DecAD: Decoupling Anomalies in Latent Space for Multi-Class Unsupervised Anomaly Detection

Supplementary Material

A. Overview

This supplementary material consists of:
- *Invertible Neural Network Details*
- *Supplementary on Comparison*
- *Supplementary on Ablation Study*
- *Visualization for Real-IAD*
- *Limitations*

B. Invertible Neural Network Details

B.1. The Structure of Invertible Mapping Unit

According to the methodology in the main text, our invertible neural network (INN) $\mathcal{F}$ consists of a series of block F_r, where the output of each block is normalized by 'ActNorm' [3] to maintain invertibility of INN and stable training dynamics. Mathematically,

$$\mathcal{F} = F_R \circ F_{R-1} \cdots \circ F_1, \tag{1}$$

where R is the number of invertible blocks (see Fig. 1). Furthermore, each layer F_r also consists of multiple invertible mapping unit f_k, i.e.,

$$F_r = f_K^r \circ f_{K-1}^r \circ \cdots f_1^r. \tag{2}$$

The concrete structure of f_k is shown in Fig. 1. Each f_k is based on invertible mapping $\mathbf{G}$, i.e.,

$$f(x) = \left(\frac{\mathbf{Id} + \mathbf{G}}{2}\right)^{-1}(x) - x. \tag{3}$$

Following INN-based work [1], we adopt i-DenseNet [11] to parametrize $\mathbf{G}$. Based on $\mathbf{G}$ and Eq. (3), we can model f_k, which is named as 'Induced Norm Linear'. To ensure that each f_k is 1-Lipchitz, we select activation function 'Pila' [1] to process the output of 'Induced Norm Linear', formulated as:

$$\text{Pila}(x) = \begin{cases} x & \text{if } x \geq 0, \\ \left(\frac{k^2}{2}x^3 - kx^2 + x\right)e^{kx} & \text{if } x < 0, \end{cases}$$

where k is treated as a fixed hyperparameter and set to 5 in our experiments. Finally, consistent with Eq. (2) and Eq. (1), the whole INN consists of $R = 3$ blocks and each block includes $K = 2$ units.

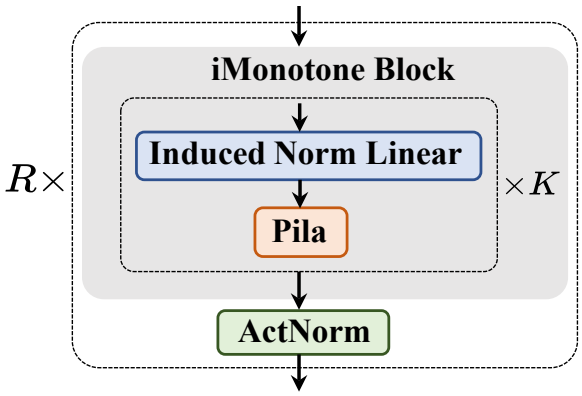

Figure 1. The structure of invertible mapping unit. The 'induced norm linear' is applied to model f_k. 'Pila' is 1-Lipchitz activation function. Each 'iMonotone block' consists of K 'induced norm linear'. The whole INN consists of R 'iMonotone blocks'.

C. Supplementary on Comparison

C.1. Comparison with More SOTA Methods

The methods shown in the main text are mainly reconstruction-based methods [2, 7, 15, 17]. Therefore, for completeness, we will compare more SOTA methods in Table 1, which includes some augmentation-based methods, such as DRAEM [16], DeSTSeg [18], and OmniAL [19] and feature-embedding-based SOTA methods, such as PatchCore [12]. We found that OmniAL achieves good anomaly localization on MVTec AD dataset, but fails on the VisA dataset. We believe that there is a huge gap between the pseudo defects generated by the traditional augmentation method of OmniAL and the fine-grained real defects in the VisA dataset, which leads to its low performance on the VisA dataset. In addition, the performance of PatchCore in multi-class task is much lower than that in single-class task, because of the increased diversity of patch embedding vectors in the core set, which makes it more difficult to deploy distance-based matching algorithms. Furthermore, we compare recent SOTA methods [4, 6, 9, 10]. Our method outperforms recent reconstruction-based methods, MoEAD [10] and OneNIP [4], by $+1.3$ and $+1.1$ in terms of image-level AUROC on MVTec AD. Compared to the latest work, MambaAD [6], our DecAD improves image-level and pixel-level AUROC by $+0.4/0.1$ on MVTec AD, and image-level AUROC $+0.4$ on VisA.

C.2. Comparison on One-Class AD Task

Anomaly leakage also exists in one-class AD task. Therefore, we compare some one-class unsupervised AD methods [5, 8, 12, 13, 18] on MVTec AD. Our DecAD shows a superior performance in terms of 98.6 P-AUROC and 95.7 AUPRO under one-class AD setting in Table 2.

Table 1. Anomaly detection results with image-level AUROC, pixel-level AUROC metrics on MVTec AD dataset. The publication information corresponding to each method is also displayed in this table. **Bold**/underline values indicate the best/runner-up. I: image-level AUROC. P: pixel-level AUROC.

Method	Pub&Year	MVTec AD		VisA	
		I	P	I	P
DRAEM [16]	ICCV2021	88.1	87.2	88.9	96.1
PatchCore [12]	CVPR2022	96.4	95.7	91.5	97.9
DeSTSeg [18]	CVPR2023	89.2	93.1	88.9	96.1
OmniAL [19]	CVPR2023	97.2	**98.3**	87.8	96.6
LafitE [14]	Arxiv2023	98.5	97.6	–	–
HVQ-Trans [9]	NIPS2023	98.0	97.3	93.2	**98.7**
MoEAD [10]	ECCV2024	97.7	97.0	93.1	**98.7**
OneNIP [4]	ECCV2024	97.9	97.9	92.5	**98.7**
MambaAD [6]	NIPS2024	98.6	97.7	94.3	98.5
Ours	–	**99.0**	97.8	**94.7**	98.5

Table 2. Anomaly detection results with image-level AUROC, pixel-level AUROC, and AUPRO on MVTecAD dataset under one-class AD setting. **Bold**/underline values indicate the best/runner-up.

Method	I-AUROC	P-AUROC	AUPRO
DeSTSeg [18]	99.0	97.9	-
SimpleNet [8]	99.6	98.1	–
MemKD [5]	99.6	98.2	94.5
RD++ [13]	99.4	98.3	95.0
PatchCore-L [12]	99.6	98.2	93.5
Ours	**99.7**	**98.6**	**95.7**

C.3. Complexity Analysis

As shown in Table 3, compared to the augmentation-based method, our DecAD outperforms DRAEM [16] on three efficiency measures. DRAEM utilizes image processing techniques such as texture cropping and pasting to generate image-level anomaly samples, which is time-consuming. However, our DecAD adopts feature-level anomaly synthesis, which only requires adding adversarial perturbations to patch embeddings. This process only involves gradient calculation, significantly reducing synthesis time. Compared with the previous CNN-based distillation method, RD4AD [2], our method has fewer learnable parameters. However, since our DecAD is based on the ViT framework, it invisibly increases the computational complexity, thereby increasing FLOPs. Finally, compared to UniAD [15], our method only requires 67 epochs (batch size is 16) for training to achieve better performance, while UniAD requires 1000 epochs for training, which invisibly increases the time cost of training.

Table 3. Complexity comparison between our DecAD and other important methods on MVTec AD. Our method requires 15,000 iterations for training, which is approximately 67 epochs. Learnable parameters, FLOPs, and train epochs are reported in this table. **Bold**/underline values indicate the best/runner-up.

Method	Learnable Parameters	FLOPs	Train Epoch
DRAEM[16]	97.4M	198G	700
RD4AD[2]	80.6M	24.6G	150
UniAD[15]	24.5M	**3.6G**	1000
Ours	**23.2M**	48.6G	**67**

D. Supplementary on Ablation Study

D.1. Block Choices in Encoder

In this work, we apply ViT-B/14 pre-trained by DINOv2 as our encoder, which consists of 12 residual attention blocks. Extensive experiments show that selecting layers with either very high or very low indexes for reconstruction will degrade detection performance, as shown in Table 4. The best performance is achieved when the selected indexes are close to the middle and the step between them is three (e.g., [3, 6, 9] in Table 4). We believe that the features extracted from blocks with lower indexes contain more geometric information, while those from higher indexes blocks encapsulate more abstract semantic information. Since defects often have irregular geometric structures and unusual colors, selecting features from the front or middle layers is more suitable for reconstruction-based frameworks.

Table 4. Impact of different indexes of residual attention blocks in the frozen encoder. Image-level AUROC and pixel-level AUROC are reported on **MVTec AD** dataset. **Bold**/underline values indicate the best/runner-up.

Indexes	I/P-AUROC	Indexes	I/P-AUROC
[1 6 12]	96.3/95.8	[2 8 12]	97.8/96.7
[2 4 8]	97.5/96.4	[3 6 10]	98.3/97.2
[4 8 10]	98.7/97.6	[3 6 9]	**99.0/97.8**

D.2. Performance Gain of DecAD as a Plugin

Our proposed DecAD consists of anomaly synthesis and anomaly decoupling and can be deployed as a plugin for existing reconstruction and distillation based methods. Therefore, we test DecAD as a plugin to enhance existing unsupervised AD methods [2, 6, 15]. We report the mean of seven metrics on MVTec AD and VisA in Table 5. Upon RD4AD [2], our method improves its performance by **+3.3** and **+1.2** on MVTec AD and VisA, respectively. For reconstruction-based UniAD [15], our DecAD obtains

Table 5. The performance gain of DecAD as a Plugin. We report the mean of seven metrics on MVTec AD and VisA datasets.

MVTec AD		VisA	
Method	**Mean**	**Method**	**Mean**
RD4AD [2]	82.3	RD4AD [2]	77.8
UniAD [15]	81.7	UniAD [15]	74.6
MambaAD [6]	86.0	MambaAD [6]	78.7
RD4AD+**Ours**	**85.6**	RD4AD+**Ours**	**79.0**
UniAD+**Ours**	**83.5**	UniAD+**Ours**	**76.5**
MambaAD+**Ours**	**86.8**	MambaAD+**Ours**	**79.9**

a gain of **+1.8** and **+1.9** on MVTec AD and VisA. Moreover, based on MambaAD [6], our work also works well and improves its performance by **+0.8** and **+1.2** on the two datasets. Since these three methods apply pre-trained encoders different from our DecAD, Table 5 provides more intuitive comparisons, demonstrating the effectiveness of our method under different network settings.

D.3. Choices of Magnitude δ in Adversarial Noise

In this section, we investigate the impact of different δ in adversarial noise $\epsilon^* = \delta * \|\nabla_{\mathbf{x}}(\mathcal{H}(\mathcal{D}(\mathbf{x})))\|_2$. Selecting $\delta = 2.5$, we obtain the best classification result of 99.0 in terms of image-level AUROC. The best anomaly localization performance is achieved through setting $\delta = 3$, where pixel-level AUROC is 97.9 and AUPRO is 92.7. When we select a lower magnitude $\delta < 2$, anomaly classification and localization are degraded, as shown in Table 6. In our main text, we select $\delta = 2.5$ to achieve the best anomaly classification.

Table 6. Impact of different magnitudes δ in adversarial noise ϵ^*. Image-level AUROC, pixel-level AUROC, and AUPRO are reported on **MVTec AD** dataset. **Bold**/underline values indicate the best/runner-up.

Magnitude δ	**I-AUROC**	**P-AUROC**	**AUPRO**
$\delta = 1.0$	98.3	97.2	91.7
$\delta = 1.5$	98.2	97.4	92.1
$\delta = 2.0$	98.6	97.6	92.5
$\delta = 2.5$	**99.0**	97.8	92.5
$\delta = 3.0$	98.6	**97.9**	**92.7**

E. Visualization for Real-IAD

In our main paper, we show anomaly localization on MVTec AD and VisA datasets. For the integrity of the results, we show the visualization for Real-IAD in Fig.2.

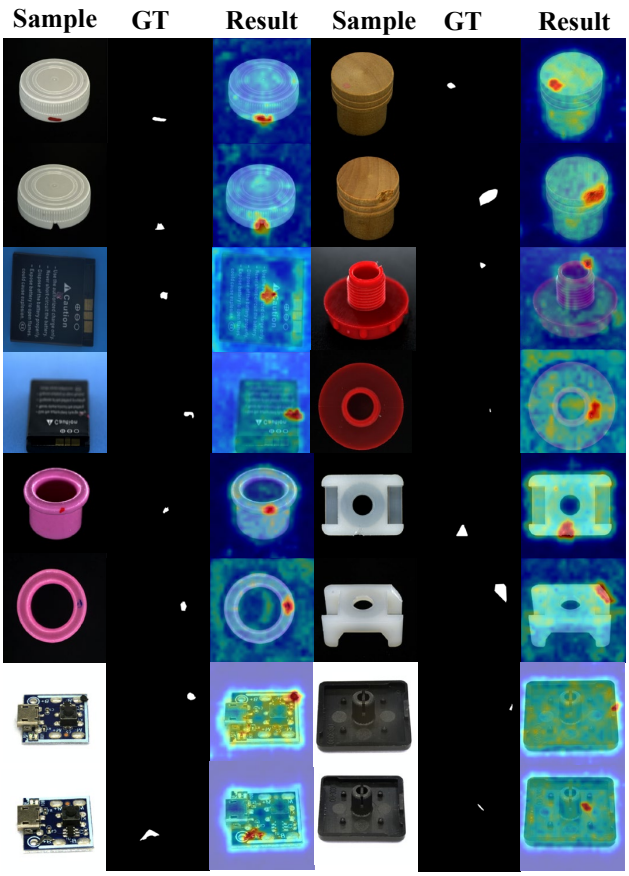

Figure 2. The visualization for Real-IAD dataset.

F. Limitations

F.1. Lack of Multi-scale Feature Information

Since our framework is based on the ViT, the size of each layer output remains consistent. Therefore, our framework lacks multi-scale feature information, which will degrade the detection performance of the model for fine-grained defects. On the contrary, the CNN-based method, RD4AD [2], can extract multi-scale feature information for distillation, which makes them perform well on datasets such as VisA with more fine-grained defects. In the future, we will focus on ViT network with multi-scale information for anomaly detection.

F.2. Common Issue in Distillation Frameworks

The existing anomaly detection distillation frameworks apply encoders pre-trained on natural datasets to extract features. However, there is a huge gap between natural and industrial data, which makes the encoded features unable to fully represent the patterns of anomalies. Although some works adopt adapters after the frozen encoder, they cannot completely overcome this issue. Our DecAD frame-

work also applies an encoder pre-trained on natural images, which essentially may not extract representative features for distillation.

References

[1] Byeongkeun Ahn, Chiyoon Kim, Youngjoon Hong, and Hyunwoo J Kim. Invertible monotone operators for normalizing flows. In *NeurIPS*, 2022. 1

[2] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, 2022. 1, 2, 3

[3] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016. 1

[4] Bin-Bin Gao. Learning to detect multi-class anomalies with just one normal image prompt. In *ECCV*, 2025. 1, 2

[5] Zhihao Gu, Liang Liu, Xu Chen, Ran Yi, Jiangning Zhang, Yabiao Wang, Chengjie Wang, Annan Shu, Guannan Jiang, and Lizhuang Ma. Remembering normality: Memory-guided knowledge distillation for unsupervised anomaly detection. In *CVPR*, 2023. 1, 2

[6] Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2404.06564*, 2024. 1, 2, 3

[7] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *AAAI*, 2024. 1

[8] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *CVPR*, 2023. 1, 2

[9] Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. In *NeurIPS*, 2023. 1, 2

[10] Shiyuan Meng, Wenchao Meng, Qihang Zhou, Shizhong Li, Weiye Hou, and Shibo He. Moead: A parameter-efficient model for multi-class anomaly detection. In *ECCV*, 2020. 1, 2

[11] Yura Perugachi-Diaz, Jakub Tomczak, and Sandjai Bhulai. Invertible densenets with concatenated lipswish. In *NeurIPS*, 2021. 1

[12] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, 2022. 1, 2

[13] Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan Duong, Chanh D Tr Nguyen, and Steven QH Truong. Revisiting reverse distillation for anomaly detection. In *CVPR*, 2023. 1, 2

[14] Haonan Yin, Guanlong Jiao, Qianhui Wu, Borje F Karlsson, Biqing Huang, and Chin Yew Lin. Lafite: Latent diffusion model with feature editing for unsupervised multi-class anomaly detection. *arXiv preprint arXiv:2307.08059*, 2023. 2

[15] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. In *NeurIPS*, 2022. 1, 2, 3

[16] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draema discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, 2021. 1, 2

[17] Jiangning Zhang, Xuhai Chen, Yabiao Wang, Chengjie Wang, Yong Liu, Xiangtai Li, Ming-Hsuan Yang, and Dacheng Tao. Exploring plain vit reconstruction for multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2312.07495*, 2023. 1

[18] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *CVPR*, 2023. 1, 2

[19] Ying Zhao. Omnial: A unified cnn framework for unsupervised anomaly localization. In *CVPR*, 2023. 1, 2