Prototypical Residual Networks for Anomaly Detection and Localization Supplemental Material

Hui Zhang[1,2] Zuxuan Wu[1,2] Zheng Wang[3] Zhineng Chen[1,2*] Yu-Gang Jiang[1,2]
[1]Shanghai Key Lab of Intell. Info. Processing, School of CS, Fudan University
[2]Shanghai Collaborative Innovation Center of Intelligent Visual Computing
[3]School of Computer Science, Zhejiang University of Technology

A. Anomaly Generation Strategies

This section details the generation of simulated anomalies, as shown in Fig. 1. A noise image is generated by a Perlin noise generator [8, 13] (Fig. 1,P), and then the noise parts within a target area are retained as the ground truth mask (Fig. 1,M). As the shape, size, and number of generated anomalous regions vary widely, we synthesize simulated anomalies (Fig. 1,S) as:

$$S = \bar{M} \odot N + (1-\beta)(M \odot A) + \beta\,(M \odot N) \quad (1)$$

where N is the normal sample, A is the source image of the anomaly, $\bar{M}$ is the inverse of M, $\odot$ is the element-wise multiplication operation, β is the opacity parameter for better combination of abnormal and normal regions. When A is an image randomly sampled from the DTD dataset [3] and is augmented (Aug_1, Fig. 5 in Section 3.4), we define S as a HEterologous Anomaly (HEA). Correspondingly, when A is an image randomly sampled from augmented normal samples, we define S as a HOmology Anomaly (HOA). In particular, the normal image is first augmented (Aug_1, Fig. 5 in Section 3.4), then is evenly divided into an 8×8 grid and randomly arranged before being reassembled [12].

Fig. 2 shows the anomalies generated by different strategies. In addition to increasing the number, extended anomalies (EA) increase the variety of seen anomalies. HEA and HOA supplement potential unseen anomalies with anomalies significantly different from seen anomalies.

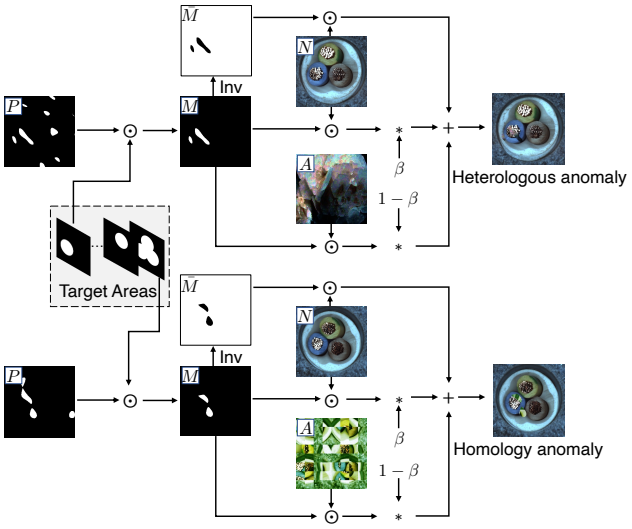

Figure 1. Generating simulated anomalies.

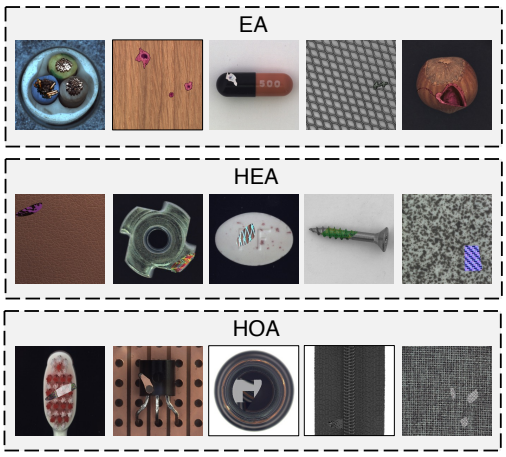

Figure 2. Examples of anomalies generated by different strategies.

B. Experiments

B.1. Dataset Split

MVTec AD [1] is a widely used anomaly detection and localization benchmark with 15 classes, each containing one to several subclasses of anomalies. Following the general setting proposed by DRA [5], the 10 labeled anomaly samples are sampled from all possible anomaly classes in the test set per dataset. These sampled anomalies are then removed from the test data. Both BTAD [7] and KolektorSDD2 [2] are real-world industrial datasets containing three product types and one product type, respectively. The general setting

Category	DRAEM [13]				CFLOW [6]				SSPCAB [9]				RD4AD [4]				PatchCore [10]				**Ours**			
	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑
Class1	86.9	75.4	56.3	20.9	91.6	**94.1**	84.1	33.4	95.3	78.9	61.2	29.9	95.2	92.8	83.0	41.6	84.4	89.6	72.8	13.1	**100**	92.7	**90.3**	**50.1**
Class2	85.8	83.7	66.1	18.2	98.2	99.6	98.2	50.2	93.9	92.0	80.4	18.3	99.7	**99.7**	99.1	**57.8**	**100**	**99.7**	**99.3**	55.5	96.0	97.1	95.6	44.8
Class3	98.0	90.3	78.2	32.9	88.3	93.7	86.0	32.9	**99.6**	90.3	79.2	31.7	81.2	93.9	85.2	31.8	94.0	**96.2**	**92.4**	**50.9**	99.2	94.2	91.3	32.4
Class4	99.3	98.6	95.5	62.4	**100**	**99.5**	**98.5**	65.1	99.9	99.1	97.6	74.7	99.9	99.1	97.7	64.6	100	99.4	98.4	**88.2**	99.7	98.2	96.7	67.2
Class5	**97.9**	56.4	39.9	21.9	86.3	94.3	84.5	**50.7**	81.1	53.6	35.9	15.5	74.1	86.7	64.3	31.2	90.6	**95.2**	77.3	29.6	96.9	94.9	**86.1**	30.2
Class6	**100**	96.0	89.3	71.5	96.5	96.1	87.9	46.9	**100**	95.4	88.3	70.0	92.0	88.3	68.9	30.3	99.4	98.1	93.5	71.2	**100**	**98.4**	**95.7**	**71.7**
Class7	**100**	96.7	90.8	58.1	98.9	96.0	91.8	61.4	100	94.8	87.0	51.1	99.8	95.2	91.4	65.7	99.9	**96.9**	**94.8**	**77.7**	**100**	95.1	91.3	51.3
Class8	**99.7**	92.9	90.4	34.2	56.7	79.9	51.0	3.2	96.4	91.1	88.9	23.2	65.2	86.2	67.6	7.0	60.6	86.4	56.5	7.8	93.4	**97.1**	**95.1**	**34.4**
Class9	50.2	49.7	13.3	0.1	99.9	**99.9**	**99.8**	**65.1**	50.9	60.4	26.1	0.1	**100**	99.8	99.4	26.5	96.4	99.4	95.7	45.9	97.1	98.7	96.8	46.4
Class10	92.7	94.2	85.4	35.7	95.7	98.0	94.4	42.9	86.5	89.1	74.7	24.4	99.6	99.0	97.9	51.1	**99.9**	**99.6**	**99.0**	49.6	**99.9**	**99.6**	**99.0**	**65.6**
Average	91.1	83.4	70.5	35.6	91.2	95.1	87.6	45.2	90.4	84.5	71.9	33.9	90.7	94.1	85.5	40.8	92.5	96.1	88.0	49.0	**98.2**	**96.6**	**93.8**	**49.4**

Table 1. Anomaly Detection and Localization on DAGM [11]. "I", "P", "O" and "A" respectively refer to the five metrics of image auroc, pixel auroc, pro and ap. The best results are highlighted in bold.

Category	DRAEM [13]				CFLOW [6]				SSPCAB [9]				RD4AD [4]				PatchCore [10]				**Ours**			
	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑
01	98.5	91.5	61.4	17.0	93.4	94.8	60.1	**39.6**	96.2	92.4	62.8	18.1	98.8	95.7	72.8	49.3	96.6	96.5	78.4	47.1	**100**	**96.6**	**81.4**	38.8
02	68.6	73.4	39.0	23.3	79.0	93.9	**56.9**	65.5	69.3	65.6	28.6	15.8	**84.9**	**96.0**	55.8	**66.1**	81.3	94.9	54.0	56.3	84.1	95.1	54.4	65.7
03	99.8	96.3	84.3	17.2	99.1	99.5	97.9	56.8	99.4	92.4	71.0	5.0	99.5	99.0	98.8	45.1	**99.9**	99.2	96.4	51.2	**99.9**	**99.6**	**98.3**	**57.4**
Average	89.0	87.1	61.6	19.2	90.5	96.1	71.6	**54.0**	88.3	83.5	54.1	13	94.4	96.9	75.8	53.5	92.6	96.9	76.3	51.5	**94.7**	**97.1**	**78.0**	**54.0**

Table 2. Anomaly Detection and Localization on BTAD [7].

used in BTAD and SDD2 is same to that used in MVTec. DAGM [11] contains 10 texture classes, and the original training set for each class consists of normal and abnormal samples. For each class, we first move all anomalous samples from the original training set to the original test set, and then randomly select ten anomalous samples from the test set as part of the new training set. These sampled anomalies are then removed from the test set.

B.2. More Detailed Comparison

Table 1 includes fine-grained anomaly detection and localization performance comparisons on all DAGM sub-datasets. We observe that PRN consistently performs well on all 10 sub-datasets and, in the average scenario, performs best across all four criteria. In particular, our approach outperforms previous methods by a large margin in two metrics, image auroc and pro.

We also compare the anomaly detection and location performance of each method in detail on the three BTAD products and report the numerical results in Table 2. It can be concluded that our method achieves consistently higher performance than the others on different categories .

B.3. More Qualitative Examples

We further qualitatively evaluate the performance of anomaly detection and location compared to state-of-the-art methods by introducing additional visualizations, as shown in Fig. 3, Fig. 4 and Fig. 5. Our method accurately detects and localizes anomalies in a wide range of sizes, shapes and numbers, as demonstrated by qualitative comparison results. Moreover, we argue that some of the localization errors can be attributed to inaccurate ground truth labels on anomalies. An example of this is shown in the second row of Fig. 5, where the ground truth does not label all anomalous regions. Another example is shown on the left in the fourth row of Fig. 4, where the ground truth labels a broad anomaly region, but our method correctly localizes the anomaly region. These imprecise annotations inevitably impact the anomaly localization scores of the evaluated methods.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, 2019. 1, 3

[2] Jakob Božič, Domen Tabernik, and Danijel Skočaj. Mixed supervision for surface-defect detection: From weakly to fully supervised learning. *Comput Ind*, 2021. 1

[3] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *CVPR*, 2014. 1

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

[5] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *CVPR*, 2022. 1

[6] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *WACV*, 2022. 2

[7] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In *ISIE*, 2021. 1, 2, 3

[8] Ken Perlin. An image synthesizer. *ACM SIGGRAPH*, 1985. 1

[9] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moes-

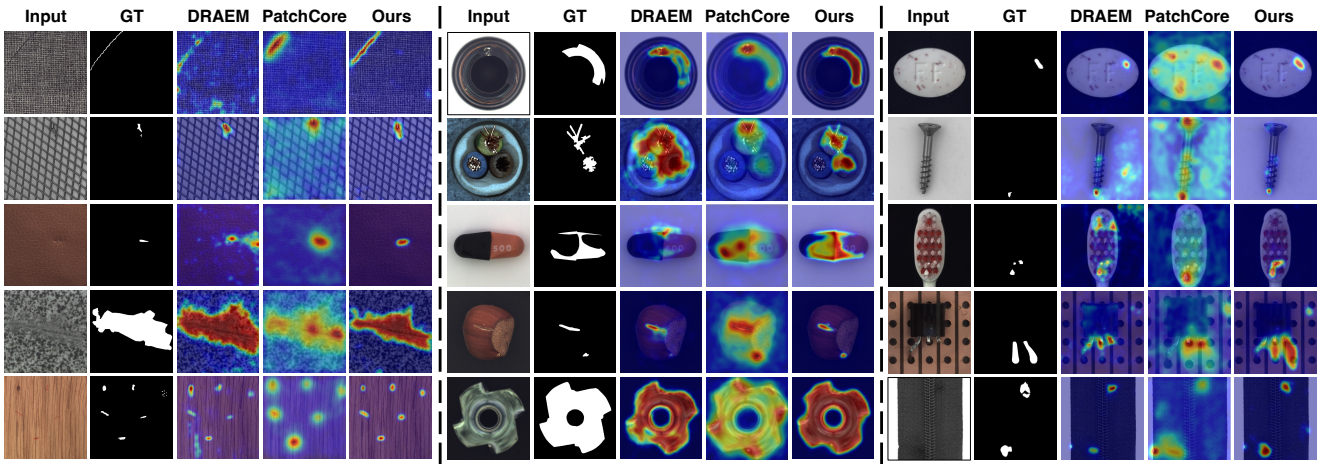

Figure 3. More qualitative examples on MVTec [1].

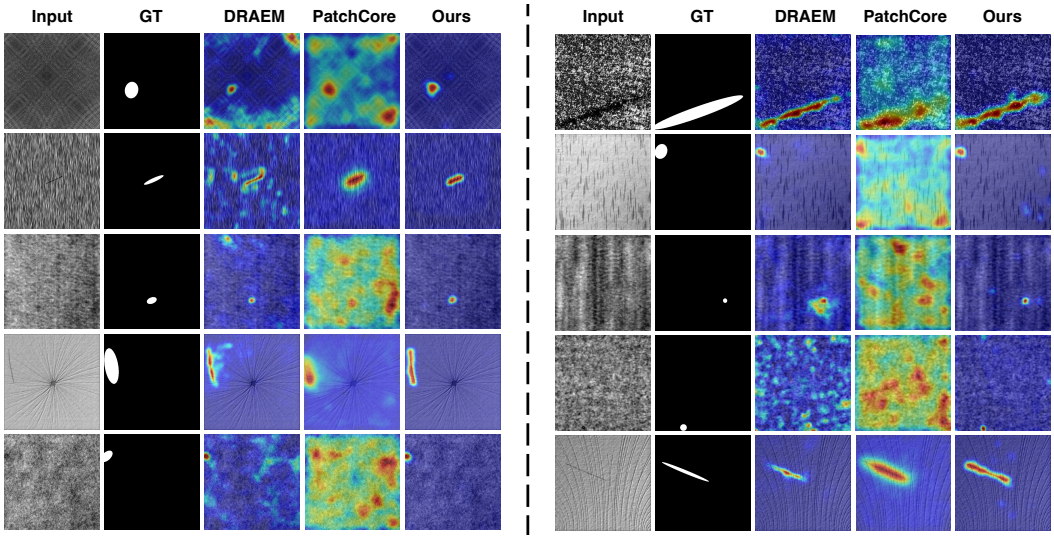

Figure 4. Qualitative examples on DAGM [11].

lund, and Mubarak Shah. Self-supervised predictive convolutional attentive block for anomaly detection. In *CVPR*, 2022. 2

[10] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, 2022. 2

[11] Matthias Wieler and Tobias Hahn. Weakly supervised learning for industrial optical inspection. 2007. 2, 3

[12] Minghui Yang, Peng Wu, Jing Liu, and Hui Feng. Memseg: A semi-supervised method for image surface defect detection using differences and commonalities. *arXiv preprint arXiv:2205.00908*, 2022. 1

[13] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, 2021. 1, 2

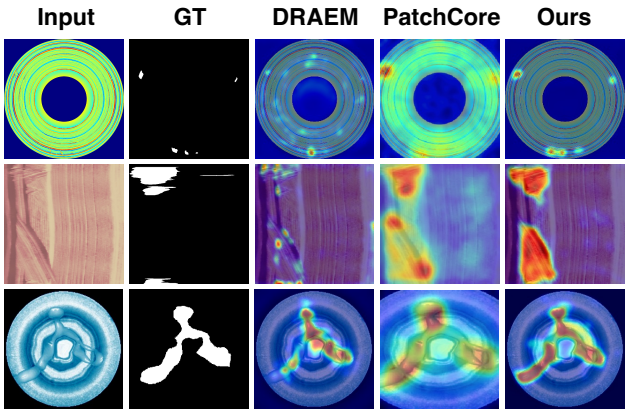

Figure 5. Qualitative examples on BTAD [7].