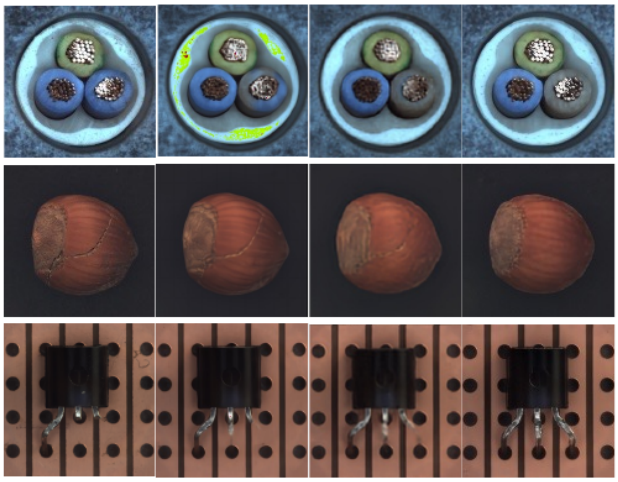
Class	Embedding-based			Reconstruction-based			
	Spade [6]	PaDiM [8]	CFlow [13]	DRAEM[40]	UniAD [36]	Ours	Ours*
bottle	(98.4, 95.5)	(98.5, 95.3)	(99.0, 96.8)	(99.1, **97.2**)	(98.0, 93.8)	(97.7,95.0)	(97.7,95.0)
cable	(97.2, 90.9)	(98.1, 91.1)	(97.7, 93.5)	(94.7, 76.0)	(97.2, 86.3)	(95.2,88.7)	(95.6,**89.5**)
capsule	(99.0, 93.7)	(98.8, 92.3)	(99.0, 93.4)	(94.3, **91.7**)	(98.7, 90.8)	(98.0,90.1)	(97.5,91.4)
carpet	(97.5, 94.7)	(98.9, 94.5)	(99.3, 97.7)	(95.5, 92.9)	(98.4, 94.5)	(98.9,95.8)	(98.9,**95.8**)
grid	(93.7, 86.7)	(96.1, 90.5)	(99.0, 96.1)	(99.7, 98.4)	(97.5, 92.6)	(99.1,98.1)	(99.1,**98.4**)
hazelnut	(99.1, 95.4)	(98.4, 84.0)	(98.9, 96.7)	(99.7, **98.1**)	(98.2, 93.0)	(97.7,89.5)	(97.3,91.1)
leather	(97.6, 97.2)	(99.2, 97.9)	(99.7, 99.4)	(98.6, 98.0)	(98.7, 97.2)	(99.5,99.1)	(99.5,**99.1**)
metal nut	(98.1, 94.4)	(98.0, 92.9)	(98.6, 91.7)	(99.5, **94.1**)	(94.9, 87.1)	(96.8,93.0)	(96.8,93.0)
pill	(96.5, 94.6)	(97.0, 95.3)	(99.0, 95.4)	(97.6, 88.9)	(96.2, **95.3**)	(92.5,94.5)	(92.5,94.5)
screw	(98.9, 96.0)	(98.7, 94.6)	(98.9, 95.3)	(97.6, **98.2**)	(98.9, 95.3)	(99.0,95.6)	(99.0,95.6)
tile	(87.4, 75.9)	(94.3, 93.7)	(98.0, 94.3)	(99.2, **98.9**)	(92.0, 79.6)	(92.1,95.1)	(92.1,95.1)
toothbrush	(97.9, 93.5)	(98.7, 94.3)	(98.9, 95.1)	(98.1, 90.3)	(98.3, 88.2)	(98.9,94.7)	(98.6,**95.7**)
transistor	(94.1, 87.4)	(97.9, 91.4)	(98.0, 81.4)	(90.9, 81.6)	(97.9, **93.9**)	(92.6,89.7)	(93.1,90.1)
wood	(88.5, 87.4)	(95.7, 89.3)	(96.7, 95.8)	(96.4, **94.6**)	(93.0, 86.0)	(94.7,92.9)	(94.5,93.0)
zipper	(96.5, 92.6)	(98.5, 95.0)	(99.0, 96.6)	(98.8, **96.3**)	(97.7, 93.2)	(97.6,93.6)	(97.6,93.6)
average	(96.0, 91.7)	(97.8, 92.8)	(98.6, 94.6)	(97.3,93.0)	(97.0, 91.1)	(96.7,93.7)	(96.7,**94.1**)

表1：MVTec-AD [3] データセットにおける最先端の異常検出手法との比較。異常局所化性能をピクセル単位のAUROCとPROメトリクスで比較し、表では(AUROC, PRO)と表記しています。再構築ベースの手法の中で最も高いPROスコアを強調表示しています。他の再構築ベースの手法と比べて大幅に劣る結果は赤色で表記しています。すべてのカテゴリで同じハイパーパラメータを使用した場合の結果を「Ours」とし、各カテゴリごとに調整された異なるハイパーパラメータを使用した場合の結果を「Ours*」として示します。



(a) Input (b) DRAEM (c) UniAD (d) Ours

図4：MVTec-ADの「ケーブル」と「ヘーゼルナッツ」クラスにおける再構築結果の比較。3つのカテゴリにおける異常の種類は、色入れ替え、ひび割れ、およびリード切断です。

地面真値、当社のピクセルレベル異常予測、当社の特徴量レベル異常予測、および元の画像上に可視化された最終結果を表します。当社のノイズ除去モデルが異常の境界を正確に推定できることを示しています。

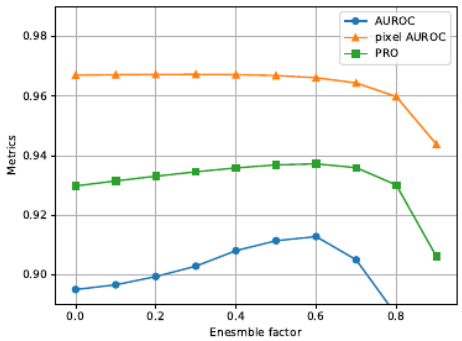


図5：異なるアンサンブル要因αの影響。

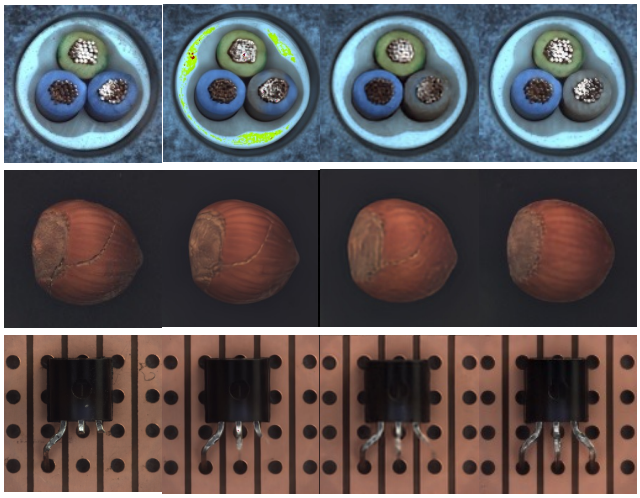
4.4. Ablation study

アンサンブル要因。ピクセルレベルと特徴レベルの両方のlの予測を組み合わせて最終的な異常スコアを取得します。アンサンブルウェイトの影響を検証するための実験を実施しました。図5に示すように、最適なアンサンブル係数α = 0.5を選択しました。PROメトリクスは、特徴量レベル異常スコアのみを使用する場合と比べて0.7%改善されました。

事前学習済み特徴抽出器。ResNet [14] と WideResNet [39] で抽出された深層特徴は高次元であり、当社のノイズ除去モデルが特徴を再構築する際に大きな困難をもたらします。代わりに、当社は

Class	Embedding-based			Reconstruction-based			
	Spade [6]	PaDiM [8]	CFlow [13]	DRAEM[40]	UniAD [36]	Ours	Ours*
bottle	(98.4, 95.5)	(98.5, 95.3)	(99.0, 96.8)	(99.1, **97.2**)	(98.0, 93.8)	(97.7,95.0)	(97.7,95.0)
cable	(97.2, 90.9)	(98.1, 91.1)	(97.7, 93.5)	(94.7, 76.0)	(97.2, 86.3)	(95.2,88.7)	(95.6,**89.5**)
capsule	(99.0, 93.7)	(98.8, 92.3)	(99.0, 93.4)	(94.3, **91.7**)	(98.7, 90.8)	(98.0,90.1)	(97.5,91.4)
carpet	(97.5, 94.7)	(98.9, 94.5)	(99.3, 97.7)	(95.5, 92.9)	(98.4, 94.5)	(98.9,95.8)	(98.9,**95.8**)
grid	(93.7, 86.7)	(96.1, 90.5)	(99.0, 96.1)	(99.7, 98.4)	(97.5, 92.6)	(99.1,98.1)	(99.1,**98.4**)
hazelnut	(99.1, 95.4)	(98.4, 84.0)	(98.9, 96.7)	(99.7, **98.1**)	(98.2, 93.0)	(97.7,89.5)	(97.3,91.1)
leather	(97.6, 97.2)	(99.2, 97.9)	(99.7, 99.4)	(98.6, 98.0)	(98.7, 97.2)	(99.5,99.1)	(99.5,**99.1**)
metal nut	(98.1, 94.4)	(98.0, 92.9)	(98.6, 91.7)	(99.5, **94.1**)	(94.9, 87.1)	(96.8,93.0)	(96.8,93.0)
pill	(96.5, 94.6)	(97.0, 95.3)	(99.0, 95.4)	(97.6, 88.9)	(96.2, **95.3**)	(92.5,94.5)	(92.5,94.5)
screw	(98.9, 96.0)	(98.7, 94.6)	(98.9, 95.3)	(97.6, **98.2**)	(98.9, 95.3)	(99.0,95.6)	(99.0,95.6)
tile	(87.4, 75.9)	(94.3, 93.7)	(98.0, 94.3)	(99.2, **98.9**)	(92.0, 79.6)	(92.1,95.1)	(92.1,95.1)
toothbrush	(97.9, 93.5)	(98.7, 94.3)	(98.9, 95.1)	(98.1, 90.3)	(98.3, 88.2)	(98.9,94.7)	(98.6,**95.7**)
transistor	(94.1, 87.4)	(97.9, 91.4)	(98.0, 81.4)	(90.9, 81.6)	(97.9, **93.9**)	(92.6,89.7)	(93.1,90.1)
wood	(88.5, 87.4)	(95.7, 89.3)	(96.7, 95.8)	(96.4, **94.6**)	(93.0, 86.0)	(94.7,92.9)	(94.5,93.0)
zipper	(96.5, 92.6)	(98.5, 95.0)	(99.0, 96.6)	(98.8, **96.3**)	(97.7, 93.2)	(97.6,93.6)	(97.6,93.6)
average	(96.0, 91.7)	(97.8, 92.8)	(98.6, 94.6)	(97.3,93.0)	(97.0, 91.1)	(96.7,93.7)	(96.7,**94.1**)

Table 1: Compare with the state-of-the-art anomaly detection approaches on MVTec-AD [3] dataset. We compare anomaly localization performance with pixel-wise AUROC and PRO metrics, denoted as (AUROC, PRO) in the table. We highlight the best PRO scores among all the reconstruction-based methods. We denote the result in Red if the result underperforms other reconstruction-based methods by a large margin. We show our results with the same hyperparameters for all categories, denoted as Ours, and different hyperparameters adjusted for each category, denoted as Ours*.



(a) Input (b) DRAEM (c) UniAD (d) Ours

Figure 4: Comparisons of the reconstruction results on class cable and hazelnut of MVTec-AD. The anomaly types are color-swap, crack, and cut-lead for the three categories.

represent ground truth, our pixel-level anomaly predictions, our feature-level anomaly predictions, and the final results visualized on the original images. We show that our denoising model is capable of precise boundary estimation of anomalies.

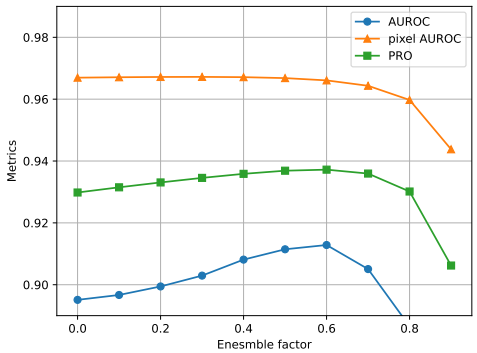


Figure 5: The effects of different ensembling factor α.

4.4. Ablation study

Ensemble factor. We combine the pixel-level and feature-level 11 predictions to get the final anomaly score. We conduct experiments to verify the effects of the ensembling weights. As Fig. 5 shows, we choose the best ensemble factor $\alpha = 0.5$. The PRO metric is improved by 0.7% compared with only using the feature-level anomaly score.

Pretrained feature extractor. We observe that the deep features extracted by ResNet [14] and WideResNet [39] are of high dimensional, which brings great difficulties for our denoising model to reconstruct the features. Instead, we