

図 6: MVTec-AD [3] における異常の局在化の視覚化。左から右の列は、異常画像、対応するグラウンドトゥース、KL 発散に基づくピクセルレベルの予測、および元の画像上の最終的な異常マップの視覚化を示しています。当社のノイズ除去拡散モデルは、境界を意識したピクセルレベルの異常スコアを生成し、最終的な異常スコアの品質を向上させます。

特徴抽出器として、特徴の次元が比較的小さい EfficientNet [30] を採用しています。ストライド (2、4、8、16)、次元 (24、32、56、160) の中間特徴マップを選択し、 $64 \times 64$  にサイズ変更して、それらを連結してテンソル  $f \in \mathbb{R}^{72 \times 64 \times 64}$  を作成します。連結した特徴の解像度を検証するための実験を実施しました。表 2 を参照してください。2 を参照してください。解像度を  $16 \times 16$  から  $64 \times 64$  に上げると、PRO は 2.5% 向上します。ストライド (2、4、8、16) の EfficientNet 特徴を組み合わせると、PRO は 3.5% 向上します。

統一モデル。当社のノイズ除去拡散モデルは複雑な現実の産業データ分布をモデル化できるため、MVTEC-ADのすべてのカテゴリで異常検出に単一のモデルを使用する実験を実施しました。詳細は表を参照してください。3. DRAEM [40]とPaDiM [8]の性能は統一設定で大幅に低下し、PROメトリクスの性能が5%以上低下しています。一方、当社のモデルはPROで93.0%を達成し、性能低下は1.1%未満です。

#### 4.5. 再構築の定性的結果。

図7では、事前訓練された特徴抽出器の指導の下で、当社のノイズ除去勾配プロセスが異常画像を正常画像に滑らかに変換できることを示しています。左の2列は ground truth と入力異常画像です

Resize	Stride				AUROC	PRO
	16	8	4	2		
64	✓				94.2	90.2
	✓	✓			95.8	91.9
	✓	✓	✓		96.4	93.2
	✓	✓	✓	✓	<b>96.7</b>	<b>93.7</b>
32	✓	✓	✓	✓	96.6	92.4
16	✓	✓	✓	✓	95.5	91.2

表 2: 特徴レベルの再構築の消去研究。実験は、EfficientNet [30] のさまざまな層の特徴を、さまざまな解像度にサイズ変更して実施しています。

Method	Base		Unified	
	AUROC	PRO	AUROC	PRO
PaDiM[8]	<b>97.8</b>	92.8	90.5	85.3
DRAEM[40]	97.3	93.0	89.4	82.2
UniAD[36]	96.6	-	<b>97.0</b>	91.1
Ours	96.7	<b>94.1</b>	96.0	<b>93.0</b>

表3: MVTEC-ADのすべてのカテゴリにおける異常検出用の単一統合モデルの比較。

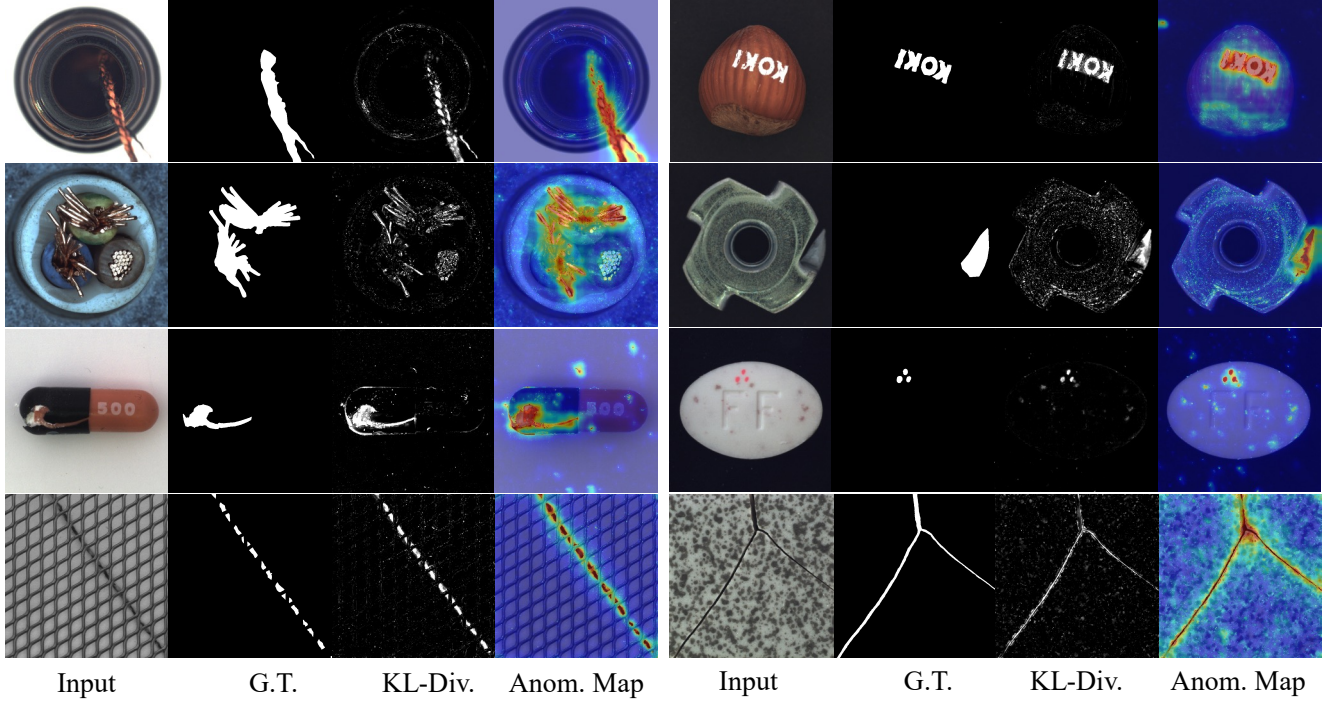


Figure 6: Visualization of anomaly localization on MVTec-AD [3]. The columns from left to right present anomalous images, the corresponding ground truths, our pixel-level predictions based on KL divergence, and the visualization of the final anomaly map on the original image. Our denoising diffusion models produce boundary-aware pixel-level anomaly scores to improve the quality of final anomaly scores.

adopt EfficientNet [30] as our feature extractor, where the dimension of features is relatively small. We select the intermediate feature map of stride (2, 4, 8, 16) with dimension (24, 32, 56, 160), resize to  $64 \times 64$  and concatenate them together into a tensor  $\mathbf{f} \in \mathbb{R}^{272 \times 64 \times 64}$ . We conduct experiments to verify the resolution of the concatenated features; see Tab. 2. Increasing resolution from  $16 \times 16$  to  $64 \times 64$  increases the PRO by 2.5%. Combining EfficientNet features of stride (2, 4, 8, 16) improves the PRO by 3.5%.

**Unified model.** Since our denoising diffusion model is capable of modeling complex real industrial data distribution, we conduct experiments to use a single model for anomaly localization for all categories of MVTec-AD, see Tab. 3. The performance of DRAEM [40] and PaDiM [8] drops greatly for the unified setting, with more than 5% degradation in the PRO metric. In contrast, our model still achieves 93.0% in PRO, with less than 1.1% performance drop.

#### 4.5. Qualitative results of reconstruction.

We show in Fig. 7 that our denoising gradient process can smoothly transform an anomalous image into a normal one under the guidance of a pre-trained feature extractor. The left two columns are ground truth and input anomalous

Resize	Stride				AUROC	PRO
	16	8	4	2		
64	✓				94.2	90.2
	✓	✓			95.8	91.9
	✓	✓	✓		96.4	93.2
	✓	✓	✓	✓	<b>96.7</b>	<b>93.7</b>
32	✓	✓	✓	✓	96.6	92.4
16	✓	✓	✓	✓	95.5	91.2

Table 2: Ablation study of feature level reconstruction. Experiments are conducted with features from different layers of EfficientNet [30], resized to different resolutions.

Method	Base		Unified	
	AUROC	PRO	AUROC	PRO
PaDiM[8]	<b>97.8</b>	92.8	90.5	85.3
DRAEM[40]	97.3	93.0	89.4	82.2
UniAD[36]	96.6	-	<b>97.0</b>	91.1
Ours	96.7	<b>94.1</b>	96.0	<b>93.0</b>

Table 3: Comparison of a single unified model for anomaly localization of all the categories on MVTec-AD.