Prototype-Aware Contrastive Knowledge Distillation for Few-Shot Anomaly Detection – The Supplementary Material

Zhihao Gu[1]
ellery-holmes@sjtu.edu.cn
Lizhuang Ma[†] [1,2]
ma-lz@cs.sjtu.edu.cn
Taihai Yang[2]
thyang@stu.ecnu.edu.cn

[1] School of EIEE
Shanghai Jiao Tong University
Shanghai, China

[2] East China Normal University
Shanghai, China

This supplementary material first elaborates on RD [4] and the proposed PACKD. Then we provide results on MVTec 3D-AD [2] and Eyecandies [3] datasets in Tab. 1. Finaly, the subset-level results on MVTec AD [1] and MPDD [6] datasets are listed from Tab. 2 to 5.

Detailed Architecture of RD and PACKD

Fig. 1 (a) and (b) show the architecture of RD [4] and the proposed framework, respectively.

In Fig. 1 (a), given a support sample X_s, the original RD produces feature representations $\{F_s^{T_i}\}_{i=1}^3$ from the first three stages. Then, $F_s^{T_1} \in R^{C_1 \times H_1 \times W_1}$ and $F_s^{T_2} \in R^{C_2 \times H_2 \times W_2}$ are processed by convolution blocks to have the same spatial dimension as $F_s^{T_3} \in R^{C_3 \times H_3 \times W_3}$ and concatenated along channel with $F_s^{T_3}$. The resulting feature passes through a transformation encoder $\mathcal{T}$ to form the one-class embedding, denoted as $F_s^{T_4} \in R^{C_4 \times H_4 \times W_4}$. Finally, based on it, the student decoders aim to reconstruct the distillation targets $\{F_s^{T_i}\}_{i=1}^3$ and give the corresponding reconstructions $\{F_s^{S_i}\}_{i=1}^3$. The knowledge distillation loss Eq. (1) in the paper is used to enforce the feature consistency between $\{F_s^{T_i}, F_s^{S_i}\}_{i=1}^3$. In inference, the similarity between $\{F^{T_i}, F^{S_i}\}_{i=1}^3$ of the test sample is calculated for final anomaly detection.

Fig. 1 (b) demonstrates how we insert the prototype extraction and integration module (PEIM) and apply contrastive distillation strategy into RD. Differently, for the support image X_s, it only goes through the whole teacher network (owning four stages) and generates representations $\{F_s^{T_i}\}_{i=1}^3$. Contrarily, the query image X_q is handled by the RD and produces teacher representations $\{F_q^{T_i}\}_{i=1}^3$, one-class embedding $F_q^{S_4}$, and student representations $\{F_q^{S_i}\}_{i=1}^3$. The PEIM extracts prototypes $P_s^{T_i}$ from $F_s^{T_i}$ and integrate $P_s^{T_i}$ into $F_q^{S_i}$, leading to $\hat{F}_q^{S_i}$,where $i = 2,3,4$. Besides, the CDS is constructed based on $\{F_s^{T_i}, F_q^{T_i}, F_q^{S_i}, F_{q^-}^{S_i}\}_{i=1}^3$, where $F_{q^-}^{S_i}$ is the student representations of a query from a different category. And the orthogonal loss is conducted on $\{P_s^{T_i}\}_{i=2}^4$. In inference, the few-shot normal sample acts as the

[†]Corresponding author. Work done when Zhihao Gu is an intern at CATL.

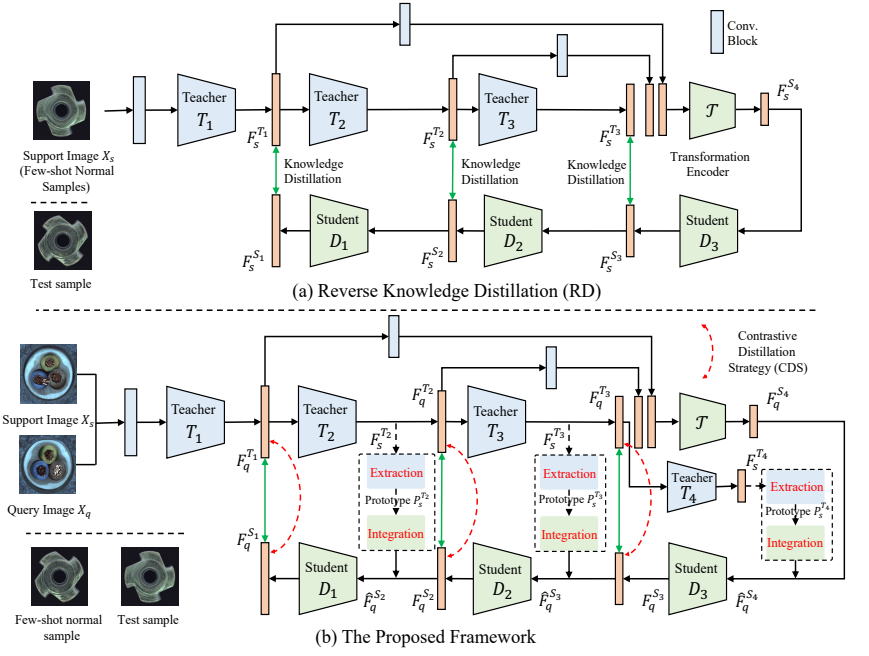

Figure 1: Overview of (a) RD and (b) the proposed PACKD framework in the few-shot setting. RD is trained on few-shot samples and evaluated on test samples of the same category. Contrarily, our PACKD is trained on existing categories while tested on unseen ones.

Method	MVTec 3D-AD [2]			Eyecandies [3]		
	k=2	k=4	k=8	k=2	k=4	k=8
PatchCore [4]	66.7	69.2	73.2	70.8	73.6	77.1
RegAD [5]	69.8	71.5	75.1	73.6	76.2	80.9
Ours	**72.7**	**74.6**	**78.9**	**77.0**	**79.1**	**83.6**

Table 1: Results on more datasets with image-level AUROC (%).

support sample, and the similarity between $\{F^{T_i}, F^{S_i}\}_{i=1}^{3}$ of the test sample are calculated for final anomaly detection.

Results on More datasets

Please note that we follow [5] to perform few-shot comparison on the MVTec AD and MPDD datasets *but* provide additional results on the VisA dataset to enhance the experiment section. To further enhance the experiments, we provide results on the MVTec 3D-AD [2] and the recently proposed Eyecandies [3] datasets in Table 1. It is observed that our method still achieves better results.

Category	k=2				k=4				k=8			
	TDG	DifferNet	RegAD	**Ours**	TDG	DifferNet	RegAD	**Ours**	TDG	DifferNet	RegAD	**Ours**
Bottle	69.3	99.3	99.4	99.4	69.6	99.3	99.4	99.5	70.3	99.4	99.8	99.8
Cable	68.3	85.3	65.1	84.7	70.3	85.2	76.1	86.9	74.7	87.9	80.6	92.5
Capsule	55.1	73.0	67.5	72.2	47.6	80.3	72.4	74.1	44.7	78.6	76.3	80.4
Carpet	66.2	78.4	96.5	97.2	68.7	78.6	97.9	98.3	78.2	78.5	98.5	99.0
Grid	83.8	62.1	84.0	85.0	86.2	60.5	91.2	88.5	87.6	78.5	91.5	95.3
Hazelnut	67.2	94.9	96.0	96.0	71.2	95.8	95.8	96.2	82.8	97.9	96.5	97.1
Leather	93.6	90.7	99.4	100.0	93.2	91.2	100.0	100.0	93.5	92.2	100.0	100.0
Metal Nut	67.1	61.9	91.4	91.9	69.2	67.3	94.6	93.8	68.7	67.6	98.3	98.7
Pill	69.2	83.2	81.3	81.8	64.7	84.0	80.8	82.0	67.9	82.1	80.6	84.2
Screw	98.8	73.4	52.5	85.0	98.8	72.5	56.6	86.1	99.0	75.0	63.4	92.6
Tile	86.3	97.0	94.3	97.3	87.2	98.0	95.5	97.9	87.4	99.6	97.4	99.0
Toothbrush	54.4	60.8	86.6	86.6	57.8	62.5	90.9	92.1	57.6	60.8	98.5	99.5
Transistor	55.9	61.8	86.0	84.9	67.7	62.2	85.2	86.5	71.5	63.3	93.4	96.6
Wood	98.4	98.1	99.2	99.2	98.3	96.4	98.6	99.5	98.4	99.4	99.4	99.6
Zipper	64.4	89.2	86.3	91.1	65.3	84.8	88.5	92.8	66.3	87.3	94.0	95.8
Average	73.2	80.6	85.7	**90.2**	74.4	81.3	88.2	**91.6**	76.7	83.2	91.2	**95.3**

Table 2: K-shot anomaly detection results in terms of image-level AUROC (%) on the MVTec AD dataset. Results are listed as the average AUROC of 10 runs.

Category	k=2			k=4			k=8		
	PatchCore	RegAD	**Ours**	PatchCore	RegAD	**Ours**	PatchCore	RegAD	**Ours**
Bottle	98.1	98.0	98.6	98.2	98.4	98.6	98.5	97.5	99.8
Cable	96.4	91.7	96.9	97.5	92.7	96.2	97.8	94.9	96.9
Capsule	96.5	97.3	97.7	96.8	97.6	97.7	97.7	98.2	98.4
Carpet	98.5	98.9	99.0	98.6	98.9	99.1	99.0	98.9	99.6
Grid	62.1	77.4	73.6	69.4	85.7	79.8	67.5	88.7	81.3
Hazelnut	96.3	98.1	97.0	97.6	98.0	97.9	96.4	98.5	98.6
Leather	99.0	98.0	95.9	99.1	99.1	99.3	99.3	98.9	99.6
Metal Nut	94.6	96.9	97.0	95.9	97.8	96.8	97.1	96.9	98.1
Pill	94.2	93.6	95.9	94.8	97.4	93.9	96.8	97.8	97.5
Screw	90.0	94.4	92.1	91.3	95.0	96.0	90.8	97.1	96.5
Tile	94.4	94.3	96.0	94.6	94.9	99.3	96.0	95.2	99.7
Toothbrush	97.5	98.2	98.8	98.4	98.5	98.2	98.2	98.7	98.9
Transistor	98.6	93.4	95.0	90.7	93.8	97.1	95.0	96.8	98.7
Wood	93.2	93.5	93.1	93.5	94.7	95.4	93.0	94.6	97.5
Zipper	98.0	95.1	98.3	98.1	94.0	98.5	98.2	97.4	99.3
Average	93.3	94.6	**95.0**	94.3	95.8	**96.2**	94.7	96.7	**97.3**

Table 3: K-shot anomaly detection results in terms of pixel-level AUROC (%) on the MVTec AD dataset. Results are listed as the average AUROC of 10 runs.

Category	k=2				k=4				k=8			
	TDG	DifferNet	RegAD	**Ours**	TDG	DifferNet	RegAD	**Ours**	TDG	DifferNet	RegAD	**Ours**
Bra. Black	46.4	56.7	63.3	68.2	48.8	59.9	63.8	70.2	51.0	69.7	67.3	70.5
Bra. Brown	54.9	61.3	59.4	64.7	57.5	64.2	66.1	67.8	65.4	66.3	69.6	69.0
Bra. White	64.0	42.2	55.6	58.5	65.4	51.8	59.3	59.7	66.8	69.1	61.4	60.2
Connector	53.1	54.1	73.0	74.5	55.8	54.8	77.2	76.5	62.9	54.5	84.9	75.3
Metal Plate	91.8	96.8	61.7	69.8	95.1	98.2	78.6	77.6	98.4	99.8	80.2	78.0
Tubes	51.8	49.8	67.1	64.1	58.5	50.7	67.5	69.4	64.9	52.6	67.9	70.4
Average	60.3	60.2	63.4	**66.6**	63.6	63.3	68.3	**69.8**	68.2	68.5	**71.9**	70.5

Table 4: K-shot anomaly detection results in terms of image-level AUROC (%) on the MPDD dataset. Results are listed as the average AUROC of 10 runs.

Category	k=2				k=4				k=8			
	RD	PatchCore	RegAD	**Ours**	RD	PatchCore	RegAD	**Ours**	RD	PatchCore	RegAD	**Ours**
Bra. Black	75.4	78.9	-	95.1	75.9	79.1	-	95.5	76.2	79.6	-	95.7
Bra. Brown	73.4	76.9	-	94.1	74.8	77.3	-	94.6	75.1	77.5	-	94.8
Bra. White	62.4	68.1	-	92.8	64.5	69.3	-	93.4	64.6	70.2	-	94.6
Connector	82.3	85.2	-	95.9	82.4	86.4	-	96.0	82.6	87.1	-	96.0
Metal Plate	76.5	86.3	-	94.8	77.2	86.7	-	95.3	77.5	86.9	-	95.6
Tubes	77.1	79.5	-	93.5	78.1	80.1	-	94.0	78.3	80.5	-	95.2
Average	74.5	79.2	93.2	**94.4**	75.5	79.8	93.9	**94.8**	75.7	80.3	95.1	**95.3**

Table 5: K-shot anomaly detection results in terms of pixel-level AUROC (%) on the MPDD dataset. Results are listed as the average AUROC of 10 runs.

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.

[2] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *arXiv preprint arXiv:2112.09045*, 2021.

[3] Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the Asian Conference on Computer Vision*, pages 3586–3602, 2022.

[4] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9727–9736, 2022.

[5] Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *Proceedings of the European Conference on Computer Vision*, pages 303–319, 2022.

[6] Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In *2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 66–71, 2021.

[7] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter V. Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14298–14308, 2022.