Supplementary Material for Registration based Few-Shot Anomaly Detection

Chaoqin Huang[1,3,4], Haoyan Guan[2], Aofan Jiang[1], Ya Zhang[1,3] ✉, Michael Spratling[2], and Yan-Feng Wang[1,3] ✉

[1] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University
{huangchaoqin, stillunnamed, ya_zhang, wangyanfeng}@sjtu.edu.cn
[2] King's College London
{haoyan.guan, michael.spratling}@kcl.ac.uk
[3] Shanghai Artificial Intelligence Laboratory
[4] National University of Singapore

1 Main Contributions

This paper targets a challenging yet practical setting for anomaly detection, with 1) a single model for all categories (*i.e., generalizable without fine-tuning*), 2) only a few images for each novel category (*i.e., few shot*), and 3) only normal samples available (*i.e., unsupervised setting*). To our best knowledge, it is the first attempt to explore such a setting, as a critical step toward practical large-scale industrial applications, a point appreciated by the other reviewers. To learn a category-agnostic model, we further propose a novel comparison-based solution, which is quite different from the popular reconstruction-based or classification-based methods. We adopt STN to align the images and Siamese network to implement the comparison. The SOTA results achieved on MVTec and MPDD show the effectiveness of our method.

2 Experiments

2.1 Ablation Studies on Support Set Augmentations.

The proposed support set augmentations are shown to be essential for both detection and localization. Table 1 further presents the ablation studies of comparing different augmentation methods for support images with $k = 2$. The experimental results have validated the effectiveness of all the proposed augmentation methods. In particular, rotation and translation are shown to perform better on MVTec, while flipping and rotation seem to perform better on MPDD.

2.2 Comparisons with Metaformer [3].

Although using different training settings, according to the reported results in [3], Metaformer achieves about 88% AUC for MVTec when k=8, while RegAD achieves 91.2% AUC, an ≈3% improvement, with the same test set and

Table 1. Ablation studies of different versions of support set augmentations on the MVTec and MPDD datasets with $k = 2$. Besides the full version of RegAD, We also provide the individual training version to reduce the influence of data augmentations on multiple categories. G, F, T, R means graying, flipping, translation, and rotation, respectively. Results are listed as the macro-average AUC in % over all categories in each dataset of 10 runs. The best-performing method is in bold.

Augmentations				Individual Training				Aggregated Training			
				MVTec		MPDD		MVTec		MPDD	
G	F	T	R	image	pixel	image	pixel	image	pixel	image	pixel
				74.7	88.6	49.6	89.5	79.1	90.5	57.6	91.0
✓				74.9	88.6	49.4	89.4	79.5	90.7	58.0	91.3
	✓			75.5	90.0	50.1	90.2	79.8	92.6	58.6	92.6
		✓		77.4	90.9	49.8	91.5	81.3	92.4	58.3	90.9
			✓	79.6	92.7	50.0	91.7	82.2	93.6	59.8	90.8
✓	✓			75.6	89.9	50.0	90.1	80.5	92.7	59.7	91.7
✓		✓		77.5	90.9	49.8	91.4	81.0	92.4	58.5	92.6
✓			✓	79.7	92.6	50.0	91.6	83.8	93.7	60.9	92.6
	✓	✓		77.7	91.5	50.1	91.7	81.6	93.2	58.2	92.3
	✓		✓	79.7	92.9	51.3	91.8	82.3	94.0	59.7	92.9
		✓	✓	81.3	93.1	49.9	92.2	84.2	94.6	60.6	91.7
✓	✓	✓		77.8	91.5	50.2	91.6	81.7	93.5	59.9	92.6
✓	✓		✓	79.8	92.9	51.2	91.7	83.9	94.2	63.0	93.0
✓		✓	✓	81.4	93.1	49.9	92.3	84.9	**94.7**	61.2	92.8
	✓	✓	✓	**81.5**	**93.3**	50.7	**92.4**	85.4	94.6	61.2	**93.2**
✓	✓	✓	✓	**81.5**	**93.3**	**50.8**	**92.4**	**85.7**	94.6	**63.4**	**93.2**

Table 2. Comparison with AD method trained by full data.

	RegAD			PatchCore [2]	CflowAD [1]
k	32	64	128	full data	full data
MVTec	94.6%	95.5%	95.9%	99.1%	98.3%
MPDD	76.8%	82.3%	83.2%	82.1%	86.1%

evaluation protocol. Metaformer achieves worse results despite three unfair advantages: (i) an additional large-scale dataset, MSRA10K, is used during its entire meta-training procedure (beyond parameter pre-training), together with additional pixel-level annotations; (ii) it performs additional fine-tuning on each novel category; (iii) it is trained with a deep transformer architecture for more epochs (100 vs. 50), with a larger batch size (64 vs. 32).

2.3 Experiments with a Large k.

To decrease the training burden, RegAD is designed to be adaptable to unseen categories without parameter fine-tuning. Without fine-tuning on the few-shot support examples, simply increasing the shot number, the performance gain saturates very soon. We further experiment with k=64 and k=128. As shown in Table 2, when k increases from 64 to 128, a limited performance gain is observed. But the results are still competitive, though with a shallow backbone, compared to those of the AD methods trained by full data.

References

1. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 98–107 (2022) 2
2. 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 (CVPR). pp. 14318–14328 (2022) 2
3. Wu, J.C., Chen, D.J., Fuh, C.S., Liu, T.L.: Learning unsupervised metaformer for anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4369–4378 (2021) 1