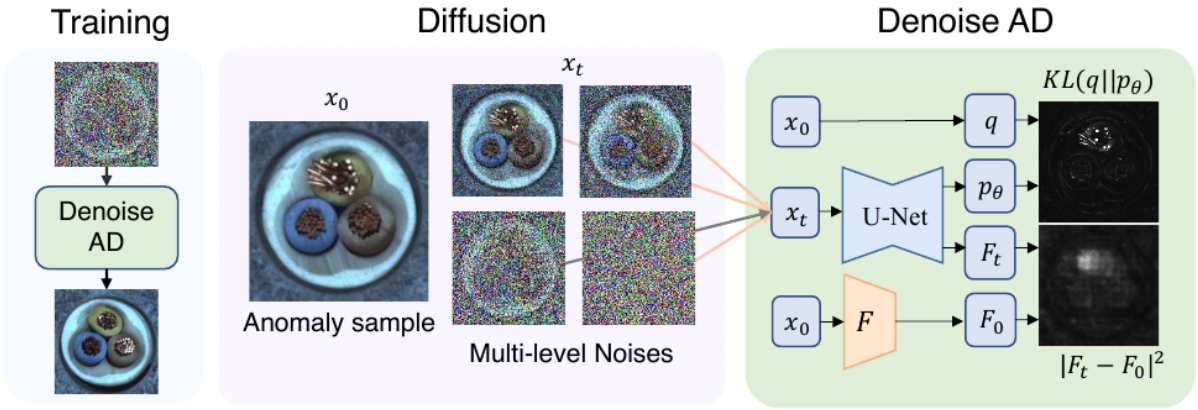

図2：当社のノイズ除去拡散モデルは、異常のない画像のみで訓練されます。推論時、異常サンプルに異なるスケールのノイズが追加されます。十分な大きさのノイズを加えると、異常ピクセルは正常ピクセルと区別できなくなり、再構築が容易になります。後方分布 q(x_{t-1} | x_t, x_0) と推定分布 p_{\theta}(x_{t-1} | x_t) の間のKLダイバージェンスをピクセルレベルの異常スコアとして採用します。特徴再構築のMSE誤差を特徴レベルのスコアとして使用します。異なるノイズスケールからの結果の平均を出力として採用します。

これにより、一連のノイズ付き画像（x_1 ，x_2 ，...，x_T ）が生成されます。導入されるガウスノイズの分散は $\{\beta_t\}_{t=1,2,\cdots,T}$ で表されます。データ分布と追加されたノイズがともにガウス分布であるため、ノイズ付き画像x_tの閉形式は：

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\boldsymbol{I}), \tag{1}$$

ここで、$\alpha^-_t = \prod_{s=1}^{Qt}$（1 − β_s ）です。拡散モデルは、p_θ（x_0 ）$=^R$ $p(x_{0:T})$ $dx_{1:T}$ で表されます。画像生成プロセス中、モデルはまず一様ガウス分布 p_T（x）¤ N（x_T ；0，I ）からサンプリングし、推定された分布 p_θ（x_t ）からサンプリングすることで画像を徐々にノイズ除去します：

$$p_\theta(\boldsymbol{x}_{0:T}) = p(\boldsymbol{x}_T)\prod_{t=1}^{T} p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t), \tag{2}$$

$$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t), \boldsymbol{\Sigma}_\theta(\boldsymbol{x}_t, t)). \tag{3}$$

拡散モデルの訓練はオートエンコーダーとして扱えます。DDPM [15] で提案されたように、拡散モデルはMSE損失を用いてノイズのスケール　を予測するように訓練されます。

$$L_{mse} = \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{\epsilon}}\left[(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t))^2\right] \tag{4}$$

変分境界に基づく追加のトレーニング損失を用い、拡散モデル自体がノイズの分散を自動的に学習するように設計されています。[23]で提案された方法です：

$$L_{vlb} = L_0 + L_1 + \cdots + L_{T-1} + L_T, \tag{5}$$

$$L_0 = -logp_\theta(\boldsymbol{x}_0|\boldsymbol{x}_1), \tag{6}$$

$$L_{t-1} = D_{KL}(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)||p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)), \tag{7}$$

$$L_T = D_{KL}(q(\boldsymbol{x}_T|\boldsymbol{x}_0)||p(\boldsymbol{x}_T)). \tag{8}$$

3.2. 異常検出用のノイズ除去モデル

自動エンコーダーに基づく従来の再構築手法[2，5]は、自動エンコーダーが訓練中に同一のマッピングに退化するため、異常の再構築に成功する問題を抱えています。しかし、ノイズを含む画像での再構築はこの問題を回避します。図2に示すように、異常画像に徐々にノイズを加えると、ノイズレベルが大きな場合、異常領域が消失し、正常サンプルのピクセルと区別できなくなります。ただし、ノイズを含む画像からノイズのない画像への直接的な再構築は、重大な再構築誤差を引き起こす可能性があります。本研究では、生成型拡散モデルDDPM [15] を利用して、画像を徐々にノイズ除去し再構築します。拡散モデルは、DDPMのトレーニング手順を用いて異常のないデータでトレーニングされます。

ピクセルレベルスコア。異常検出のため、画像x_0にランダムなガウスノイズを加えてx_tを取得します。従来の再構築ベースの手法では、再構築画像と元の入力のRGB空間での差分を異常スコアとして使用します。しかし、このアプローチは$p(x_0 | x_t)$の推定が困難であり、結果に大きなノイズを導入します。この制限を克服するため、事後分布$q(x_{t-1} | x_t , x_0)$と推定分布$p\theta(x_{t-1} | x_t)$のKLダイバージェンスを異常スコアとして採用します。

$$\boldsymbol{s}_t = KL(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)||p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)). \tag{9}$$

図6では、KLダイバージェンスがノイズの少ない入力ピクセルの確率を適切に測定することを示しています。

特徴量レベルスコア。拡散モデルの結果は境界部で鋭いものの、頑健性がないことが観察されます

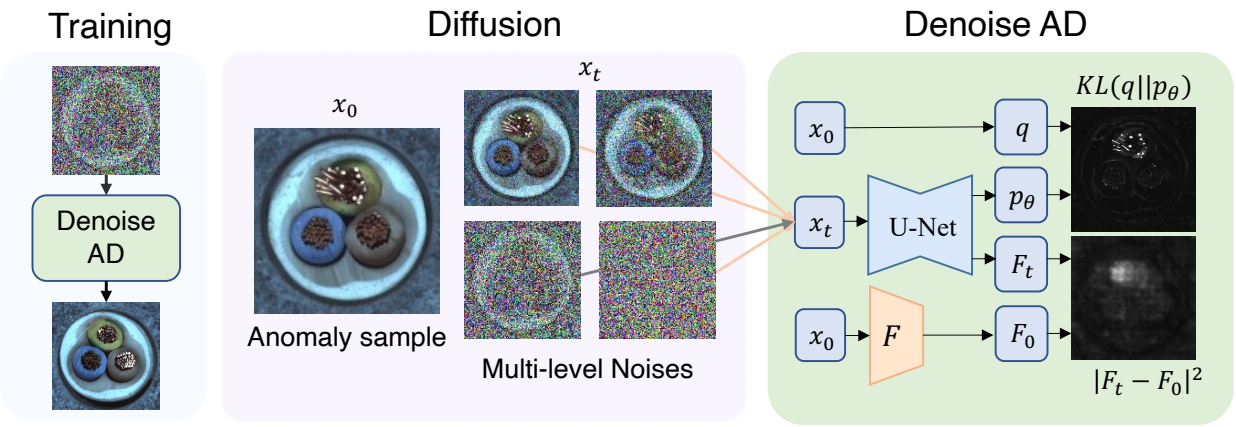

Figure 2: Our denoising diffusion model is trained with only anomaly-free images. During inference, noises of different scales are added to the anomaly sample. With large enough noises, the anomalous pixels become indistinguishable from the normal pixels and easier for reconstruction. We take the KL-divergence between the posterior distribution $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$ and estimated distribution $p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ as the pixel-level anomaly score. The MSE error of feature reconstruction is used as a feature-level score. We take the average of results from different noise scales as the outputs.

which leads to a series of noised images $(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T)$. The variances of Gaussian noises introduced are denoted as $\{\beta_t\}_{t=1,2,\cdots,T}$. Since the data distribution and noises added are both Gaussian, the closed form of a noised image $\boldsymbol{x}_t$ is:

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\boldsymbol{I}), \tag{1}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$. The diffusion models are then represented with $p_\theta(\boldsymbol{x}_0) = \int p(\boldsymbol{x}_{0:T})\,\mathrm{d}\boldsymbol{x}_{1:T}$. During the image generation process, the model first samples from uniform Gaussian distribution $p_T(\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{x}_T; \boldsymbol{0}, \boldsymbol{I})$ and gradually denoises the image by sampling from the estimated distribution $p_\theta(\boldsymbol{x}_t)$:

$$p_\theta(\boldsymbol{x}_{0:T}) = p(\boldsymbol{x}_T)\prod_{t=1}^{T} p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t), \tag{2}$$

$$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t), \boldsymbol{\Sigma}_\theta(\boldsymbol{x}_t, t)). \tag{3}$$

The training of diffusion models can be treated as an autoencoder. As proposed in DDPM [15], the diffusion models are trained with an MSE loss to predict the scale of noises ϵ.

$$L_{mse} = \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{\epsilon}}\left[(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t))^2\right] \tag{4}$$

An additional training loss based on the variational bound is used to automatically learn the variance of noises by the diffusion model itself, as proposed by [23]:

$$L_{vlb} = L_0 + L_1 + \cdots + L_{T-1} + L_T, \tag{5}$$

$$L_0 = -logp_\theta(\boldsymbol{x}_0|\boldsymbol{x}_1), \tag{6}$$

$$L_{t-1} = D_{KL}(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)||p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)), \tag{7}$$

$$L_T = D_{KL}(q(\boldsymbol{x}_T|\boldsymbol{x}_0)||p(\boldsymbol{x}_T)). \tag{8}$$

3.2. Denoising Model for Anomaly detection

Previous reconstruction methods based on AutoEncoder [2, 5] suffer from the successful reconstruction of anomalies because the AutoEncoder easily degrades to an identical mapping during training. However, reconstruction with noisy images prevents the issue. As illustrated in Fig 2, gradually adding noise to an anomalous image causes the anomalous regions to vanish for large noise levels, making them indistinguishable from the pixels of normal samples. Nevertheless, direct reconstruction from noisy to noise-free images can result in significant reconstruction errors. In this study, we utilize a generative diffusion model DDPM [15] to gradually denoise and reconstruct the image. The diffusion model is trained on anomaly-free data using the training procedure of DDPM.

Pixel-level score. For anomaly detection, we begin by corrupting an image $\boldsymbol{x}_0$ with random Gaussian noises to obtain $\boldsymbol{x}_t$. Previous reconstruction-based methods employ the difference between the reconstructed image and the original input in RGB space as the anomaly score. However, this approach entails a difficult estimation of $p(\boldsymbol{x}_0|\boldsymbol{x}_t)$ and introduces significant noise to the results. To address this limitation, we employ the KL divergence of the posterior distribution $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$ and the estimated distribution $p\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ as the anomaly score,

$$\boldsymbol{s}_t = KL(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)||p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)). \tag{9}$$

We show in Fig. 6 that the KL divergence correctly measures the likelihood of input pixels with much less noise.

Feature-level score. We observe that the results from the diffusion model are usually sharp in boundary but not robust