One-for-More: Continual Diffusion Model for Anomaly Detection

Supplementary Material

A. Experimental details

Data pre-processing. We employ the data pre-processing pipeline specified in DiAD [3] for both the MVTec [1] and VisA [11] datasets to mitigate potential train-test discrepancies. This involves channel-wise standardization using precomputed mean $[0.485, 0.456, 0.406]$ and standard deviation $[0.229, 0.224, 0.225]$ after normalizing each RGB image to $[0, 1]$.

Patch perturbation. We adopt the method proposed by NSA [8] for patch perturbation on the original images. The NSA method builds upon the Cut-paste technique [4] and enhances it by incorporating the Poisson image editing method [7] to alleviate the discontinuities caused by pasting image patches. The cut-paste method is commonly used in the anomaly detection domain to generate simulated anomalous images. It involves randomly cropping a patch from one image and pasting it onto a random location in another image, thus creating a simulated anomaly. The Poisson-based pasting method seamlessly blends the cloned object from one image into another by solving Poisson partial differential equations, thereby better simulating a realistic anomalous region. In this paper, the number of patches is set as a random value from 1 to 4, and the patch size is a random value from 0.03 to 0.4 of the original image size. The visualization of patch perturbation is shown in Figure 1.

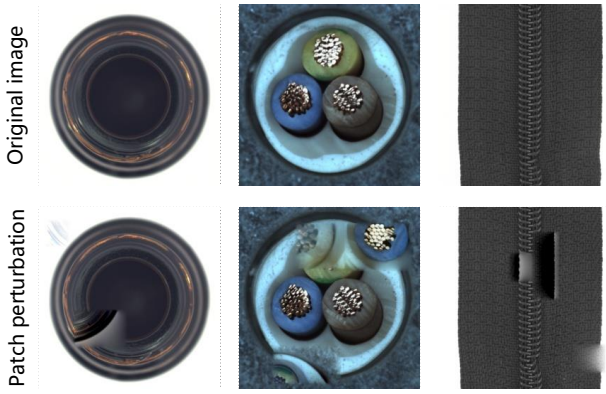

Figure 1. Qualitative results of logical anomaly detection.

Evaluation metrics. We follow the literature [1] in reporting the Area Under the Receiver Operation Characteristic (AUROC) for both image-level and pixel-level anomaly detection. To measure the performance of the model in continuous learning, referring to DNE [5], we calculated the average AUROC (A-AUROC) and the forgetting measure (FM) for N continual steps. Specially, we define $A^{\text{pix}}_{N,i}$ and $A^{\text{img}}_{N,i}$ as the test AUROC of task i after training on task N.

$$\text{A-AUROC} = \begin{cases} \frac{1}{N}\sum_{i=1}^{N-1} A^{\text{pix}}_{N,i} \\ \frac{1}{N}\sum_{i=1}^{N-1} A^{\text{img}}_{N,i} \end{cases}, \tag{1}$$

$$\text{FM} = \begin{cases} \frac{1}{N-1}\sum_{i=1}^{N-1} \max_{b\in\{1,\cdots,N-1\}} \left(A^{\text{pix}}_{b,i} - A^{\text{pix}}_{N,i}\right) \\ \frac{1}{N-1}\sum_{i=1}^{N-1} \max_{b\in\{1,\cdots,N-1\}} \left(A^{\text{img}}_{b,i} - A^{\text{img}}_{N,i}\right) \end{cases} \tag{2}$$

In addition to the results for the AUROC documented in the body of the paper, We also supplement the image-level Precision-Recall (AUPR) results and pixel-level Per-region-overlap (PRO) [1, 2] results. Referring to Equation (1) and Equation (2), we calculate A-AUPR, A-PRO, and their FM to evaluate our method. The results are shown in Table 1 2 4 5. Our method still achieves an advanced level in the above metrics.

B. Memory Analysis of iSVD

In Section 3.2, considering the storage of the original matrix and $\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V}$ during SVD, the memory overhead of SVD is $d\Lambda + d^2 + \Lambda^2 + \min(d, \Lambda)$, while iSVD uses a memory overhead of $d(m+k) + d^2 + (m+k)^2 + \min(d, m+k)$. It is known that $\Lambda \gg d$, $m > d > k$ and $\Lambda = mn$, thus, the memory saving rate of iSVD is about:

$$\begin{aligned} &\frac{d\Lambda + d^2 + \Lambda^2 + d - [d(m+k) + d^2 + (m+k)^2 + d]}{d\Lambda + d^2 + \Lambda^2 + d} \\ =&\frac{d\Lambda + \Lambda^2 - d(m+k) - (m+k)^2}{\Lambda^2} \Big/ \frac{d\Lambda + d^2 + \Lambda^2 + d}{\Lambda^2} \\ =&\frac{d}{\Lambda} + 1 - \frac{d(m+k)}{\Lambda^2} - \frac{(m+k)^2}{\Lambda^2} \\ \approx&1 - \frac{m^2 + 2mk + k^2}{m^2 n^2} \\ =&1 - \frac{1}{n^2} - \frac{2k}{m\Lambda} - \frac{k^2}{\Lambda^2} \approx \frac{n^2 - 1}{n^2}. \end{aligned} \tag{3}$$

In practice, the actual memory saving rate differs from the theoretical value due to factors such as memory sharing. Taking the intermediate features of ten images as an example, Figure 2 shows the actual and theoretical memory saving rate of splitting the feature matrix into n blocks for iSVD. Although there are some differences between the theoretical value and the actual value, the general trend is consistent.

Method	14 – 1 with 1 Step		10 – 5 with 1 Step		3 × 5 with 5 Steps		10 – 1 × 5 with 5 Steps	
	A-AUPR (↑)	**FM (↓)**	**A-AUPR (↑)**	**FM (↓)**	**A-AUPR (↑)**	**FM (↓)**	**A-AUPR (↑)**	**FM (↓)**
UCAD* [6]	95.8	0.26	95.0	0.98	93.1	2.02	95.5	0.07
IUF [9]	97.8	0.25	95.4	1.92	91.1	2.86	95.3	0.16
ControlNet [10]	97.2	1.55	96.7	1.76	86.7	6.40	89.0	7.43
DiAD [3]	97.4	0.71	96.4	1.85	89.1	4.31	91.4	4.83
CDAD	**98.4**	**0.08**	**98.3**	**0.55**	**95.8**	**1.88**	**98.4**	**0.02**

Table 1. Image-level A-AUPR of our method on MVTec under 4 continual anomaly detection settings. The best and second-best results are marked in **blod** and underline. $*$ indicates memory limited.

Method	14 – 1 with 1 Step		10 – 5 with 1 Step		3 × 5 with 5 Steps		10 – 1 × 5 with 5 Steps	
	A-PRO (↑)	**FM (↓)**	**A-PRO (↑)**	**FM (↓)**	**A-PRO (↑)**	**FM (↓)**	**A-PRO (↑)**	**FM (↓)**
UCAD [6]	86.3	1.16	80.7	2.89	71.1	7.48	80.8	1.19
IUF [9]	88.6	0.62	85.0	3.22	72.9	5.79	84.3	2.41
ControlNet [10]	88.5	1.75	85.8	4.70	71.5	10.0	77.9	7.11
DiAD [3]	88.9	0.85	87.4	3.91	72.1	9.23	83.1	2.93
CDAD	**89.8**	**0.47**	**88.9**	**2.57**	**83.8**	**4.05**	**89.2**	**1.16**

Table 2. Pixel-level A-PRO of our method on MVTec under 4 continual anomaly detection settings. The best and second-best results are marked in **blod** and underline. $*$ indicates memory limited.

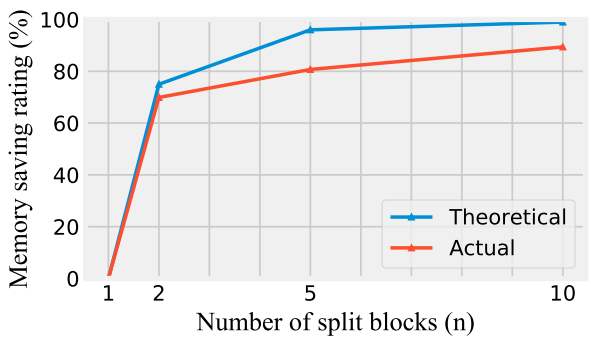

Figure 2. The theoretical and actual values of the memory saving ratio when the different number of split blocks is used.

In this paper, we determine the number of blocks n according to the number of images of the old task. Specifically, we first sample the old task dataset, randomly retain 10% of the images, and then group according to the number of sampled images (denoted as N_{img}). The number of groups in iSVD is $n = \frac{N_{img}}{e}$. In this paper, e is set to 1 by default, that is, the intermediate features of each image are divided into a separate group for iSVD operation.

In addition, we analyze the impact of different e on the model and the time and memory overhead. Table 3 records, for setting different e, the anomaly detection results of our model on MVTec setting 2, as well as the time and memory overhead for computing the significant representation of the old task. When e is set to different values, the anomaly detection results and forgetting rate of the model will not have much influence. Although we discussed in Section 4.5 that the large number of split blocks will affect the performance of iSVD, it will not impair its representation ability of core information, so it can still ensure the continuous learning ability of the model. Table 3 also shows that with the increase of e, the memory consumption increases, but the time cost decreases, which indicates that although iSVD can greatly alleviate the pressure of memory, it will bring extra time cost.

	A-AUROC	**FM**	**Memory**	**Times**
e=1	94.2 / 95.3	2.05 / 2.40	16.7GB	33.5h
e=2	94.1 / 95.4	2.10 / 2.32	30.3GB	28.3h
e=4	94.4 / 95.6	1.95 / 2.21	57.9GB	22.6h
e=6	94.5 / 95.7	1.91 / 2.23	85.9GB	20.2h

Table 3. The impact of different e on the model and the time and memory overhead of iSVD under MVTec Setting 2.

C. Qualitative Results

We supplement the qualitative results on MVTec and VisA datasets, which show the localization image reconstruction results and anomaly localization results for the seven tasks, respectively, as shown in Figure 3-9. Our method not only overcomes the "faithfulness hallucination" problem of the diffusion model but also shows excellent anomaly localization results.

Method	11 – 1 with 1 Step		8 – 4 with 1 Step		8 – 1 × 4 with 4 Steps	
	A-AUROC (↑)	**FM (↓)**	**A-AUROC (↑)**	**FM (↓)**	**A-AUROC (↑)**	**FM (↓)**
UCAD* [6]	88.1	0.29	83.2	5.17	82.9	2.16
IUF [9]	**91.6**	-0.04	83.4	7.51	83.0	6.87
ControlNet [10]	85.2	2.38	78.8	6.25	72.4	4.56
DiAD [3]	74.9	5.42	70.1	12.29	59.5	9.55
CDAD	89.4	**-0.77**	**85.3**	**3.1**	**84.7**	**1.83**

Table 4. Image-level A-AUPR of our method on VisA under 3 continual anomaly detection settings. The best and second-best results are marked in **blod** and underline. ∗ indicates memory limited.

Method	11 – 1 with 1 Step		8 – 4 with 1 Step		8 – 1 × 4 with 4 Steps	
	A-PRO (↑)	**FM (↓)**	**A-PRO (↑)**	**FM (↓)**	**A-PRO (↑)**	**FM (↓)**
UCAD* [6]	80.4	2.02	72.4	7.46	70.5	9.83
IUF [9]	82.0	1.04	63.9	20.8	57.0	23.95
ControlNet [10]	62.3	2.45	61.0	1.81	51.7	10.38
DiAD [3]	69.9	4.33	67.7	8.29	55.0	11.74
CDAD	**81.6**	**-0.22**	**78.9**	**1.59**	**77.7**	**1.66**

Table 5. Pixel-level A-PRO of our method on VisA under 3 continual anomaly detection settings. The best and second-best results are marked in **blod** and underline. ∗ indicates memory limited.

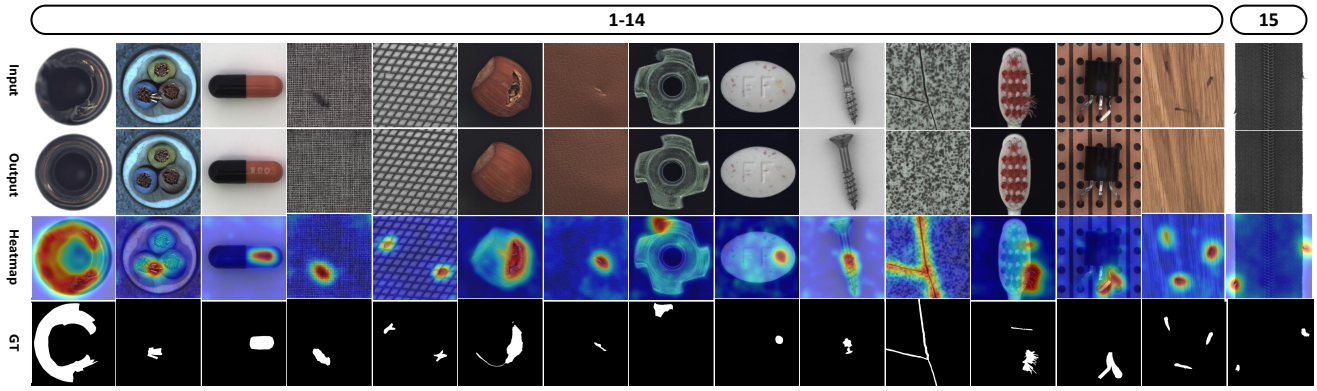

Figure 3. Qualitative comparison results under setting 1 of MVTec, the numbers represent continual training classes.

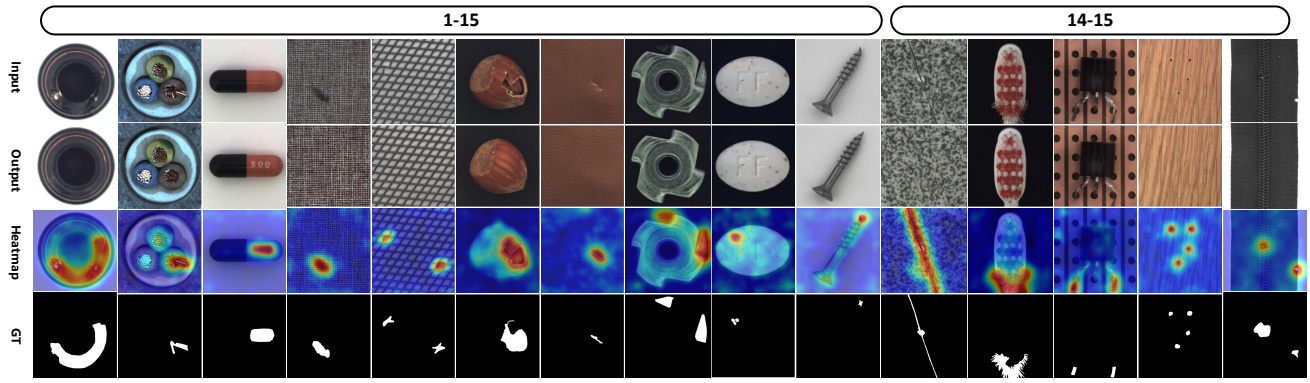

Figure 4. Qualitative comparison results under setting 2 of MVTec, the numbers represent continual training classes.

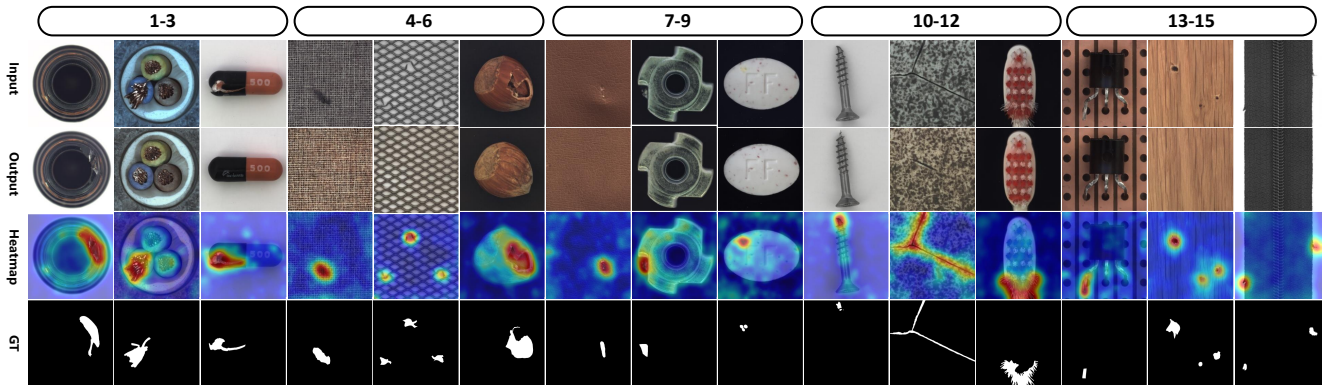

Figure 5. Qualitative comparison results under setting 3 of MVTec, the numbers represent continual training classes.

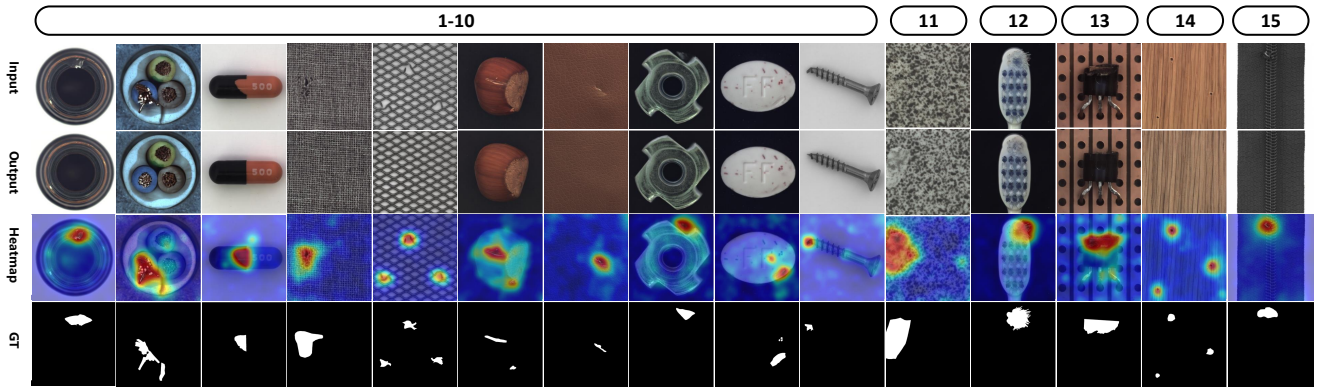

Figure 6. Qualitative comparison results under setting 4 of MVTec, the numbers represent continual training classes.

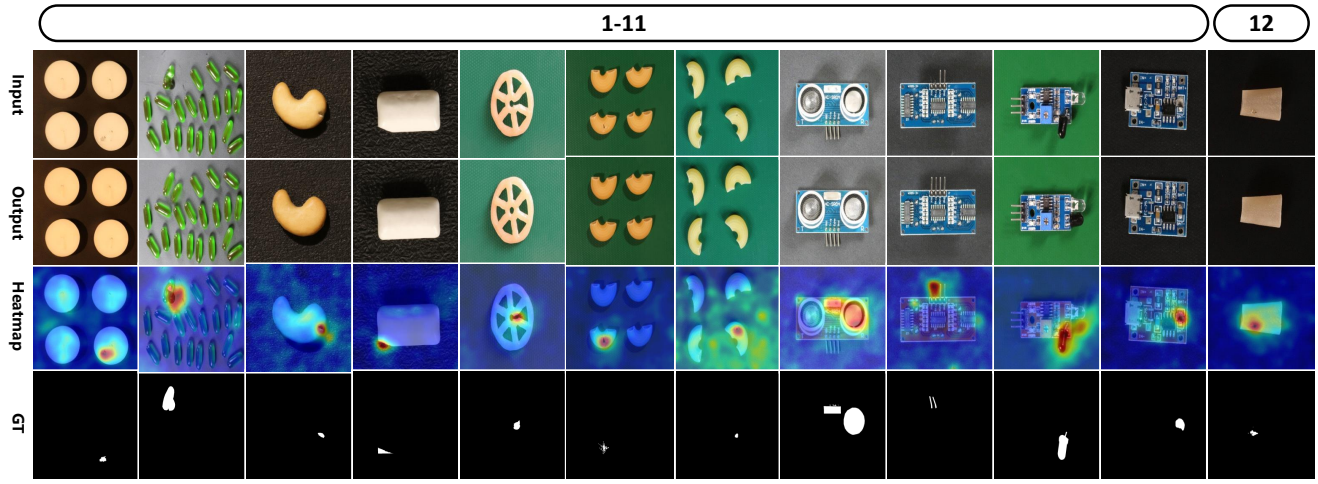

Figure 7. Qualitative comparison results under setting 5 of VisA, the numbers represent continual training classes.

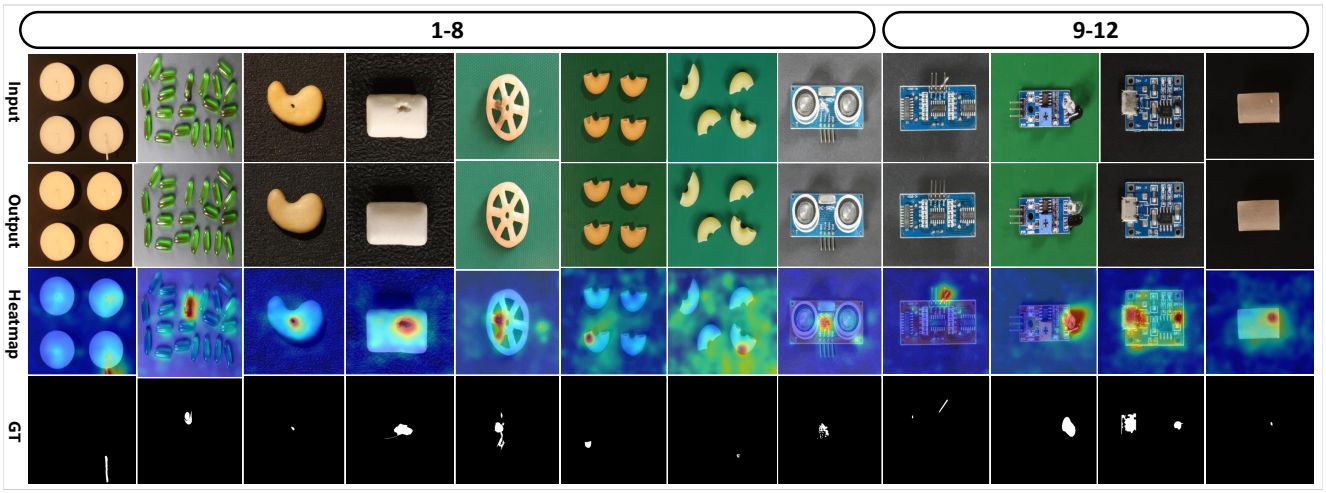

Figure 8. Qualitative comparison results under setting 6 of VisA, the numbers represent continual training classes.

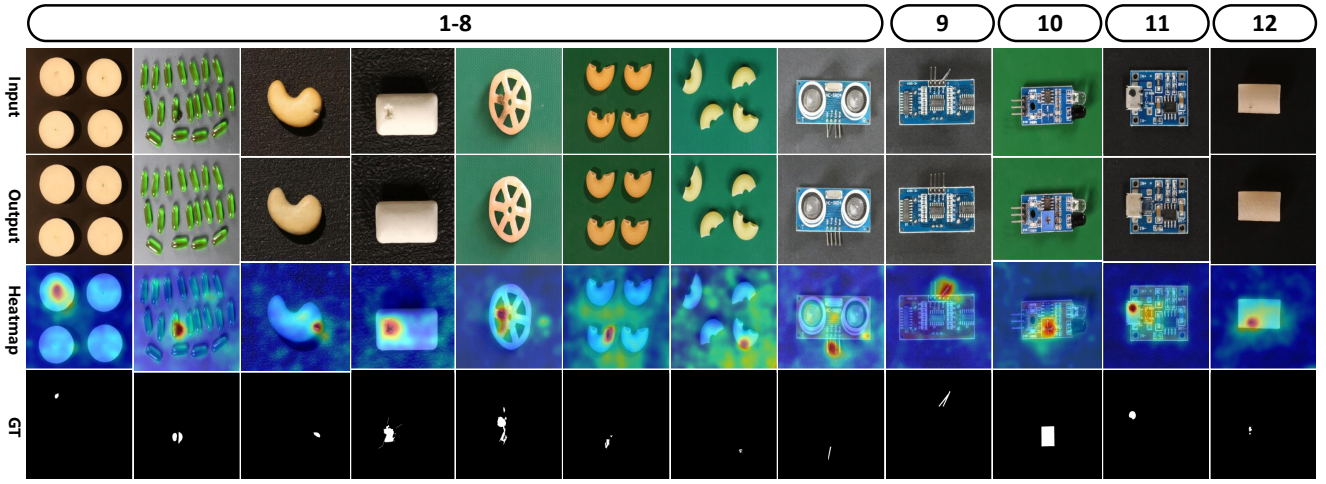

Figure 9. Qualitative comparison results under setting 7 of VisA, the numbers represent continual training classes.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 1

[2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4183–4192, 2020. 1

[3] 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8472–8480, 2024. 1, 2, 3

[4] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pages 9664–9674. Computer Vision Foundation / IEEE, 2021. 1

[5] Wujin Li, Jiawei Zhan, Jinbao Wang, Bizhong Xia, Bin-Bin Gao, Jun Liu, Chengjie Wang, and Feng Zheng. Towards continual adaptation in industrial anomaly detection. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2871–2880, 2022. 1

[6] Jiaqi Liu, Kai Wu, Qiang Nie, Ying Chen, Bin-Bin Gao, Yong Liu, Jinbao Wang, Chengjie Wang, and Feng Zheng. Unsupervised continual anomaly detection with contrastively-learned prompt. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3639–3647, 2024. 2, 3

[7] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *Seminal Graphics Papers: Pushing the*

Boundaries, Volume 2, pages 577–582. 2023. 1

[8] Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *Proceedings of the European conference on computer vision*, pages 474–489. Springer, 2022. 1

[9] Jiaqi Tang, Hao Lu, Xiaogang Xu, Ruizheng Wu, Sixing Hu, Tong Zhang, Tsz Wa Cheng, Ming Ge, Ying-Cong Chen, and Fugee Tsung. An incremental unified framework for small defect inspection. In *European Conference on Computer Vision*, pages 307–324. Springer, 2025. 2, 3

[10] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 2, 3

[11] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *Proceedings of the European conference on computer vision*, pages 392–408. Springer, 2022. 1