OmniAL: A unified CNN framework for unsupervised anomaly localization Supplementary Material

Ying Zhao
Ricoh Software Research Center (Beijing) Co., Ltd.
zy_deepwhite_zy@hotmail.com

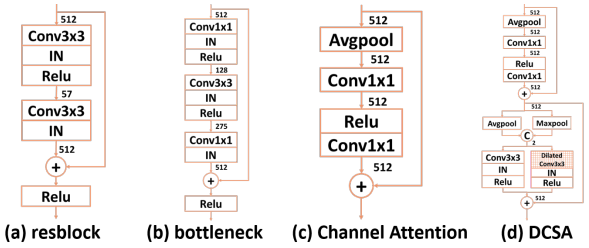

Figure 7. Modules comparison. To make them have similar numbers of parameters with our DCSA module, we adjust the feature map dimensions for the resblock and bottleneck.

Table 6. Ablation study on modules in Fig.7. Our DCSA module achieves the best performance. M-P: Module Parameters, Rec-P: Reconstruction sub-network Parameters, Loc-P: Localization sub-network Parameters, I-AUC: Image-level AUROC, P-AUC: Pixel-level AUROC, P-AP: Pixel-level AP.

Modules	M-P	Rec-P	Loc-P	I-AUC	P-AUC	P-AP
resblock	**527,019**	**63,906,674**	**30,852,340**	93.3	94.2	60.0
bottleneck	524,966	62,852,068	30,847,397	95.7	96.1	62.5
CA	524,288	62,851,972	28,700,493	96.8	98.2	72.7
DCSA	524,328	62,852,092	28,700,653	**97.2**	**98.3**	**73.4**

In this supplementary material, we present the additional details and results are not covered by the main paper.

1. Ablation study

To illustrate the efficiency of proposed DCSA module, we carry on experiments by replacing DCSA with the modules having similar number of parameters, as shown in Fig. 7. Table 6 presents that DCSA performs better than the others. Fig. 8 demonstrates that the relation between performance and the separate path balance factor used in DiffNeck. The network achieves the best performance by setting the balance factor as 0.1.

2. Adversarial anomaly samples

In general, the adversarial anomaly sample is defined as

$$A^{adv} = A + \delta \tag{1}$$

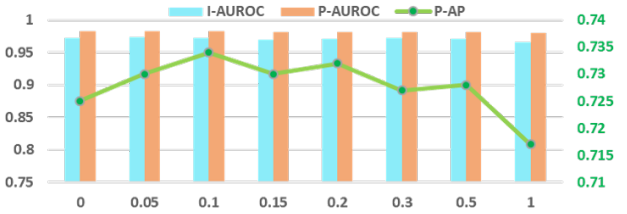

Figure 8. Ablation study on DiffNeck balance factor. Our method achieves the best performance with factor 0.1.

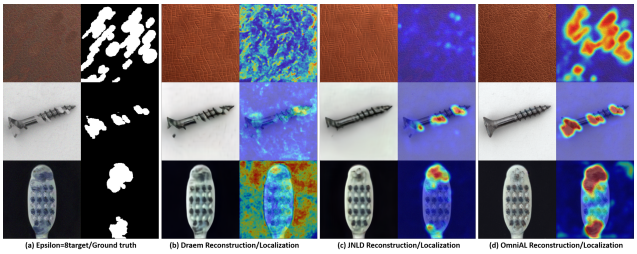

Figure 9. Robustness comparison with unified Draem [4] and JNLD [5] on adversarial samples.($\epsilon = 8target$)

Where A is the clean normal or synthesized anomaly image, δ is the perturbation and A^{adv} is the adversarial sample. Small amount of visually unperceivable perturbations δ is the fundamental premise for adversarial attacks. By injecting the perturbations δ to the synthesized anomaly images, we get the adversarial samples that are imperceptible to human visual system but deteriorate the anomaly localization methods. Since our JND-based synthesized anomaly are also visually unperceivable, the generated adversarial anomaly samples disguise to be well without arouse suspicion.

PGD [2] is a powerful optimization-based method and extensively used for evaluating the robustness in various computer vision tasks. Following [3], we generate the adversarial perturbations based on PGD [2]. We denote $f(A|\theta)$ as our anomaly localization model (Detailed structure are shown in Table 1.) with parameter θ, Y indicates the ground truth and D as the metric to measure the predic-

Table 7. Network for adversarial perturbations. PA:Panel anomaly synthesis, Rec:Reconstruction subnetwork, Seg:Segmentation subnetwork, BB:Basic Block, BN:Batch Norm, IN:Instance Norm, DC:Dilated Convolution, CA:Channel Attention, DSA:Dilated Spatial Attention, I:Image-level classification, P:Pixel-level localization on clean MVTecAD [1].

PA	Rec-BB			DCSA		Seg-BB		DiffNeck	I-AUROC	P-AUROC	P-AP
-	BN	IN	DC	CA	DSA	BN	DC	-	Unified/Separate		
-	√	-	-	√	-	√	-	-	86.7/98.8	86.4/98.4	44.2/75.0

tion degradation. To get suitable δ, the prediction degradation should be maximized. Thus, the objective of adversarial attacks is defined as

$$\delta = \underset{\delta}{\operatorname{argmax}}\, D(f(A+\delta|\theta), Y)$$

$$||\delta||_p = (\sum_i ||\delta_i||^p)^{\frac{1}{p}} \leq \epsilon \tag{2}$$

PGD [2] iteratively solves the maximization problem and get the perturbations δ.

$$\omega^{t+1} = \delta^t + \alpha sgn(\nabla_\delta D(f(A+\delta|\theta), Y)) \tag{3}$$

$$\delta^{t+1} = clip_{[-\epsilon,\epsilon]\cap[-A,1-A]}(\omega^{t+1}) \tag{4}$$

Where α is the adversarial step size, sgn is the sign function that extracts the sign of gradients for perturbation ∇_δ, ϵ indicates the maximum perturbation injected to each clean pixel. The clip operation also guarantees that the adversarial sample within [0,1]. The perturbation δ is initialized with the uniform distribution $Uniform(-\epsilon, \epsilon)$. For degradation measurement, we use the same objective functions as training, that is, MSE, SSIM, and focal loss. To fully evaluate the model robustness, we get the final perturbation for each adversarial sample after 5 iterations with $\epsilon = \{1/255, 2/255, 4/255, 8/255\}$ respectively. We also evaluate the robustness to the targeted attack $\epsilon = (8/25, targeted)$ by maximizing the probability of the normal-anomaly reversed segmentation ground truth. Fig.9 shows the examples of existing methods performance on targeted attack samples. Examples of different levels of adversarial anomaly samples for 15 classes of MVTecAD are shown in Fig.10. Table 7-11 illustrate that our method is more robust on anomaly localization and reconstruction robustness comparing with unified Draem [4] and JNLD [5] against 5-level adversarial samples.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec AD - A comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, pages 9592–9600, 2019. Computer Vision Foundation / IEEE. 2, 3

[2] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *ICLR, 2018*. OpenReview.net. 1, 2

[3] Yi Yu, Wenhan Yang, Yap-Peng Tan, and Alex C Kot. Towards robust rain removal against adversarial attacks: A comprehensive benchmark analysis and beyond. In *CVPR*, pages 6013–6022, 2022. 1

[4] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. Dræm - A discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pages 8310–8319, 2021. IEEE. 1, 2, 4, 5

[5] Ying Zhao. Just noticeable learning for unsupervised anomaly localization and detection. In *ICME*, pages In Press, 2022. 1, 2, 4, 5

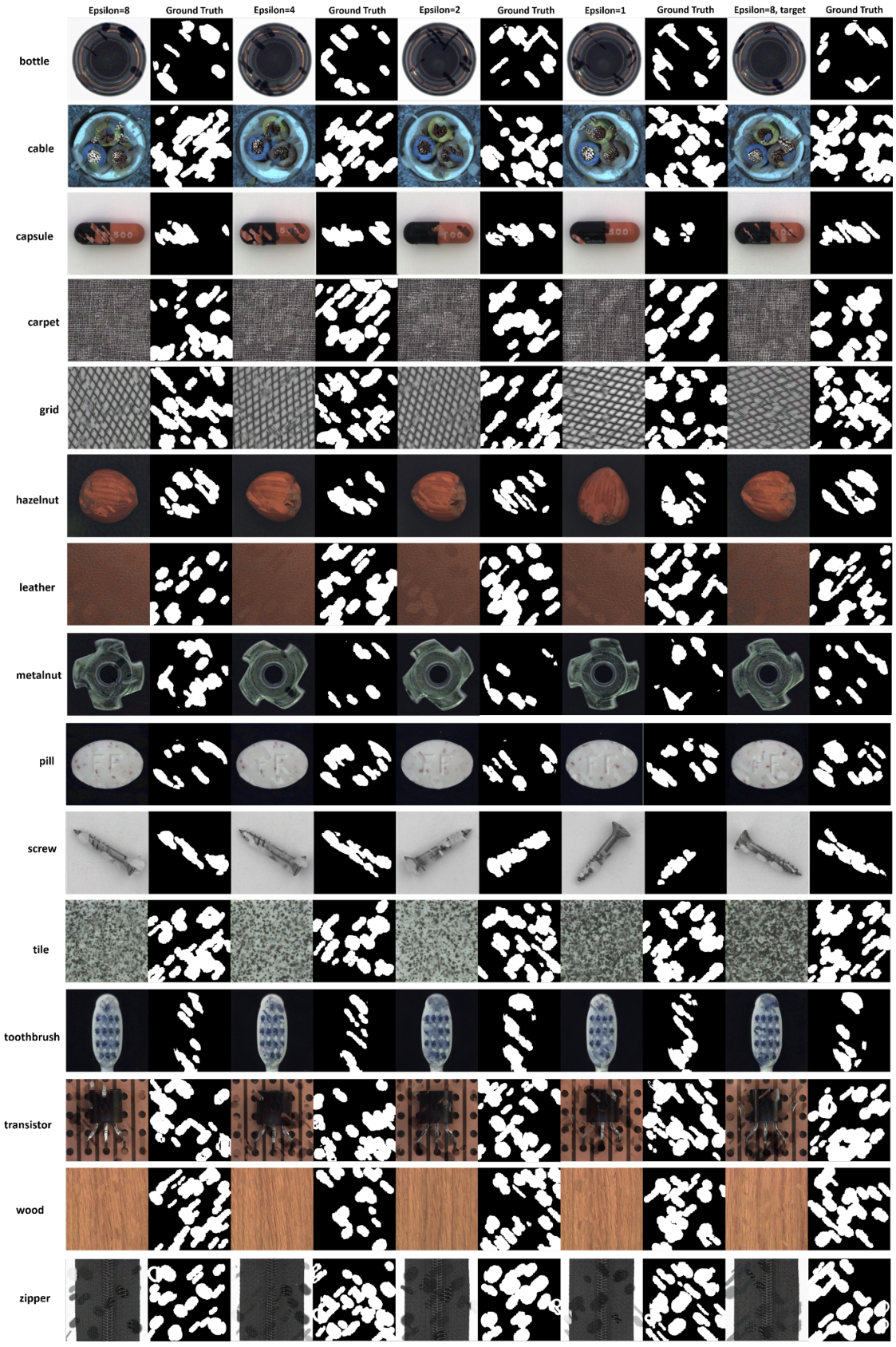

Figure 10. An overview of proposed 5-level adversarial samples($\epsilon = 1, 2, 4, 8, 8target$) of 15 categories of MVTecAD [1].

Table 8. Anomaly localization(**Pixel-AP**) robustness comparison with unified Draem [4] and JNLD [5] on adversarial samples.

Category	Epsilon=8t	Epsilon=8	Epsilon=4	Epsilon=2	Epsilon=1
	Draem / JNLD / OmniAL				
bottle	16.4 / **91.9** / 62.3	29.0 / **91.2** / 65.3	21.2 / **94.8** / 76.8	27.0 / **97.4** / 96.3	53.6 / 98.0 / **98.6**
cable	19.6 / 69.7 / **88.7**	32.9 / 62.4 / **78.3**	24.2 / 78.7 / **90.4**	40.3 / 94.0 / **97.4**	55.0 / 95.8 / **98.5**
capsule	23.7 / 79.2 / **86.1**	58.7 / 74.7 / **82.3**	18.5 / 87.7 / **91.8**	33.1 / 94.0 / **98.2**	33.8 / 96.8 / **98.6**
carpet	35.2 / 96.3 / **97.3**	41.3 / 93.5 / **90.8**	68.7 / **99.0** / 96.9	90.8 / **99.1** / **99.1**	95.8 / 99.1 / **99.3**
grid	54.5 / 98.3 / **98.8**	58.6 / 98.1 / **98.6**	84.6 / **99.3** / 99.0	91.0 / 98.8 / **99.1**	91.6 / 98.9 / **99.1**
hazelnut	5.4 / 13.8 / **20.0**	5.5 / 14.9 / **18.2**	6.0 / 29.0 / **80.8**	7.8 / 67.9 / **96.4**	9.4 / 84.8 / **97.1**
leather	15.6 / 22.8 / **74.1**	14.9 / 17.8 / **40.4**	16.9 / 65.2 / **91.6**	20.8 / 94.7 / **97.9**	40.4 / 97.2 / **98.3**
metal nut	6.9 / 27.8 / **35.7**	7.2 / 24.2 / **28.1**	7.5 / 58.7 / **63.5**	9.4 / **92.7** / 91.1	14.9 / 96.0 / **97.4**
pill	6.7 / **24.1** / 12.3	6.1 / **19.8** / 11.0	7.1 / **53.2** / 40.6	7.8 / 89.4 / **93.5**	13.7 / 95.1 / **97.9**
screw	37.6 / 44.6 / **97.0**	21.9 / 45.5 / **96.3**	29.7 / 88.3 / **98.0**	37.6 / 96.5 / **98.4**	37.6 / 97.6 / **98.6**
tile	20.4 / 24.1 / **42.8**	19.4 / 19.0 / **20.7**	24.9 / 31.0 / **39.7**	34.7 / **86.5**/ 81.8	44.7 / 97.2 / **97.8**
toothbrush	6.2 / 69.0 / **89.6**	6.1 / 70.5 / **88.3**	9.7 / 85.8 / **95.1**	15.7 / 92.8 / **97.3**	33.3 / 93.5 / **98.0**
transistor	27.7 / 84.6 / **96.0**	22.0 / 76.6 / **92.5**	28.0 / 92.0 / **97.4**	36.2 / 97.7 / **98.7**	55.4 / 98.3 / **98.9**
wood	17.4 / 25.8 / **36.5**	21.4 / 33.3 / **40.9**	25.1 / 54.2 / **73.1**	31.9 / 83.9 / **90.0**	36.3 / 91.2 / **94.1**
zipper	55.2 / 89.0 / **95.8**	57.8 / 90.5 / **93.4**	81.5 / 96.7 / **98.0**	91.2 / 97.6 / **98.8**	92.6 / **98.6** / 99.0
average	22.3 / 57.4 / **68.9**	26.9 / 55.5 / **63.0**	30.2 / 74.2 / **82.2**	38.4 / 92.2 / **95.6**	47.2 / 95.9 / **98.1**

Table 9. Anomaly detection(**Pixel-AUROC**) robustness comparison with unified Draem [4] and JNLD [5] on adversarial samples.

Category	Epsilon=8t	Epsilon=8	Epsilon=4	Epsilon=2	Epsilon=1
	Draem / JNLD / OmniAL				
bottle	63.9 / **97.4** / 90.5	79.9 / **97.6** / 92.8	71.7 / **98.5** / 95.4	75.8 / 99.3 / **99.4**	83.2 / 99.5 / **99.7**
cable	29.7 / 81.1 / **92.2**	66.3 / 80.5 / **88.8**	42.1 / 87.5 / **93.6**	57.1 / 96.3 / **98.4**	70.1 / 97.7 / **99.3**
capsule	74.2 / 89.6 / **97.9**	89.4 / 89.2 / **97.6**	72.2 / 96.0 / **98.8**	76.5 / 98.6 / **99.7**	75.5 / 99.4 / **99.9**
carpet	59.1 / 98.2 / **98.8**	69.6 / **97.1** / 95.5	86.8 / **99.6** / 98.7	95.9 / 99.6 / **99.7**	98.3 / 99.7 / **99.7**
grid	76.2 / 99.2 / **99.5**	78.4 / 99.1 / **99.4**	93.7 / **99.7** / 99.6	96.0 / 99.5 / **99.6**	96.7 / 99.6 / **99.7**
hazelnut	10.4 / **63.3** / 63.0	11.2 / 67.1 / **61.6**	19.8 / 75.8 / **96.0**	36.2 / 89.4 / **99.3**	38.8 / 96.3 / **99.5**
leather	19.4 / 47.2 / **85.1**	13.2 / 31.6 / **61.3**	22.1 / 79.4 / **95.0**	40.3 / 97.6 / **99.2**	64.3 / 98.9 / **99.4**
metal nut	16.5 / 72.8 / **78.0**	22.3 / 65.0 / **70.5**	25.1 / 87.8 / **91.6**	34.7 / 98.1 / **98.3**	56.4 / 99.0 / **99.5**
pill	19.8 / **72.7** / 58.6	20.0 / **67.8** / 48.0	27.8 / **83.4** / 77.9	36.8 / 97.6 / **98.7**	61.1 / 99.1 / **99.7**
screw	74.2 / 73.8 / **99.6**	73.3 / 73.3 / **99.5**	77.6 / 96.9 / **99.8**	79.7 / 99.5 / **99.8**	78.5 / 99.7 / **99.8**
tile	36.6 / 47.2 / **65.6**	32.1 / 31.0 / **38.1**	45.2 / 54.7 / **67.1**	58.9 / **93.5**/ 91.7	65.8 / 98.7 / **99.0**
toothbrush	24.0 / 93.1 / **98.1**	19.0 / 93.3 / **97.8**	47.2 / 96.3 / **99.3**	64.5 / 98.8 / **99.6**	83.9 / 99.1 / **99.8**
transistor	45.0 / 90.6 / **98.0**	39.2 / 86.0 / **96.2**	51.1 / 95.6 / **98.7**	63.5 / 99.0 / **99.5**	76.6 / 99.3 / **99.6**
wood	28.5 / 45.0 / **53.8**	40.4 / 55.4 / **59.1**	45.4 / 72.1 / **82.5**	52.9 / 91.6 / **94.9**	55.0 / 95.4 / **97.0**
zipper	76.0 / 93.2 / **98.0**	76.4 / 93.9 / **96.4**	89.0 / 98.0 / **99.0**	95.2 / 98.8/ **99.5**	96.1 / 99.4 / **99.6**
average	43.6 / 77.6 / **85.1**	48.7 / 75.2 / **80.2**	54.5 / 88.1 / **92.9**	64.3 / 97.1 / **98.5**	73.4 / 98.7 / **99.4**

Table 10. Anomaly reconstruction(**PSNR**) robustness comparison with unified Draem [4] and JNLD [5] on adversarial samples.

Category	Epsilon=8t	Epsilon=8	Epsilon=4	Epsilon=2	Epsilon=1
	Draem / JNLD / OmniAL				
bottle	30.6 / 26.2 / **31.1**	31.0 / 26.6 / **32.3**	31.2 / 26.5 / **32.1**	30.9 / 32.0 / **34.0**	30.8 / 32.0 / **34.6**
cable	27.3 / 23.8 / **29.2**	26.7 / 23.8 / **30.2**	28.5 / 24.4 / **30.8**	27.5 / 28.4 / **32.0**	26.9 / 27.8 / **32.1**
capsule	28.0 / 24.6 / **29.6**	28.4 / 25.0 / **31.3**	30.0 / 25.4 / **31.8**	29.6 / 30.4 / **33.2**	29.4 / 30.1 / **33.6**
carpet	26.1 / 23.1 / **28.6**	26.8 / 23.7 / **30.1**	27.4 / 23.7 / **30.3**	26.8 / 27.4 / **31.2**	26.4 / 26.9 / **31.4**
grid	26.3 / 23.2 / **28.9**	26.9 / 23.6 / **30.1**	27.7 / 23.9 / **30.7**	27.5 / 27.9 / **31.9**	27.3 / 27.7 / **32.2**
hazelnut	27.3 / 24.3 / **29.4**	27.5 / 24.4 / **30.3**	28.8 / 25.5 / **31.5**	28.8 / 29.4 / **32.9**	28.9 / 29.4 / **33.3**
leather	27.4 / 24.9 / **29.6**	27.5 / 24.8 / **30.3**	28.9 / 26.0 / **31.6**	28.9 / 29.5 / **33.0**	29.0 / 29.5 / **33.5**
metal nut	27.5 / 25.0 / **29.5**	27.5 / 24.9 / **30.0**	28.9 / 26.1 / **31.5**	28.9 / 29.5 / **32.9**	29.0 / 29.5 / **33.4**
pill	27.8 / 25.4 / **29.6**	27.8 / 25.3 / **30.0**	29.5 / 26.7 / **31.8**	29.8 / 30.3 / **33.5**	30.1 / 30.6 / **34.1**
screw	28.1 / 25.6 / **29.9**	27.9 / 25.3 / **30.2**	29.7 / 26.8 / **32.1**	30.1 / 30.7 / **33.9**	30.4 / 31.0 / **34.5**
tile	27.7 / 25.3 / **29.8**	27.6 / 25.1 / **30.0**	29.2 / 26.5 / **31.8**	29.5 / 30.1/ **33.5**	30.0 / 30.5 / **34.2**
toothbrush	27.7 / 25.3 / **29.8**	27.6 / 25.1 / **29.9**	29.2 / 26.5 / **31.8**	29.5 / 30.1 / **33.5**	29.9 / 30.4 / **34.2**
transistor	27.7 / 25.3 / **29.8**	27.8 / 25.1 / **29.9**	29.3 / 26.4 / **31.8**	29.6 / 30.1 / **33.5**	29.9 / 30.5 / **34.2**
wood	27.8 / 25.4 / **29.8**	27.7 / 25.3 / **29.9**	29.2 / 26.5 / **31.9**	29.5 / 30.0 / **33.5**	29.9 / 30.4 / **34.2**
zipper	27.8 / 25.3 / **29.9**	27.7 / 25.1 / **30.0**	29.2 / 26.3 / **31.9**	29.4 / 30.0/ **33.4**	29.8 / 30.4 / **34.1**
average	27.7 / 24.8 / **29.6**	27.8 / 24.9 / **30.3**	29.1 / 25.8 / **31.6**	29.1 / 29.7 / **33.1**	29.2 / 29.8 / **33.6**

Table 11. Anomaly reconstruction(**SSIM**) robustness comparison with unified Draem [4] and JNLD [5] on adversarial samples.

Category	Epsilon=8t	Epsilon=8	Epsilon=4	Epsilon=2	Epsilon=1
	Draem / JNLD / OmniAL				
bottle	0.921 / 0.879 / **0.932**	0.940 / 0.893 / **0.953**	0.940 / 0.900 / **0.952**	0.956 / 0.959 / **0.971**	0.961 / 0.965 / **0.977**
cable	0.875 / 0.823 / **0.921**	0.890 / 0.833 / **0.947**	0.911 / 0.856 / **0.950**	0.910 / 0.920 / **0.967**	0.906 / 0.918 / **0.972**
capsule	0.861 / 0.820 / **0.898**	0.896 / 0.846 / **0.940**	0.916 / 0.871 / **0.945**	0.925 / 0.932 / **0.966**	0.926 / 0.934 / **0.972**
carpet	0.841 / 0.782 / **0.907**	0.872/ 0.810 / **0.937**	0.879 / 0.821 / **0.944**	0.876 / 0.887 / **0.960**	0.872 / 0.884 / **0.964**
grid	0.860 / 0.798 / **0.916**	0.878 / 0.815 / **0.939**	0.896 / 0.833 / **0.951**	0.897 / 0.905 / **0.965**	0.896 / 0.905 / **0.970**
hazelnut	0.862 / 0.810 / **0.911**	0.875 / 0.821 / **0.930**	0.903 / 0.854 / **0.951**	0.909 / 0.916 / **0.967**	0.910 / 0.918 / **0.972**
leather	0.850 / 0.804 / **0.907**	0.861 / 0.813 / **0.923**	0.892 / 0.848 / **0.947**	0.895 / 0.903 / **0.963**	0.898 / 0.906 / **0.968**
metal nut	0.849 / 0.806 / **0.903**	0.858 / 0.812 / **0.917**	0.892 / 0.850 / **0.945**	0.895 / 0.903 / **0.961**	0.899 / 0.907 / **0.967**
pill	0.851 / 0.810 / **0.901**	0.857 / 0.813 / **0.911**	0.897 / 0.860 / **0.946**	0.905 / 0.912 / **0.963**	0.910 / 0.917 / **0.969**
screw	0.855 / 0.816 / **0.901**	0.859 / 0.818 / **0.910**	0.902 / 0.866 / **0.947**	0.911 / 0.918 / **0.965**	0.918 / 0.924 / **0.971**
tile	0.850 / 0.810 / **0.904**	0.854 / 0.811/ **0.909**	0.894 / 0.860 / **0.946**	0.900 /0.909 / **0.964**	0.908 / 0.916 / **0.971**
toothbrush	0.850 / 0.810 / **0.904**	0.854 / 0.811 / **0.909**	0.894 / 0.858 / **0.946**	0.901 / 0.909 / **0.964**	0.909 / 0.917 / **0.971**
transistor	0.852 / 0.812 / **0.904**	0.856/ 0.812 / **0.908**	0.897 / 0.859 / **0.947**	0.905 / 0.913 / **0.965**	0.912 / 0.919 / **0.971**
wood	0.850 / 0.812 / **0.903**	0.853 / 0.812 / **0.907**	0.894 / 0.858 / **0.945**	0.900 / 0.908 / **0.963**	0.909 / 0.916 / **0.970**
zipper	0.850 / 0.808 / **0.903**	0.853 / 0.808/ **0.907**	0.893 / 0.853/ **0.945**	0.897 / 0.908 / **0.962**	0.905 / 0.916 / **0.969**
average	0.858 / 0.813 / **0.908**	0.870 / 0.822 / **0.923**	0.900 / 0.856 / **0.947**	0.905 / 0.913 / **0.964**	0.909 / 0.917 / **0.970**