 [6, 8, 25] は、異常検出と局所化のための意味のある特徴を抽出するためにニューラルネットワークを採用しています。Spade [6] は、ImageNetで事前訓練された深層ネットワークを用いた異常検出手法を初めて提案しました。この手法は、異常特徴をK番目に近い正常特徴と一致させるためにK-NN検索を使用します。PaDiM [8] は多変量ガウス分布を構築し、マハラノビス距離を異常スコアとして使用します。PatchCore [25] は、正常な特徴のコアセットを保存するためのメモリバンクを提案し、時間とメモリの複雑さを改善します。最近、UniAD [36] は、モデルが恒等関数に収束するのを防ぐためにマスク付き自己注意を用いた特徴の再構築にトランスフォーマーネットワークを提案しました。これにより、単一のモデルで全カテゴリの異常を検出可能です。

データ分布 p(x) に対し、ノイズ除去拡散確率モデル（DDPM）[15] はマルコフ連鎖ノイズ除去プロセスを用いて分布を学習します。訓練中、モデルは現実の画像 x_0 にランダムなガウスノイズを徐々に追加していきます。

拡散モデル拡散モデル[27, 15]は、画像生成タスクで最先端の性能を達成する強力な生成モデルです。最近の方法[15, 23, 10]は、ランダムなガウスノイズを段階的に除去して現実的な画像を生成するように提案されています。尤度訓練により、拡散モデルはデータ密度を学習する能力を獲得します。DDIM[28]は、非マルコフ逆サンプリングにより拡散サンプリングを高速化します。スコアベースモデル [29] は、同様の拡散プロセスを有する別のノイズ除去生成モデルです。
AnoDDPM [35]は医療画像の異常セグメンテーションに拡散モデルを導入しましたが、拡散モデルを高品質な再構築モデルとしてのみ使用しました。RGB空間での再構築誤差を異常スコアに依存すると、ノイズの多い予測と多くの産業用異常検出アプリケーションでの性能制限が生じます。

3. Methods

3.1. Preliminary

拡散モデルは、データ分布を近似し現実的な画像を生成できる強力な生成モデルです。

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 $\boldsymbol{x}_0$,