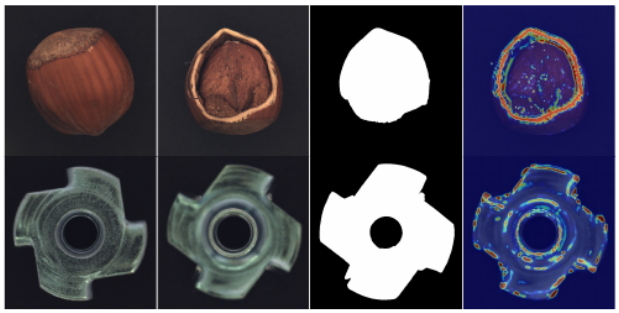


(a) Normal (b) Anomalous (c) Label (d) Prediction

図3：ラベルの曖昧さ。拡散モデルは、成功した再構築のために変更が必要な異常ピクセルに焦点を当てます。大規模なヘーゼルナッツの割れ目と金属ナットの反転を正しくセグメント化するには、セマンティック情報が必要です。

異常検出の再現率において。図3に示すように、異常領域は正常ピクセルと類似した色を持つため、ノイズ除去モデルにより高い確率で検出されます。拡散モデルは、再構築に成功するために変更が必要な異常ピクセルを優先的に処理しますが、この問題に対処するにはセマンティック情報が必要です。

異常検出の精度を向上させるため、ピクセル空間と特徴空間の両方を考慮した共同分布アプローチを提案します。このアプローチは、P(x, f) で表されます。入力画像の深層特徴を抽出するために、事前訓練された特徴抽出器を採用しています。拡散モデルは、ノイズのない画像と破損した画像のピクセルとセマンティック特徴を同時に再構築するように訓練されます。トレーニング損失および異常スコアとして、平均二乗誤差（MSE）損失関数を採用します。これは次のように定義されます：

$$s_t^f = L_{mse}^f = \frac{1}{C \times H \times W} \sum |f(\boldsymbol{x}_0) - f(\boldsymbol{x}_t)|^2, \quad (10)$$

ここで、f は形状$R^{C\times H\times W}$ の特徴を抽出する事前訓練された特徴抽出器です。x_0 と x_t はそれぞれノイズのない画像と対応するランダムノイズを含む破損画像を表します。最終的な異常スコアは、ピクセルレベルと特徴レベルの結果の加重和です。

多階層ノイズ。異なる異常は、異なるノイズ階層に対して異なる感度を示すことが観察されています。一部の異常は容易に検出可能ですが、他の異常は異常なピクセルを覆い隠すのに十分な大きなノイズが必要です。私たちは、さまざまなノイズ階層に対する異常スコアを測定し、結果を平均化します。KLダイバージェンススコアは時間ステップtに応じて大きく変動するため、平均化する前に正規化します。最終的な異常スコアは次のように算出されます：

$$A = \sum_{i=1,2,\ldots,n} \alpha \hat{s}_{t_i} + (1-\alpha) s_{t_i}^f, \quad (11)$$

アルゴリズム1：勾配ノイズ除去再構築。

入力：画像 x_0 ，ガウス分布（μ，Σ）

Output: $\boldsymbol{x}_N$

for $t = 1, \cdots, N$ **do**

$\boldsymbol{f}_t = F(\boldsymbol{x}_t)$

$\boldsymbol{g} = \nabla_{\boldsymbol{x}_t}(\boldsymbol{f}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{f}_t - \boldsymbol{\mu})$

$\boldsymbol{x}_t = \sqrt{1-\hat{\beta}_t}\boldsymbol{x}_{t-1} + \sqrt{\hat{\beta}_t}\boldsymbol{g}$;

if $t \% N_d = 0$ **then**

$\boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}_t), \boldsymbol{\sigma}_\theta(\boldsymbol{x}_t))$

end

end

ここで、$\hat{s}_{ti}$ は平均と標準偏差で正規化されたスコアであり、$T = \{t_1, t_2, \cdots, t_n\}$ は拡散モデルのフォワードプロセスで選択されたタイムステップです。私たちは第4.4節のアンサンブル要因αの効果。統一モデル。拡散モデルの拡散ネットワークの容量は、任意の複雑な分布をモデル化するのに十分であることが証明されています[15, 10]。UniAD[36]と同様に、単一の拡散モデルで複数のカテゴリの分布を学習する実験を実施しました。表3は、単一の統合モデル設定下で、当社の統合モデルの性能が他の手法を大幅に上回っていることを示しています。結果は、拡散モデルを異常局所化に活用する有効性を確認しています。

3.3. 再構築のための勾配ノイズ除去

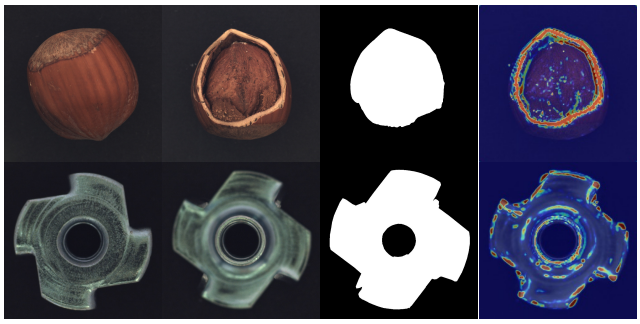
異常領域は、拡散モデルで除去可能な特殊なノイズとして解釈できます。私たちは、DDPM [15] の逆拡散プロセスに簡単な調整を加えた勾配ノイズ除去プロセスを提案します。異常画像は滑らかに正常画像に変換され、異常検出結果の解釈可能な説明を提供します。

まず、異常再構築のための勾配降下最適化プロセスを導入します。異常のないデータの深層特徴量に対して、PaDiM [8] で近似された多変量ガウス分布 N（μ, Σ）を用います。再構築のため、PaDiM の特徴量抽出器で埋め込み f（x_0 ）を抽出し、マハラノビス距離を用いて勾配降下で画像を最適化します：

$$L = (f(\boldsymbol{x}_0) - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(f(\boldsymbol{x}_0) - \boldsymbol{\mu}), \quad (12)$$

$$\boldsymbol{x}_{t+1} = \omega \boldsymbol{x}_t - s\nabla_{\boldsymbol{x}_t} L, \quad (13)$$

ここでωは重み減衰因子、sは学習率です。このプロセスは、PaDiMの異常スコアを最小化するよう画像を最適化します。しかし、ノイズの多い勾配∇_{xt} Lは、いくつかの反復後に画像を破損し、画像に大きなノイズを導入します。私たちは、高品質な再構築のため、拡散モデルを用いて勾配のノイズ除去を提案します。



(a) Normal (b) Anomalous (c) Label (d) Prediction

Figure 3: Ambiguity of label. The diffusion model focuses on the anomalous pixels that need to be altered for successful reconstruction. It requires semantic information to correctly segment large-area hazelnut cracking and metal nut flipping.

in anomaly recall. As demonstrated in Fig. 3, the anomaly regions with a similar color to the normal pixels are assigned with a high likelihood by the denoising model. The diffusion model prioritizes the anomalous pixels that need to be altered for successful reconstruction, which requires semantic information to address the issue.

To enhance the accuracy of anomaly detection, we propose a joint distribution approach that considers both the pixel space and feature space, represented by $P(\boldsymbol{x}, \boldsymbol{f})$. We employ a pre-trained feature extractor to extract the deep features of the input image. The diffusion model is trained to concurrently reconstruct the pixels and semantic features of the noise-free image with the corrupted image. We adopt the Mean Squared Error (MSE) loss function as the training loss and anomaly score, which is defined as follows:

$$s_t^f = L_{mse}^f = \frac{1}{C \times H \times W} \sum |f(\boldsymbol{x}_0) - f(\boldsymbol{x}_t)|^2, \quad (10)$$

where f is a pre-trained feature extractor to extract features with shape $\mathbb{R}^{C \times H \times W}$, $\boldsymbol{x}_0$ and $\boldsymbol{x}_t$ represent a noise-free image and the corresponding corrupted image with random noises, respectively. The final anomaly score is the weighted sum of the pixel-level and feature-level results.

Multi-scale noises. We have observed that different anomalies exhibit varying sensitivities to different noise scales. While some anomalies can be detected easily, others require sufficiently large noise to overwhelm the anomalous pixels. We measure the anomaly score for various noise scales and average the results. Since the KL-divergence score varies significantly with the timestep t, we normalize it before averaging. The final anomaly score is obtained as follows:

$$A = \sum_{i=1,2,\ldots,n} \alpha \hat{s}_{t_i} + (1-\alpha) s_{t_i}^f, \quad (11)$$

Algorithm 1: Gradient Denoising Reconstruction.

Input: Image $\boldsymbol{x}_0$, Gaussian $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
Output: $\boldsymbol{x}_N$
for $t = 1, \cdots, N$ **do**
  $\boldsymbol{f}_t = F(\boldsymbol{x}_t)$
  $\boldsymbol{g} = \nabla_{\boldsymbol{x}_t} (\boldsymbol{f}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{f}_t - \boldsymbol{\mu})$
  $\boldsymbol{x}_t = \sqrt{1-\hat{\beta}_t} \boldsymbol{x}_{t-1} + \sqrt{\hat{\beta}_t} \boldsymbol{g}$;
  if $t \% N_d = 0$ **then**
    $\boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}_t), \boldsymbol{\sigma}_{\theta}(\boldsymbol{x}_t))$
  end
end

where $\hat{s}_{t_i}$ is the normalized score by mean and standard deviation, $T = \{t_1, t_2, \cdots, t_n\}$ are the selected timesteps of the forward-process of the diffusion model. We analyze the effects of ensembling factor α in Sec. 4.4.

Unified model. It has been proved that the diffusion model's capacity of the diffusion network is large enough for modeling any complex distributions [15, 10]. Like UniAD [36], we conduct experiments to learn distributions of multiple categories with a single diffusion model. Table 3 illustrates that the performance of our unified model outperforms the other methods by a large margin under the single unified model setting. The results confirm the effectiveness of utilizing the diffusion model for anomaly localization.

3.3. Gradient Denosing for Reconstruction

An image's anomalous regions can be viewed as a special type of noise that can be removed using the diffusion model. We propose a gradient denoising process to remove the anomalies with simple adjustments to the reverse diffusion process of DDPM [15]. An anomalous image can be smoothly transformed into a normal one, providing an interpretable explanation of the anomaly detection results.

We first introduce a gradient descending optimization process for anomaly reconstruction. We take the multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ approximated by PaDiM [8] on the deep features of anomaly-free data. For reconstruction, we extract embedding $f(\boldsymbol{x}_0)$ with the feature extractor of PaDiM and use the Mahalanobis distance to optimize the image with gradient descending:

$$L = (f(\boldsymbol{x}_0) - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (f(\boldsymbol{x}_0) - \boldsymbol{\mu}), \quad (12)$$

$$\boldsymbol{x}_{t+1} = \omega \boldsymbol{x}_t - s \nabla_{\boldsymbol{x}_t} L, \quad (13)$$

where ω is weight decay factor and s is the learning rate. The process optimizes the image such that the anomaly score of PaDiM is minimized. However, the noisy gradients $\nabla_{\boldsymbol{x}_t} L$ will corrupt the image after some iterations, introducing significant noises to the image. We propose to leverage the diffusion model to denoise the gradients for high-quality reconstruction.