Dual Memory Networks Guided Reverse Distillation for Unsupervised Anomaly Detection

Chi Dai Tran[0000−0001−6345−8658], Long Hoang Pham[0000−0002−3240−657X], Duong Nguyen-Ngoc Tran[0000−0001−7537−6377], Quoc Pham-Nam Ho[0009−0006−7256−4798], and Jae Wook Jeon[0000−0003−0037−112X]

Department of Electrical and Computer Engineering,
Sungkyunkwan University, Suwon, South Korea
{tdc2000, phlong, duongtran, hpnquoc, jwjeon}@skku.edu

1 Experimental Results

In this section, we provide detailed anomaly detection and localization results for three standard benchmark datasets, namely MVTec [2], BTAD [12], and VisA [19]. Table 1 presents the AD results regarding I-AUC on the MVTec dataset. DM-GRD outperforms its memory bank counterparts, such as ReconFA, PaDiM, PatchCore, and CFA in most categories. These findings strengthen our hypothesis that relying on a single memory module can compromise model performance on challenging datasets. Despite using only pseudo-anomaly samples for the separation loss, DM-GRD achieves better results than supervised methods. For KD approaches, DM-GRD surpasses MKD, RD, and RD++ by 11.76%, 1.53%, and 0.06%, respectively. Compared to MemKD, DM-GRD delivers competitive results, while our dual memory networks are designed in an efficient manner, reducing the inference time to less than 318 milliseconds.

Table 2 and Table 3 depict the anomaly localization performance of DM-GRD and other advanced models on the MVTec dataset in P-AUC and PRO metrics. In terms of P-AUC, our technique outperforms all other models except for PRNet, which is trained in a supervised setting. For the PRO criterion, DM-GRD surpasses its competitors by a significant margin while keeping an inference time adequate for real-time anomaly detection in the industrial sector.

Table 4, Table 5, and Table 6 reflect the anomaly detection and localization results of DM-GRD and other cutting-edge approaches on the BTAD dataset. According to our observations, most approaches produce competitive I-AUC and P-AUC values. However, since this is a small dataset with relatively small surface abnormalities, most methods perform poorly on the PRO measure. Meanwhile, with two memory modules to retain crucial areas on the surface of objects, DM-GRD has little difficulty accurately segmenting small aberrant regions.

Table 7 summarizes the anomaly detection and localization performance of memory bank and KD methods in terms of I-AUC and PRO. Except for SPADE and PaDiM, most approaches performed well on I-AUC. Nonetheless, most memory bank models scored poorly on the PRO criterion. Similarly, RD struggled to segment anomalies on a large and demanding dataset like VisA due to not being trained on bogus anomaly samples. These observations further corroborate

Table 1: Anomaly detection results in terms of I-AUC on the MVTec dataset. The best results are bolded, and the second-best are underlined.

	Memory Bank				Supervised			Knowledge Distillation				
Category	ReconFA [20]	PaDiM [4]	PatchCore [14]	CFA [8]	DevNet [13]	DRA [6]	PRNet [18]	RD [5]	RD++ [16]	MKD [15]	MemKD [7]	**Ours**
Carpet	98.10	99.80	98.70	97.30	82.50	92.50	99.70	98.90	**100.0**	79.25	99.60	**100.0**
Grid	98.70	96.70	98.20	99.20	90.60	98.60	99.40	**100.0**	**100.0**	78.01	100.0	**100.0**
Leather	100.0	100.0	**100.0**	**100.0**	92.20	98.90	**100.0**	**100.0**	**100.0**	95.05	100.0	**100.0**
Tile	98.90	98.10	98.70	99.40	99.90	**100.0**	**100.0**	99.30	99.70	91.57	100.0	99.70
Wood	99.90	99.20	99.20	99.70	97.90	99.10	**100.0**	99.20	99.30	94.29	99.50	99.91
Bottle	100.0	99.90	**100.0**	**100.0**	99.70	**100.0**	**100.0**	**100.0**	**100.0**	99.39	100	**100.0**
Cable	98.10	92.70	99.50	**99.80**	98.70	94.20	98.90	95.00	99.20	89.19	99.20	99.68
Capsule	93.90	91.30	98.10	97.30	71.90	95.10	98.00	96.30	**99.00**	80.46	98.80	97.30
Hazelnut	98.00	92.00	**100.0**	**100.0**	99.70	**100.0**	**100.0**	99.90	**100.0**	98.37	100.0	**100.0**
Metal nut	99.60	98.70	**100.0**	**100.0**	98.80	99.10	**100.0**	**100.0**	**100.0**	73.58	100.0	**100.0**
Pill	98.70	93.30	96.60	97.90	87.10	88.30	**99.30**	96.60	98.40	82.70	98.30	97.46
Screw	94.50	85.80	98.10	97.30	97.20	**99.50**	95.90	97.00	98.90	83.31	99.10	98.54
Toothbrush	100.0	96.10	**100.0**	**100.0**	79.20	87.50	**100.0**	99.50	**100.0**	92.17	100.0	**100.0**
Transistor	100.0	97.40	**100.0**	**100.0**	89.10	88.30	99.70	96.70	98.50	85.55	100.0	99.96
Zipper	96.40	90.30	99.40	99.60	99.10	99.70	99.70	98.50	98.60	93.24	99.30	**99.89**
Average	98.30	95.50	99.10	99.17	92.20	96.10	99.40	98.46	99.44	87.74	99.60	**99.50**

Table 2: Anomaly localization performance in terms of P-AUC on MVTec AD.

	Memory Bank				Supervised			Knowledge Distillation				
Category	SPADE [3]	PaDiM [4]	PatchCore [14]	CFA [8]	DevNet [13]	DRA [6]	PRNet [18]	RD [5]	RD++ [16]	MKD [15]	MemKD [7]	**Ours**
Carpet	96.50	99.10	99.00	99.28	97.20	98.20	99.00	98.90	99.20	95.64	99.10	**99.31**
Grid	92.70	97.30	98.70	98.12	87.90	86.00	98.40	99.30	99.30	91.78	99.20	**99.32**
Leather	96.60	99.20	99.30	99.37	94.20	93.80	**99.70**	99.40	99.40	98.05	99.50	99.62
Tile	86.40	94.10	95.40	95.25	92.70	92.30	**99.60**	95.60	96.60	82.77	95.70	97.13
Wood	87.50	94.90	95.00	91.53	86.40	82.90	**97.80**	95.30	95.80	84.80	95.30	96.68
Bottle	97.40	98.30	98.60	98.84	93.90	91.30	**99.40**	98.70	98.80	96.32	98.80	98.84
Cable	96.20	96.70	98.40	**98.97**	88.80	86.60	98.80	97.40	98.40	82.40	98.30	98.15
Capsule	98.00	98.50	98.80	**99.11**	91.80	89.30	98.50	98.70	98.80	95.86	98.80	98.85
Hazelnut	98.10	98.20	98.70	98.85	91.10	89.60	**99.70**	98.90	99.20	94.62	99.10	99.30
Metal nut	97.10	97.20	98.40	99.15	77.80	79.50	**99.70**	97.30	98.10	86.38	97.20	97.14
Pill	95.50	95.70	97.40	98.93	82.60	84.50	**99.50**	98.20	98.30	89.63	98.30	98.51
Screw	97.90	98.50	99.40	98.91	60.30	54.00	97.50	99.60	**99.70**	95.96	99.60	99.50
Toothbrush	96.90	98.80	98.70	98.96	84.60	75.50	**99.60**	99.10	99.10	96.12	98.90	99.40
Transistor	93.10	97.50	96.30	98.06	56.00	79.10	**98.40**	92.50	94.30	76.45	96.40	96.30
Zipper	95.50	98.50	98.80	**99.02**	93.70	96.90	98.80	98.20	98.80	93.90	98.50	98.03
Average	95.50	97.50	98.06	98.15	85.30	85.30	**99.00**	97.81	98.25	90.71	98.20	98.41

Table 3: Anomaly localization results in terms of PRO on the MVTec dataset.

	Memory Bank				Supervised			Knowledge Distillation				
Category	SPADE [3]	PaDiM [4]	PatchCore [14]	CFA [8]	DevNet [13]	DRA [6]	PRNet [18]	RD [5]	RD++ [16]	MKD [15]	MemKD [7]	**Ours**
Carpet	93.70	96.20	96.60	96.54	85.80	92.20	97.00	97.00	97.70	92.50	97.50	**99.22**
Grid	85.70	94.60	96.00	94.04	79.80	71.50	95.90	97.60	97.70	72.90	96.90	**99.29**
Leather	96.20	97.80	98.90	97.43	88.50	84.00	99.20	99.10	99.20	97.50	99.20	**99.64**
Tile	74.90	86.00	87.30	89.26	78.90	81.50	**98.20**	90.60	92.40	74.30	91.10	97.81
Wood	86.40	91.10	89.40	90.54	75.40	69.70	95.90	90.90	93.30	76.50	91.20	**97.98**
Bottle	91.50	94.80	96.20	95.76	83.50	77.6	97.00	96.60	97.00	88.60	97.10	**98.97**
Cable	89.90	88.80	92.50	94.17	80.90	77.70	97.20	91.00	93.90	66.20	93.40	**98.04**
Capsule	92.70	93.50	95.50	93.66	83.60	79.10	92.50	95.80	96.40	90.10	96.20	**98.49**
Hazelnut	91.40	92.60	93.80	95.75	83.60	79.10	92.50	95.50	96.30	94.30	95.70	**99.11**
Metal nut	91.40	85.60	91.40	94.54	76.90	76.70	95.80	92.30	93.00	76.90	90.80	**97.98**
Pill	89.60	92.70	93.20	97.19	69.20	77.00	97.20	96.4	97.00	86.40	96.60	**98.84**
Screw	90.00	94.40	97.90	95.23	31.10	30.10	92.40	98.20	98.60	85.20	98.20	**99.49**
Toothbrush	92.50	93.10	91.50	91.14	33.50	56.10	95.60	94.50	94.20	87.30	92.20	**98.45**
Transistor	86.40	84.50	83.70	95.35	39.10	49.00	94.80	78.0	81.80	68.10	85.30	**97.56**
Zipper	91.60	95.90	97.10	95.95	81.30	91.00	95.50	95.40	96.30	86.50	95.90	**98.11**
Average	90.70	92.10	93.40	94.44	71.40	73.30	96.10	93.93	94.99	82.90	94.10	**98.60**

our prediction that training primarily on anomalous samples leads to "over-generalization", making it harder to handle anomalous samples effectively.

Table 4: Anomaly detection results on the BTAD dataset at I-AUC.

	Memory Bank					Supervised		Knowledge Distillation		
Class	PaDiM [4]	SPADE [3]	PatchCore [14]	CFA [8]	REB [11]	DRA [6]	PRNet [18]	RD [5]	RD++ [16]	**DM-GRD**
01	99.80	91.40	96.70	98.10	**99.60**	-	100.0	96.30	96.80	96.30
02	82.00	71.40	81.38	85.50	88.50	-	84.10	86.60	**90.10**	88.47
03	99.40	99.90	99.95	99.00	99.80	-	99.90	**100.0**	**100.0**	**100.0**
Average	93.70	87.60	92.68	94.20	**96.00**	94.20	94.70	94.30	95.63	94.92

Table 5: Anomaly localization results on the BTAD dataset at P-AUC.

	Memory Bank						Supervised			Knowledge Distillation		
Class	PNI [1]	PaDiM [4]	SPADE [3]	PatchCore [14]	CFA [8]	REB [11]	DRA [6]	BGAD [17]	PRNet [18]	RD [5]	RD++ [16]	**DM-GRD**
01	97.40	97.30	97.00	97.03	95.90	94.70	-	**98.20**	96.60	96.60	96.20	96.73
02	97.00	94.40	96.00	95.83	96.00	95.60	-	**97.90**	95.10	96.70	96.40	97.18
03	99.00	99.10	98.80	99.19	98.60	99.70	-	99.80	99.60	99.70	99.70	**100.0**
Average	97.80	96.90	97.30	97.35	96.83	97.20	75.40	**98.60**	97.10	97.67	97.43	97.97

Table 6: Anomaly localization results on the BTAD dataset at PRO.

	Memory Bank		Supervised				Knowledge Distillation		
Class	PatchCore [14]	CFA [8]	SemiREST [9]	DRA [6]	BGAD [17]	PRNet [18]	RD [5]	RD++ [16]	**DM-GRD**
01	64.92	72.00	93.10	-	83.00	81.40	75.30	73.20	**91.26**
02	47.27	53.20	81.40	-	64.80	54.40	68.20	71.30	**84.90**
03	67.72	94.10	99.40	-	99.30	98.30	87.80	87.40	**99.44**
Average	59.97	73.10	91.30	56.20	82.40	78.00	77.10	77.30	**91.71**

Table 7: AD and localization results on the VisA dataset at I-AUC/PRO.

	Memory Bank				Knowledge Distillation			
Category	DMAD [10]	PatchCore [14]	SPADE [3]	PaDiM [4]	RD [5]	RD++ [16]	MemKD [7]	**DM-GRD**
Candle	92.70/90.60	**98.60**/94.00	91.00/93.20	91.60/**95.70**	92.20/92.20	96.40/93.80	95.90/93.80	95.00/94.91
Capsules	88.00/88.40	81.60/85.50	61.40/36.10	70.70/76.90	90.10/56.90	92.10/95.80	94.70/88.20	**92.47**/**98.43**
Cashew	95.00/88.80	97.30/**94.50**	97.80/57.40	93.00/87.90	**99.60**/79.00	97.80/91.20	99.40/97.50	95.36/92.40
Chewing gum	97.40/73.90	99.10/84.60	85.80/93.90	98.80/83.50	**99.70**/92.50	96.40/88.10	99.80/98.80	96.70/**99.60**
Fryum	98.00/92.20	96.20/85.30	88.60/91.30	88.60/80.20	96.60/81.00	95.80/90.00	98.80/96.60	**97.16**/**95.31**
Macaroni1	94.30/97.10	97.50/95.40	95.20/61.30	87.00/92.10	98.40/71.30	94.00/**96.90**	98.00/92.70	**98.50**/95.74
Macaroni2	90.40/98.50	78.10/94.40	87.90/63.40	70.50/75.40	**97.60**/68.00	88.00/97.70	92.00/84.80	**97.60**/**98.43**
PCB1	95.80/96.20	**98.50**/94.30	72.10/38.40	94.70/91.30	97.60/43.20	97.00/95.80	96.90/96.90	93.87/**96.97**
PCB2	96.90/89.30	**97.30**/89.20	50.70/42.20	88.50/88.70	91.10/46.40	97.20/90.60	98.00/94.90	92.99/**91.37**
PCB3	98.30/93.60	**97.90**/90.90	90.50/80.30	91.00/84.90	95.50/80.30	96.80/93.10	97.80/96.60	96.44/**97.42**
PCB4	99.70/91.40	99.60/90.10	83.10/71.60	97.50/81.60	96.50/72.20	**99.80**/91.90	99.80/99.90	99.34/**97.55**
Pipe fryum	99.00/95.30	**99.80**/95.70	81.10/61.70	97.00/92.50	97.00/68.30	99.60/95.60	100.0/99.00	97.80/**98.26**
Average	95.50/91.30	95.10/91.20	82.10/65.90	89.10/85.90	96.00/70.90	95.90/93.40	97.60/94.90	**96.10**/**96.37**

Table 8 provides the anomaly detection and localization outcomes on several ResNet backbones on the MVTec dataset. In terms of P-AUC and PRO, DM-GRD beats its KD opponents across all backbones, demonstrating the generalizability of our method to pinpoint anomaly regions of varying sizes. Notably, most KD techniques perform comparably on smaller backbones, highlighting their suitability for real-time industrial applications.

Table 8: Study on different backbones on the MVTec dataset.

	I-AUC				P-AUC				PRO			
Backbone	MemKD	RD	RD++	**DM-GRD**	MemKD	RD	RD++	**DM-GRD**	MemKD	RD	RD++	**DM-GRD**
ResNet18	98.60	97.90	**98.63**	98.40	97.30	97.10	97.64	**97.82**	92.40	91.20	93.65	**97.65**
ResNet50	99.00	98.40	**99.05**	98.61	97.90	97.70	98.17	**98.21**	93.80	93.10	94.78	**98.20**
WResNet50	99.60	98.46	99.44	**99.50**	98.20	97.81	98.25	**98.41**	94.50	93.93	94.99	**98.60**

2 Anomaly Detection and Localization Visualization

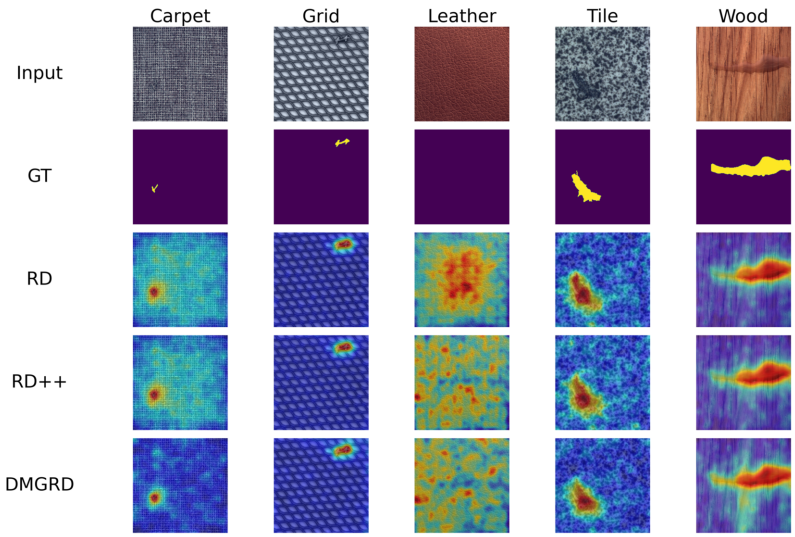

Fig. 1: Visualization of KD methods for anomaly localization in MVTec AD.

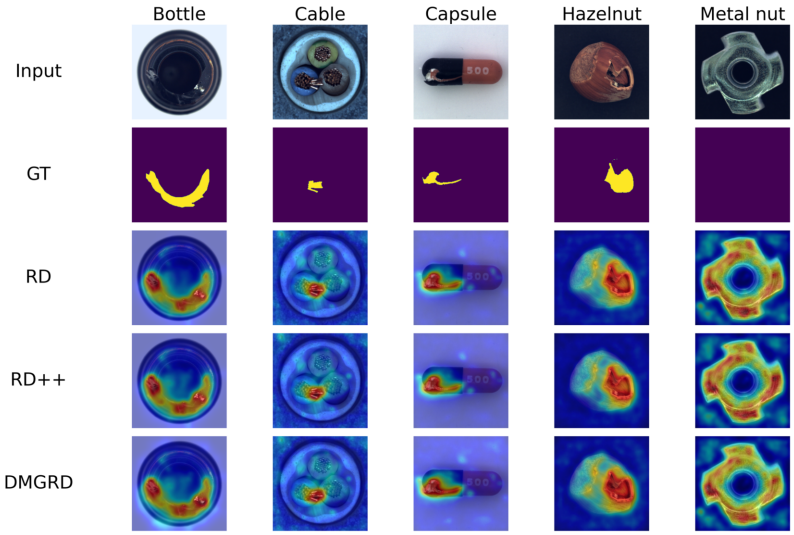

Fig. 2: Visualization of KD methods for anomaly localization in the MVTec dataset.

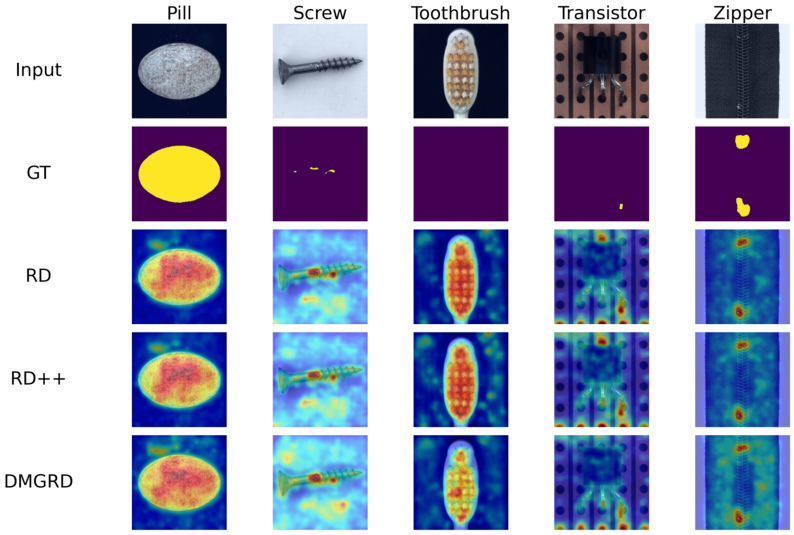

Fig. 3: Visualization of KD methods for anomaly localization in the MVTec dataset.

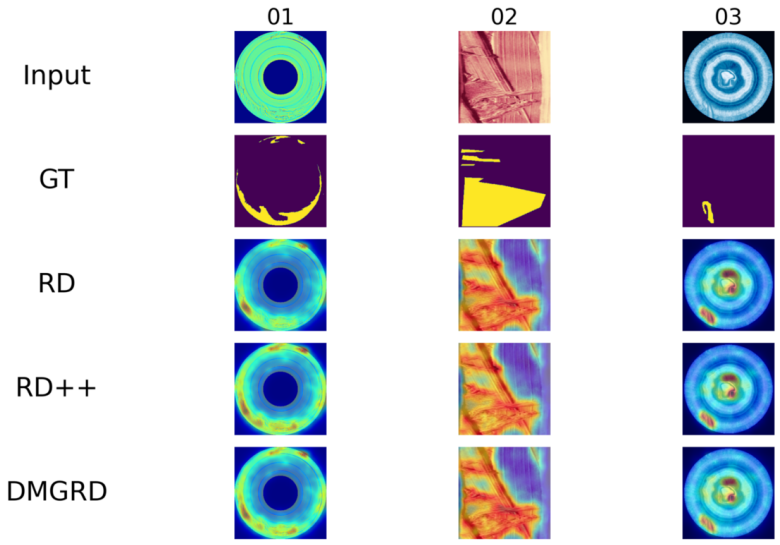

Fig. 4: Visualization of KD methods for anomaly localization in the BTAD dataset.

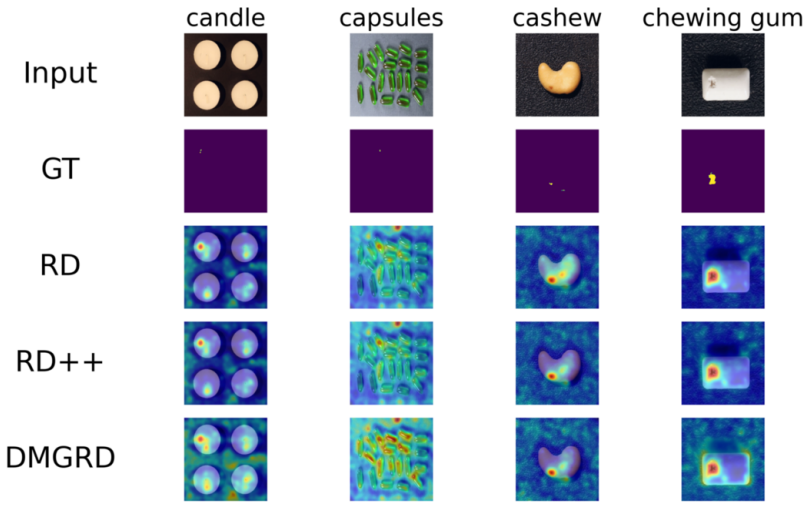

Fig. 5: Visualization of KD methods for anomaly localization in the VisA dataset.

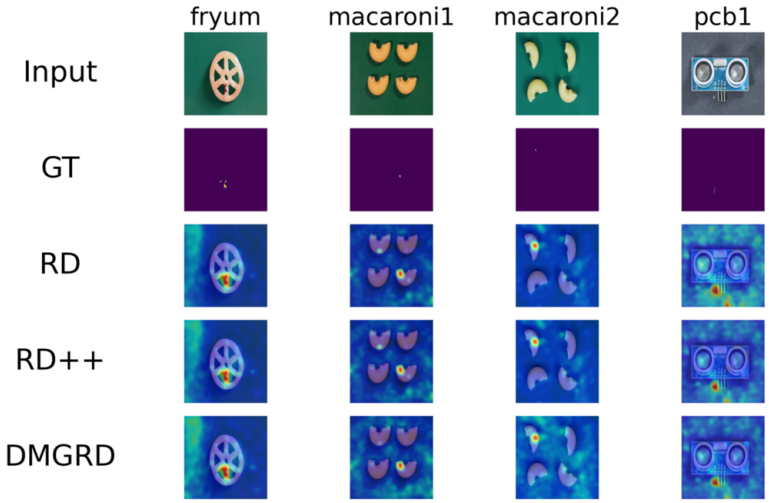

Fig. 6: Visualization of KD methods for anomaly localization in the VisA dataset.

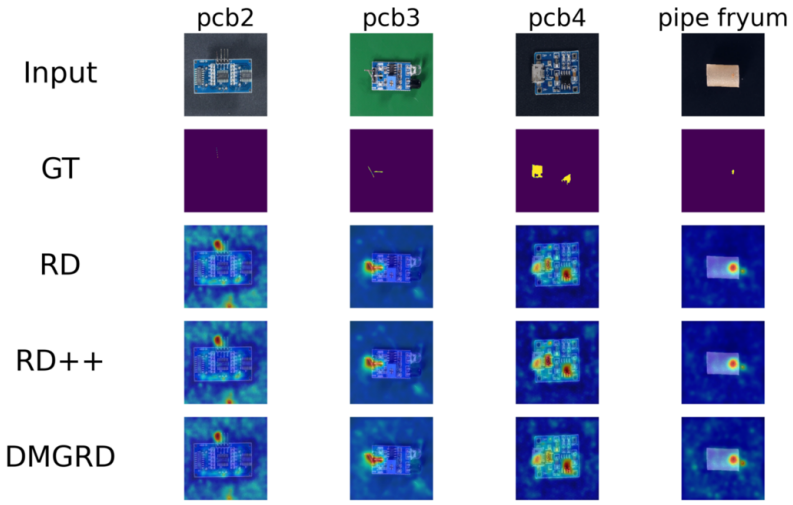

Fig. 7: Visualization of KD methods for anomaly localization in the VisA dataset.

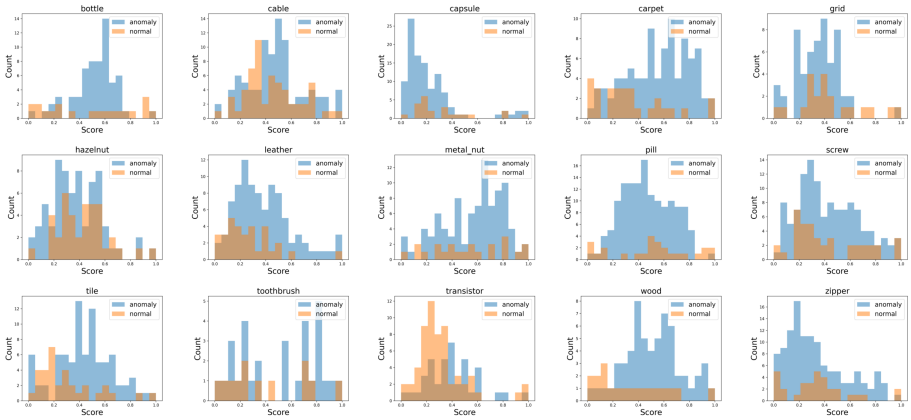

Fig. 8: Histogram of normal and anomaly scores for all categories in MVTec AD.

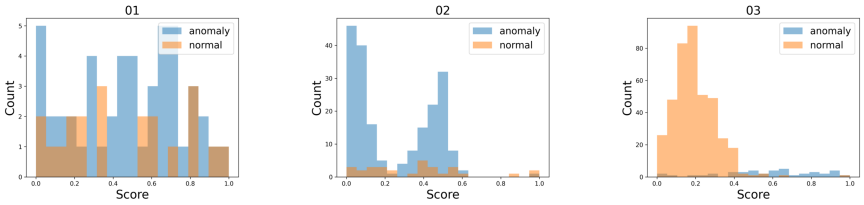

Fig. 9: Histogram of normal and anomaly scores for all classes in the BTAD dataset.

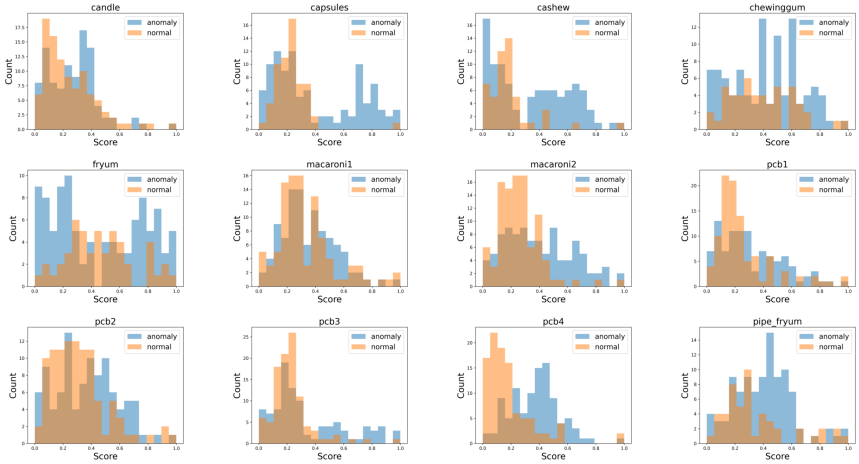

Fig. 10: Histogram of normal and anomaly scores for all classes in the VisA dataset.

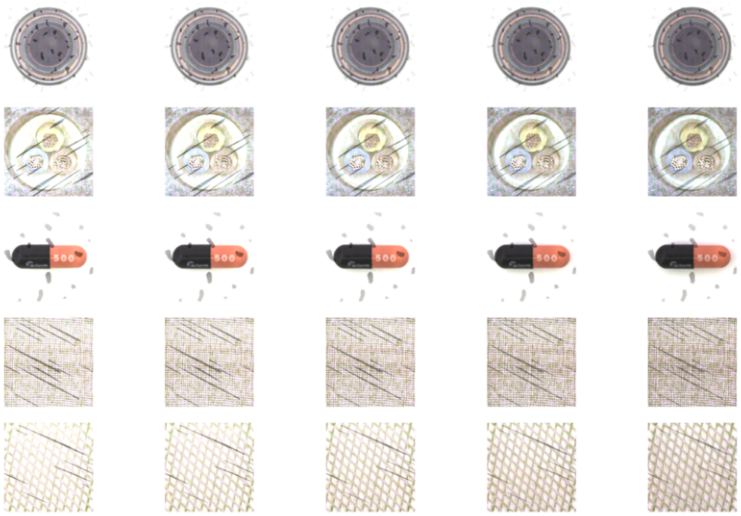

Fig. 11: Simulated anomalous regions MVTec AD with $\beta \in [0.1, 0.2, 0.3, 0.4, 0.5]$. From left to right: bottle, cable, capsule, carpet, and grid.

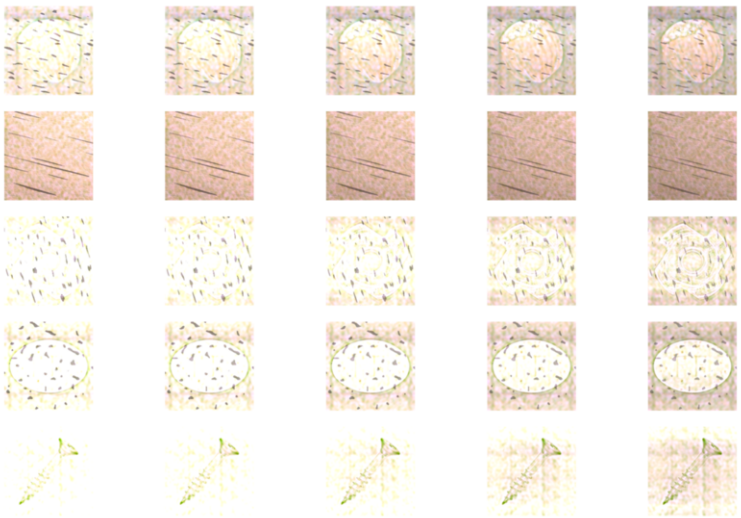

Fig. 12: Simulated anomalous regions in MVTec AD with $\beta \in [0.1, 0.2, 0.3, 0.4, 0.5]$. From left to right: hazelnut, leather, metal nut, pill, and screw.

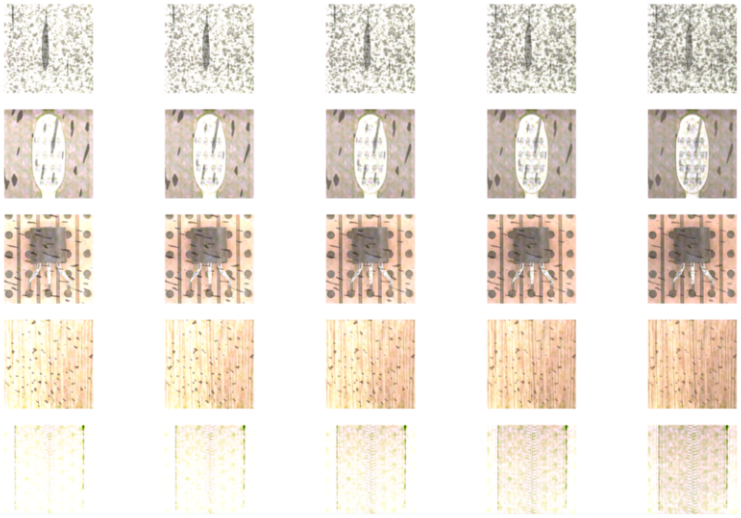

Fig. 13: Simulated anomalous regions in MVTec AD with $\beta \in [0.1, 0.2, 0.3, 0.4, 0.5]$. From left to right: tile, toothbrush, transistor, wood, and zipper.

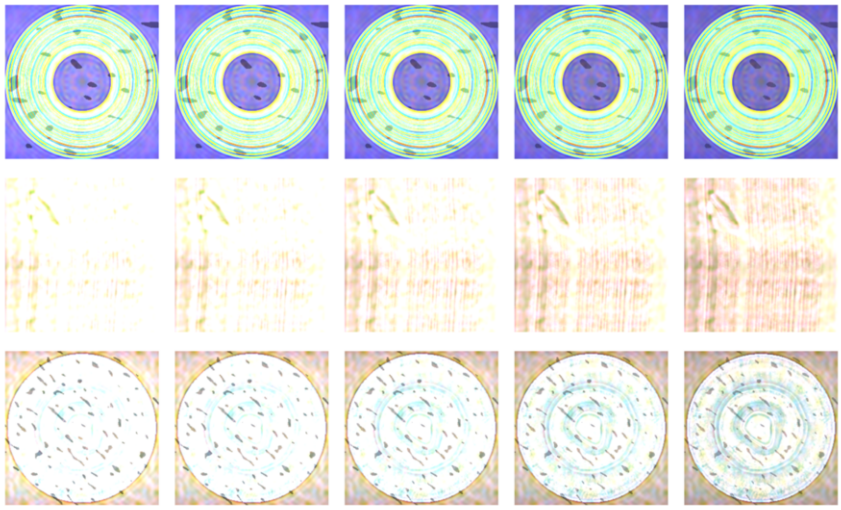

Fig. 14: Simulated anomalous regions in the BTAD dataset with $\beta \in [0.1, 0.2, 0.3, 0.4, 0.5]$. From left to right: 01, 02, and 03 classes.

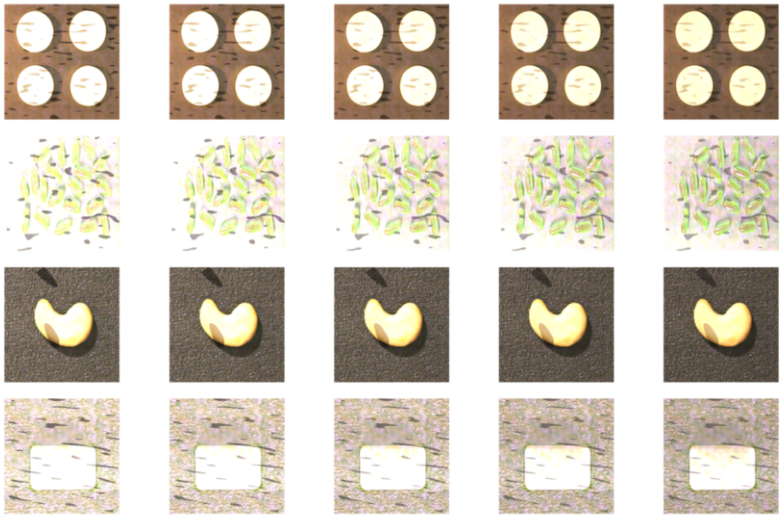

Fig. 15: Simulated anomalous regions in the VisA dataset with $\beta \in [0.1, 0.2, 0.3, 0.4, 0.5]$. From left to right: candle, capsules, cashew, and chewing gum.

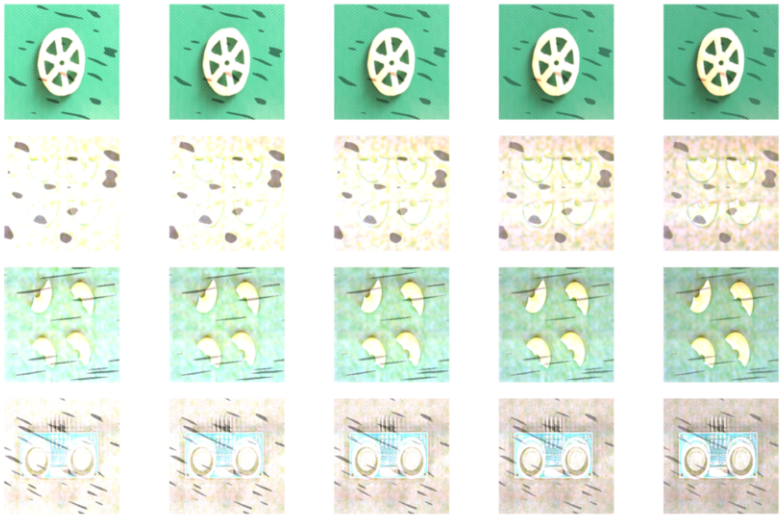

Fig. 16: Simulated anomalous regions in the VisA dataset with $\beta \in [0.1, 0.2, 0.3, 0.4, 0.5]$. From left to right: fryum, macaroni1, macaroni2, and pcb1.

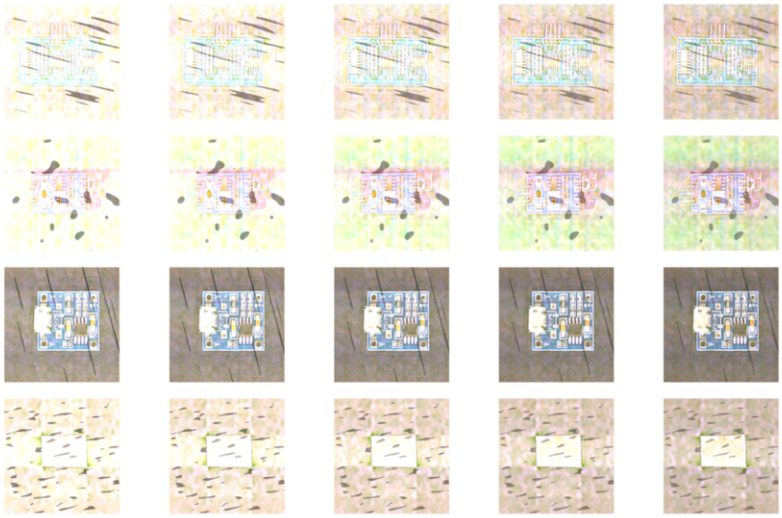

Fig. 17: Simulated anomalous regions in the VisA dataset with $\beta \in [0.1, 0.2, 0.3, 0.4, 0.5]$. From left to right:pcb2, pcb3, pcb4, and pipe fryum.

References

1. Bae, J., Lee, J.H., Kim, S.: Pni: Industrial anomaly detection using position and neighborhood information. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6373–6383 (2023)
2. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9592–9600 (2019)
3. Cohen, N., Hoshen, Y.: Sub-image anomaly detection with deep pyramid correspondences. arXiv preprint arXiv:2005.02357 (2020)
4. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: International Conference on Pattern Recognition. pp. 475–489. Springer (2021)
5. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9737–9746 (2022)
6. Ding, C., Pang, G., Shen, C.: Catching both gray and black swans: Open-set supervised anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7388–7398 (2022)
7. Gu, Z., Liu, L., Chen, X., Yi, R., Zhang, J., Wang, Y., Wang, C., Shu, A., Jiang, G., Ma, L.: Remembering normality: Memory-guided knowledge distillation for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16401–16409 (2023)
8. Lee, S., Lee, S., Song, B.C.: Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. IEEE Access **10**, 78446–78454 (2022)
9. Li, H., Wu, J., Chen, H., Wang, M., Shen, C.: Efficient anomaly detection with budget annotation using semi-supervised residual transformer. arXiv preprint arXiv:2306.03492 (2023)
10. Liu, W., Chang, H., Ma, B., Shan, S., Chen, X.: Diversity-measurable anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12147–12156 (2023)
11. Lyu, S., Mo, D., keung Wong, W.: Reb: Reducing biases in representation for industrial anomaly detection. Knowledge-Based Systems p. 111563 (2024)
12. Mishra, P., Verk, R., Fornasier, D., Piciarelli, C., Foresti, G.L.: Vt-adl: A vision transformer network for image anomaly detection and localization. In: 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE). pp. 01–06. IEEE (2021)
13. Pang, G., Ding, C., Shen, C., Hengel, A.: Explainable deep few-shot anomaly detection with deviation networks. arxiv 2021. arXiv preprint arXiv:2108.00462
14. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14318–14328 (2022)
15. Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M.H., Rabiee, H.R.: Multiresolution knowledge distillation for anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14902–14912 (2021)
16. Tien, T.D., Nguyen, A.T., Tran, N.H., Huy, T.D., Duong, S., Nguyen, C.D.T., Truong, S.Q.: Revisiting reverse distillation for anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 24511–24520 (2023)

17. Yao, X., Li, R., Zhang, J., Sun, J., Zhang, C.: Explicit boundary guided semi-push-pull contrastive learning for supervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 24490–24499 (2023)
18. Zhang, H., Wu, Z., Wang, Z., Chen, Z., Jiang, Y.G.: Prototypical residual networks for anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16281–16291 (2023)
19. Zou, Y., Jeong, J., Pemula, L., Zhang, D., Dabeer, O.: Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In: European Conference on Computer Vision. pp. 392–408. Springer (2022)
20. Zuo, Z., Wu, Z., Chen, B., Zhong, X.: A reconstruction-based feature adaptation for anomaly detection with self-supervised multi-scale aggregation. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5840–5844. IEEE (2024)