Remembering Normality: Memory-guided Knowledge Distillation for Unsupervised Anomaly Detection – Supplementary Materials

First Author
Institution1
Institution1 address
firstauthor@i1.org

Second Author
Institution2
First line of institution2 address
secondauthor@i2.org

1. Full Anomaly Detection and Localization Results on MVTec AD, VisA, and MPDD Datasets.

This section provides more detailed results for all sub-categories on MVTec AD [10], VisA [16], and MPDD [7] datasets. In MVTec AD, we report image-level I-AUC for anomaly detection in Tab. 1 and pixel-level P-AUC and P-PRO for anomaly localization in Tab. 2. It is observed that the proposed MemKD outperforms its SOTA competitors, *i.e.*, memory-bank-based PatchCore [11] and knowledge-distillation-based RD [5] on the majority of categories and performs best in average.

Tab. 3 (a) and (b) present the results for each subset of VisA and MPDD datasets, respectively. We observe that our MemKD outperforms the SPD [16] in Tab. 3 (a) by a large margin not only on the most difficult cases i.e., Macaroni1, Macaroni2, Capsules, and Candles but also on other relatively easier cases. In Tab. 3 (b), although RD [5] outperforms our method on P-AUC by 0.3%, the proposed MemKD significantly improves the baseline on metrics like I-AP and I-AUC.

Category/Method		MKD [12]	US [1]	DA [6]	MF [15]	PaDiM [4]	CutPaste [9]	DRAEM [13]	CFA [8]	RD [5]	PatchCore [11]	Ours
Textures	Carpet	98.1	91.6	86.6	94.0	**99.8**	97.0	93.9	97.3	98.9	98.7	99.6
	Grid	78.0	81.0	95.7	85.9	96.7	100.0	99.9	98.2	100.0	98.2	**100.0**
	Leather	95.1	88.2	86.2	99.2	100.0	100.0	100.0	99.0	100.0	100.0	**100.0**
	Tile	91.6	99.1	88.2	99.0	98.1	94.6	99.6	98.4	99.3	98.7	**100.0**
	Wood	94.3	97.7	98.2	99.2	99.2	99.1	99.1	**99.7**	99.2	99.2	99.5
	Average	*87.7*	*91.5*	*91.0*	*95.5*	*98.8*	*97.5*	*98.5*	*98.5*	*99.5*	*99.0*	***99.8***
Objects	Bottle	99.4	99.0	97.6	99.1	99.9	98.2	99.2	100.0	100.0	100.0	**100.0**
	Cable	89.2	86.2	84.4	97.1	92.7	81.2	92.8	**99.8**	95.0	99.5	99.2
	Capsule	80.5	86.1	76.7	87.5	91.3	98.2	98.5	97.3	96.3	98.1	**98.8**
	Hazelnut	98.4	93.1	92.1	99.4	92.0	98.3	100.0	100.0	99.9	100.0	**100.0**
	Metal Nut	73.6	82.0	75.8	96.2	98.7	99.9	98.7	100.0	100.0	100.0	**100.0**
	Pill	82.7	87.9	90.0	90.1	93.3	94.9	**98.9**	96.9	97.3	96.6	98.3
	Screw	83.3	54.9	98.7	97.5	85.8	88.7	94.9	97.3	97.0	98.1	**99.1**
	Toothbrush	92.2	95.3	99.2	100.0	96.1	99.4	100.0	100.0	99.5	100.0	**100.0**
	Transistor	85.6	81.8	87.6	94.4	97.4	96.1	94.1	100.0	96.7	100.0	**100.0**
	Zipper	93.2	91.9	85.9	98.6	90.3	99.9	**100.0**	98.6	98.5	99.4	99.3
	Average	*87.8*	*85.8*	*88.8*	*96.0*	*93.8*	*95.5*	*97.8*	*99.2*	*98.0*	*99.1*	***99.5***
Total Average		*87.8*	*87.7*	*89.5*	*95.8*	*95.5*	*96.1*	*98.1*	*98.7*	*98.4*	*99.2*	***99.6***

Table 1. Quantitative results for anomaly detection on MVTec AD [10]. We report comparison results on each class for anomaly detection. Methods achieving the top I-AUC (%) are highlighted in bold.

2. Forward Distillation with the Proposed Framework

Apart from the reverse distillation paradigm [5], we also apply the proposed framework to the forward distillation [12], which adopts a dual encoder architecture. Here, memories play a role in providing normality for encoders. Specifically, we exploit the WideResNet50 as the teacher and the vanilla ResNet50 as the student. Features from stage$_1$ to stage$_3$ are used for AD, and two NR Memory modules (M_1 and M_2) are inserted after the first two stages. Fig. 1 demonstrates its detailed structure.

Category/Method		MKD [12]	US [1]	MF [15]	SPADE [3]	PaDiM [4]	RIAD [14]	CP [9]	RD [5]	PatchC. [11]	Ours
Textures	Carpet	95.6/-	-/87.9	-/87.8	96.5/93.7	**99.1**/96.2	96.3/-	98.3/-	98.9/97.0	99.0/96.6	**99.1/97.5**
	Grid	91.8/-	-/95.2	-/86.5	92.7/85.7	97.3/94.6	98.8/-	97.5/-	**99.3/97.6**	98.7/96.0	99.2/96.9
	Leather	98.1/-	-/94.5	-/95.9	96.6/96.2	99.2/97.8	99.4/-	99.5/-	99.4/99.1	99.3/98.9	**99.5/99.2**
	Tile	82.8/-	-/**94.6**	-/88.1	86.4/74.9	94.1/86.0	89.1/-	90.5/-	95.6/90.6	95.6/87.3	**95.7**/91.1
	Wood	84.8/-	-/91.1	-/84.8	87.5/86.4	94.9/91.1	85.8/-	**95.5**/-	95.3/90.9	95.0/89.4	95.3/**91.2**
	Average	*90.6/-*	*-/92.7*	*-/88.6*	*91.9/87.4*	*96.9/93.2*	93.9/-	*96.3/-*	*97.7/95.0*	*97.5/93.6*	***97.8/95.2***
Objects	Bottle	96.3/-	-/93.1	-/88.8	97.4/91.5	98.3/94.8	98.4/-	97.6/-	98.7/96.6	98.6/96.2	**98.8/97.1**
	Cable	82.4/-	-/81.8	-/**93.7**	96.2/89.9	96.7/88.8	84.2/-	90.0/-	97.4/91.0	**98.4**/92.5	98.3/93.4
	Capsule	95.9/-	-/**96.8**	-/87.9	98.0/92.7	98.5/93.5	92.8/-	97.4/-	98.7/95.8	98.8/95.5	**98.8**/96.2
	Hazelnut	94.6/-	-/**96.5**	-/88.6	98.1/91.4	98.2/92.6	96.1/-	97.3/-	98.9/95.5	98.7/93.8	**99.1**/95.7
	Metal Nut	86.4/-	-/94.2	-/86.9	97.1/**91.4**	97.2/85.6	92.5/-	93.1/-	97.3/92.3	**98.4**/91.4	97.2/90.8
	Pill	89.6/-	-/96.1	-/93.0	95.5/89.6	95.7/92.7	95.7/-	95.7/-	98.2/96.4	97.4/93.2	**98.3/96.6**
	Screw	96.0/-	-/94.2	-/95.4	97.9/90.0	98.5/94.4	98.8/-	96.7/-	99.6/98.2	99.4/97.9	**99.6/98.2**
	Toothbrush	96.1/-	-/93.3	-/87.7	96.9/92.5	98.8/93.1	98.9/-	98.1/-	**99.1/94.5**	98.7/91.5	98.9/92.2
	Transistor	76.5/-	-/66.6	-/**92.6**	93.1/86.4	**97.5**/84.5	87.7/-	93.0/-	92.5/78.0	96.3/83.7	96.4/85.3
	Zipper	93.9/-	-/95.1	-/93.6	95.5/91.6	98.5/95.9	97.8/-	99.3/-	98.2/95.4	**98.8/97.1**	98.5/95.9
	Average	*90.8/-*	*-/90.8*	*-/90.8*	*96.6/93.4*	*97.8/91.6*	94.3/-	*95.8/-*	*97.9/93.4*	98.3/93.3	**98.4/94.1**
Total Average		*90.7/-*	*-/91.4*	*-/90.1*	*95.5/90.7*	*97.5/92.1*	94.2/-	*96.0-*	*97.8/93.9*	98.1/93.4	**98.2/94.5**

Table 2. Quantitative results for anomaly localization of P-AUC and P-PRO on MVTec AD [10]. P-AUC represents the pixel-wise comparison, while P-PRO focuses on region-based behavior. The best results for them are in bold. Remarkably, our approach is robust and presents state-of-the-art performance under both metrics.

	SPD [16]					Ours				
	I-AP	I-AUC	P-AP	P-AUC	P-PRO	I-AP	I-AUC	P-AP	P-AUC	P-PRO
PCB1	83.5	86.8	13.1	96.4	-	99.7	96.9	82.1	99.8	96.9
PCB2	76.9	76.6	9.6	96.3	-	94.8	98.0	25.2	96.0	94.9
PCB3	72.0	72.2	10.5	96.2	-	99.1	97.8	35.6	99.3	96.6
PCB4	93.8	95.2	8.5	86.7	-	98.6	99.8	44.3	98.6	99.9
Macaroni1	74.4	75.7	3.4	97.7	-	99.6	98.0	23.2	99.6	92.7
Macaroni2	62.3	66.8	0.6	94.3	-	99.2	92.0	13.0	99.5	84.8
Capsules	72.5	62.0	2.7	87.5	-	99.0	94.7	58.2	99.2	88.2
Candles	83.7	85.3	3.5	93.7	-	99.1	95.9	23.1	99.0	93.8
Cashew	93.8	86.6	9.5	86.3	-	98.7	99.4	58.2	96.6	97.5
Chewing Gum	98.2	96.7	28.5	97.0	-	99.1	99.8	60.3	98.6	98.8
Fryum	89.4	83.6	11.7	89.0	-	97.0	98.8	49.3	96.9	96.6
Pipe Fryum	93.6	87.1	11.7	91.6	-	99.2	100.0	56.2	99.2	99.0
Mean	82.8	81.2	9.4	92.7	-	**98.6**	**97.6**	**44.1**	**98.4**	**94.9**

(a) Anomaly detection and localization performance on the VisA [16] dataset.

	RD [5]					Ours				
	I-AP	I-AUC	P-AP	P-AUC	P-PRO	I-AP	I-AUC	P-AP	P-AUC	P-PRO
Bracket Black	92.0	86.0	12.1	98.0	94.6	94.4	91.2	10.7	97.8	94.5
Bracket Brown	95.2	91.9	25.3	97.6	95.6	97.3	95.2	20.5	96.3	95.2
Bracket White	85.9	82.0	1.3	98.6	93.5	93.8	92.7	15.9	98.8	97.3
Connector	100.0	100.0	66.0	99.5	97.5	100.0	100.0	60.6	99.4	96.4
Metal Plate	100.0	100.0	92.5	99.0	95.1	100.0	100.0	94.2	99.1	95.2
Tubes	98.6	96.3	76.1	99.2	95.9	98.5	96.2	74.9	99.2	97.3
Mean	95.3	92.7	45.5	**98.7**	95.3	**97.3**	**95.4**	**46.1**	98.4	**95.9**

(b) Anomaly detection and localization performance on the MPDD [7] dataset.

Table 3. Quantitative results on (a) VisA [16] and (b) MPDD [7] datasets. We report comparisons for each class. Best results are highlighted.

3. Additional Ablations

Study on different backbones. Tab. 4 illustrates qualitative comparisons of different backbones as the teacher network. The default WideResNet50 has a stronger representative capacity since it is deeper and wider, which facilitates the precise detection of anomalies. Built the proposed framework upon smaller neural networks such as ResNet18 and ResNet34, our MemKD still owns competitive performance.

Study on distilled features. Features from multiple stages contain patch embeddings with different receptive fields and can help the model identify anomalies from different levels. Tab. 5 explores their importance. We observe that combining S_1 and S_2 produces the worst results and S_1 gives limited improvement to the combination of S_2 and S_3, implying that semantics

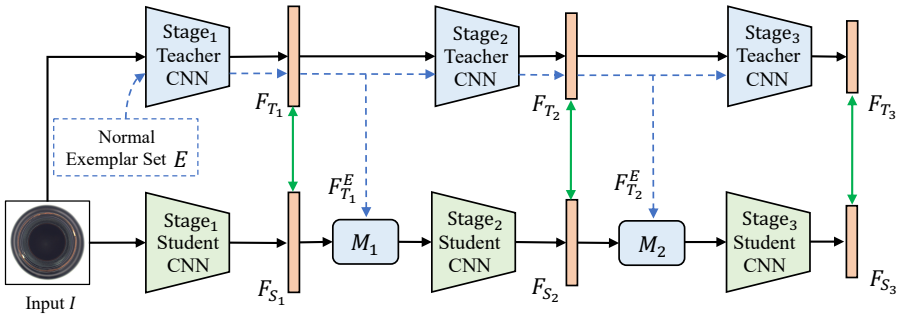


Figure 1. Overview of MemMKD, which derives from applying the proposed framework to forward distillation [12] paradigm. Two memory modules (M_1 and M_2) are inserted and dotted lines mean the normality embedding learning paths.

is more important. Overall, complementation of them helps cover various types of anomalies and thus exploiting them all yields satisfactory results.

Backbone	I-AUC	P-AUC	P-PRO
Res18	98.6	97.3	92.4
Res34	98.8	97.4	92.8
Res50	99.0	97.9	93.8
WideRes50	**99.6**	**98.2**	**94.5**

Table 4. Ablation study on different backbones.

Study on the number of NR memory. The inserted NR Memory modules (denoted as M_1, M_2, and M_3) are responsible for addressing the normality of the corresponding level. We study their importance in Tab. 6. Combining M_2 and M_3 has a similar performance on I-AUC and P-AUC as that of the combination of M_1 and M_3. Nonetheless, the former outperforms the latter by a clear margin on P-PRO. These observations imply that M_2 is more powerful than M_1. Without considering the high-level normal information (M_3), the model achieves the worst results. Overall, the high-level memory M_3 plays a more significant role and, no doubt, adopting them all contributes to the SOTA results.

S_1	S_2	S_3	I-AUC	P-AUC	P-PRO
√	√		98.6	97.3	93.8
√		√	98.9	98.0	94.0
	√	√	99.0	97.9	94.3
√	√	√	**99.6**	**98.2**	**94.5**

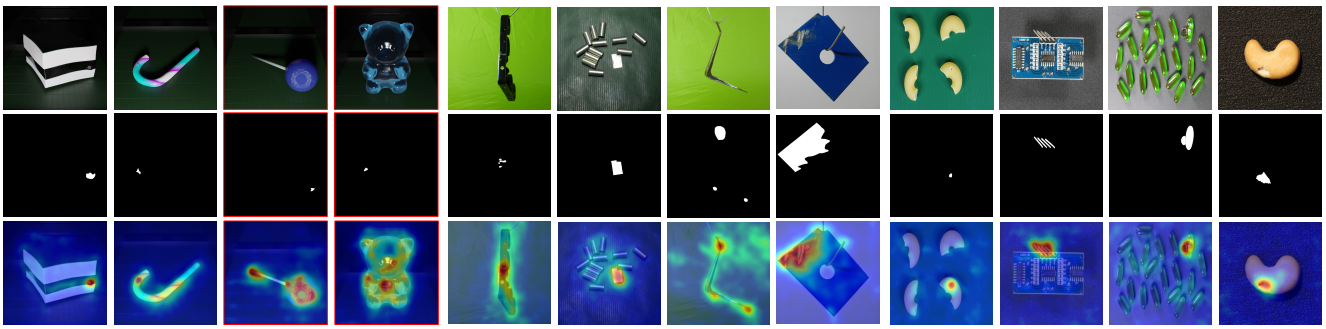
Table 5. Ablation study on distilled features.

M_1	M_2	M_3	I-AUC	P-AUC	P-PRO
	√	√	99.0	97.9	94.3
√		√	98.9	98.0	93.8
√	√		98.6	97.3	93.3
√	√	√	**99.6**	**98.2**	**94.5**

Table 6. Ablation study on the number of NR memory.

4. More Visualization on Anomaly Localization

We provide additional visualization on Eyecandies [2], MPDD [7] and VisA [16] benchmarks in Fig. 2 (a), (b) and (c), respectively. The proposed MemKD accurately localizes tiny and conspicuous anomalies, *e.g.*, the tiny hunch in the second column of Fig. 2 (a) and the large scratches in the fourth column of Fig. 2 (b). Besides, on the more challenging VisA dataset, it still gives satisfactory results, as shown in Fig. 2 (c). However, several issues exist. First of all, although the anomaly mask is accurate, it is not precise enough compared to the ground truth, which is also observed in other unsupervised AD methods [12, 5, 11]. We guess that the lack of supervision signals from anomalies results in imprecise localization results. Second, some anomalies are imperceivable since the single modality of RGB data can not provide sufficient information for anomaly detection, as demonstrated in the last columns of Fig. 2 (a).



(a) Visualization on Eyecandies [2] dataset. (b) Visualization on MPDD [7] dataset. (c) Visualization on VisA [16] dataset.

Figure 2. More qualitative results for anomaly localization on (a) Eyecandies [2] datasets, (b) MPDD [7] and (c) VisA [16]. From top to bottom: RGB image, ground truth, and the predicted anomaly map. The MemKD localizes tiny and conspicuous anomalies in these benchmarks. However, some anomalies can not be perceived only from the RGB data in the last two columns of (a).

5. Future Work

As mentioned in the paper, the RGB modality is insufficient to detect some geometrical anomalies. Therefore, considering multiple modalities for anomaly detection may give better results. In addition, the parameters of pre-trained teachers need to be fixed in both the training and testing phases, limiting the teacher's representation capacity on the target benchmarks. We believe that carefully devised adaptation methods would help produce better results.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *CVPR*, 2020. 1, 2

[2] Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. 2022. 3, 4

[3] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. In *ArXiv*, 2020. 2

[4] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *ICPR*, 2021. 1, 2

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

[6] Jinlei Hou, Yingying Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, and Hong Zhou. Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection. In *ICCV*, 2021. 1

[7] Stepan Jezek, Martin Jonák, Radim Burget, Pavel Dvorak, and Milos Skotak. Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. 2021. 1, 2, 3, 4

[8] Sungwook Lee, Seunghyun Lee, and Byung Cheol Song. Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. In *Access*, 2022. 1

[9] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, 2021. 1, 2

[10] Bergmann Paul, Fauser Michael, Sattlegger David, and Steger Carsten. Mvtec ad – a comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, 2019. 1, 2

[11] 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, 3

[12] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *CVPR*, 2021. 1, 2, 3

[13] Zavrtanik Vitjan, Kristan Matej, and Skočaj Danijel. Draem – a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, 2021. 1

[14] Zavrtanik Vitjan, Kristan Matej, and Skočaj Danijel. Reconstruction by inpainting for visual anomaly detection. In *PR*, 2021. 2

[15] Jhih-Ciang Wu, Ding-Jie Chen, Chiou-Shann Fuh, and Tyng-Luh Liu. Learning unsupervised metaformer for anomaly detection. In *ICCV*, 2021. 1, 2

[16] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. 2022. 1, 2, 3, 4