Supplementary Material for "A Prototype-Based Neural Network for Image Anomaly Detection and Localization"

Chao Huang[1], Zhao Kang[1] and Hong Wu[1*]

[1*]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, Sichuan, China.

*Corresponding author(s). E-mail(s): hwu@uestc.edu.cn;
Contributing authors: chaohuang@std.uestc.edu.cn;
zkang@uestc.edu.cn;

1 Implementation Details

We implemented ProtoAD in Python 3.6 [1] and PyTorch [2]. Our experiments are run on Intel(R) Xeon (R) CPU E5-2678 v3 @ 2.50GHz and a single 12GB NVIDIA GeForce RTX 2080 Ti. Following [3–5], We use the WideResNet50 [6] pretrained on ImageNet as the feature extractor (from torchvision). For each category of MVTec AD [7], we resize the image to 256×256 and center crop them to 224×224, and no data augmentation is applied.

We adopt the feature maps output by stage 1-3 of WideResNet50 and combine them as the patch features. Specifically, we upsampled features at level 2 and level 3 to the same resolution as level 1 (1/4 of the input image) and concatenate them along the channel dimension. After the patch features of normal images are extracted, we use FINCH [8] to learn prototypes for them, and choose the first partition generated so far having less than 10,000 clusters as the clustering result. The mean vectors of clusters serve as the prototypes for the normal patch features. Finally, we construct an image anomaly localization network (ProtoAD) by appending the feature extraction network with a 1×1 convolutional layer, a channel max-pooling, and a subtraction operation. We use the prototypes as the kernels (no bias used) of the 1×1 convolutional layer; therefore, our neural network does not need a training phase. When a test image passes through ProtoAD network, an anomaly score map will be generated, and it will be up-sampled to the image size with bilinear interpolation and be smoothed using the Gaussian filter with parameter $\delta = 4$ like in [4].

2 Full Performance with Different FINCH Partitions

FINCH is a hierarchical agglomerative clustering algorithm. It recursively merges clusters from the bottom up and provides a set of partitions in a hierarchical structure. Each successive partition is a super-set of its preceding partitions, and the number of clusters in it is smaller than those in the preceding partitions.

We report the performance (on each MVTec category and averaged) of our method with different partitions generated by FINCH (from the second (P2) to the 6-th (P6) partition) in table 1. We do not include the first partition because it has a huge number of clusters. The results in table 1 indicate the average performance decreases along with the merging process. We also give the best performance, which FINCH can achieve by selecting the best partition for each category respectively. This best performance is the upper bound that our method can achieve. In this paper, we adopts a simple rule which selects the first partition generated so far by FINCH with less than 10,000 clusters as the clustering result, and give its results in the last column of table 1.

Table 1 Anomaly detection and localization performance of ProtoAD with different FINCH partitions. Each tuple shows image-level AUROC, pixel-level AUROC, and average cluster numbers.

Category	P2	P3	P4	P5	P6	Best	Our
Carpet	(99.6, 99.2, 35674)	(99.5, 99.2, 1816)	(100.0, 99.2, 56)	(100.0, 99.2, 13)	(96.5,98.6, 3)	(100.0, 99.2, 56)	(99.5, 99.2, 1816)
Grid	(96.7, 98.8, 67994)	(94.7, 98.5, 16112)	(93.7, 98.0, 3983)	(92.4, 96.8, 1025)	(85.7,94.8, 260)	(96.7, 98.8, 67994)	(93.7, 98.0, 3983)
Leather	(100.0, 99.4, 13702)	(100.0, 99.4, 672)	(100.0, 99.4,99)	(100.0, 99.4, 24)	(100.0, 99.4, 6)	(100.0, 99.4, 6)	(100.0, 99.4, 672)
Tile	(99.4, 95.4, 18101)	(99.2, 95.2, 826)	(99.1, 95.0, 51)	(98.6, 94.6, 10)	(97.9, 94.0, 3)	(99.4, 95.4, 18101)	(99.2, 95.2, 826)
Wood	(98.9, 95.7, 105187)	(99.0, 95.6, 19585)	(99.1, 95.6, 1639)	(99.4, 95.5, 98)	(98.8, 95.2, 9)	(99.1, 95.6, 1639)	(99.1, 95.6, 1639)
Texture	(98.9, 97.7, 48132)	(98.5, 97.6, 7802)	(98.4, 97.4, 1166)	(98.1, 97.1, 234)	(95.8, 96.4, 56)	(99.0, 97.7, 17559)	(98.3, 97.5, 1787)
Bottle	(99.9, 98.4, 61108)	(99.8, 98.4, 11946)	(99.9, 98.3, 2106)	(99.8, 98.3, 595)	(99.6, 98.0, 189)	(99.9, 98.4, 61108)	(99.9, 98.3, 2106)
Cable	(99.5, 97.5, 96829)	(98.8, 97.7, 21466)	(98.3, 97.5, 2611)	(96.2, 97.0, 535)	(94.8, 96.3, 136)	(99.5, 97.5, 96829)	(98.3, 97.5, 2611)
Capsule	(94.8, 98.3, 79466)	(94.2, 98.3, 12666)	(93.1, 98.2, 2517)	(87.6, 98.0, 2517)	(84.5, 97.6, 235)	(94.8, 98.3, 79466)	(93.1, 98.2, 2517)
Hazelnut	(100.0, 98.8, 173989)	(100.0, 98.8, 46894)	(100.0, 98.8, 7127)	(100.0, 98.8, 805)	(99.4, 98.6, 171)	(100.0, 98.8, 805)	(100.0, 98.8, 7127)
Metal Nut	(100.0, 96.7, 91055)	(100.0, 96.9, 23558)	(99.9, 96.8, 4969)	(99.8, 96.6, 1204)	(99.3, 95.7, 310)	(100.0, 96.9, 23558)	(99.9, 96.8, 4969)
Pill	(96.8, 94.4, 96509)	(96.7, 94.6, 17891)	(95.8, 94.2, 2202)	(91.8, 94.0, 519)	(90.8, 94.5, 147)	(96.7, 94.6, 17891)	(95.8, 94.2, 2202)
Screw	(97.6, 99.0, 129710)	(95.5, 99.1, 33990)	(94.9, 98.9, 5504)	(89.4, 98.3, 1140)	(78.8, 97.1, 258)	(97.6, 99.0, 129710)	(94.9, 98.9, 5504)
Toothbrush	(100.0, 98.8, 24794)	(99.7, 98.8, 5704)	(99.4, 98.8, 1478)	(98.3, 98.7, 417)	(97.8, 98.4, 108)	(100.0, 98.8, 24794)	(99.7, 98.8, 5704)
Transistor	(98.1, 92.9, 104018)	(98.3, 93.3, 22724)	(97.7, 92.5, 4195)	(96.6, 90.6, 999)	(93.5, 87.1, 271)	(98.3, 93.3, 22724)	(97.7, 92.5, 4195)
Zipper	(95.6, 96.7, 71559)	(94.6, 96.7, 9323)	(94.7, 96.6, 1260)	(92.5, 96.2, 288)	(83.7, 95.6, 79)	(95.6, 96.7, 71559)	(94.6, 96.7, 9323)
Object	(98.2, 97.2, 92904)	(97.8, 97.3, 20616)	(97.4, 97.1, 3397)	(95.2, 96.6, 902)	(92.2, 95.9, 190)	(98.2, 97.2, 52844)	(97.4, 97.1, 4626)
All	(98.5, 97.3, 77980)	(98.0, 97.4, 16345)	(97.7, 97.2, 2653)	(96.1, 96.8, 679)	(93.4, 96.1, 146)	(98.5, 97.4, 41083)	(97.7, 97.2, 3680)

3 Qualitative Results

Figure 1 - 5 gives qualitative anomaly localization results for different categories of the MVTec AD dataset. From top down are the anomaly images, the corresponding ground-truth masks, and the anomaly score map predicted by ProtoAD.

References

[1] Van, R.G., Drake, F.: Python 3 reference manual. Scotts Valley, CA: CreateSpace **10**, 1593511 (2009)

[2] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (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 (2021). Springer

[5] 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)

[6] Zagoruyko, S., Komodakis, N.: Wide residual networks. In: British Machine Vision Conference 2016 (2016). British Machine Vision Association

[7] 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)

[8] Sarfraz, S., Sharma, V., Stiefelhagen, R.: Efficient parameter-free clustering using first neighbor relations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8934–8943 (2019)

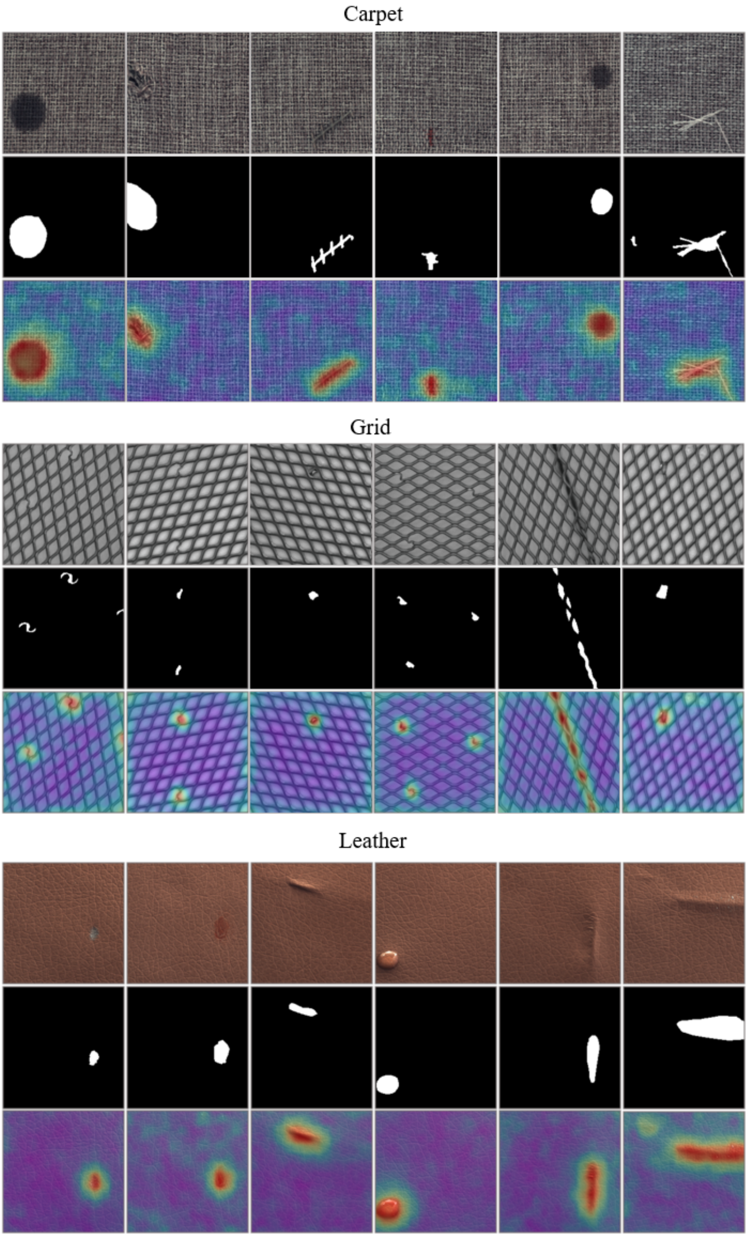

Fig. 1 Anomaly localization on carpet, grid and leather class of MVTec AD dataset. From top to bottom are input images, ground-truth localization masks, and ProtoAD prediction.

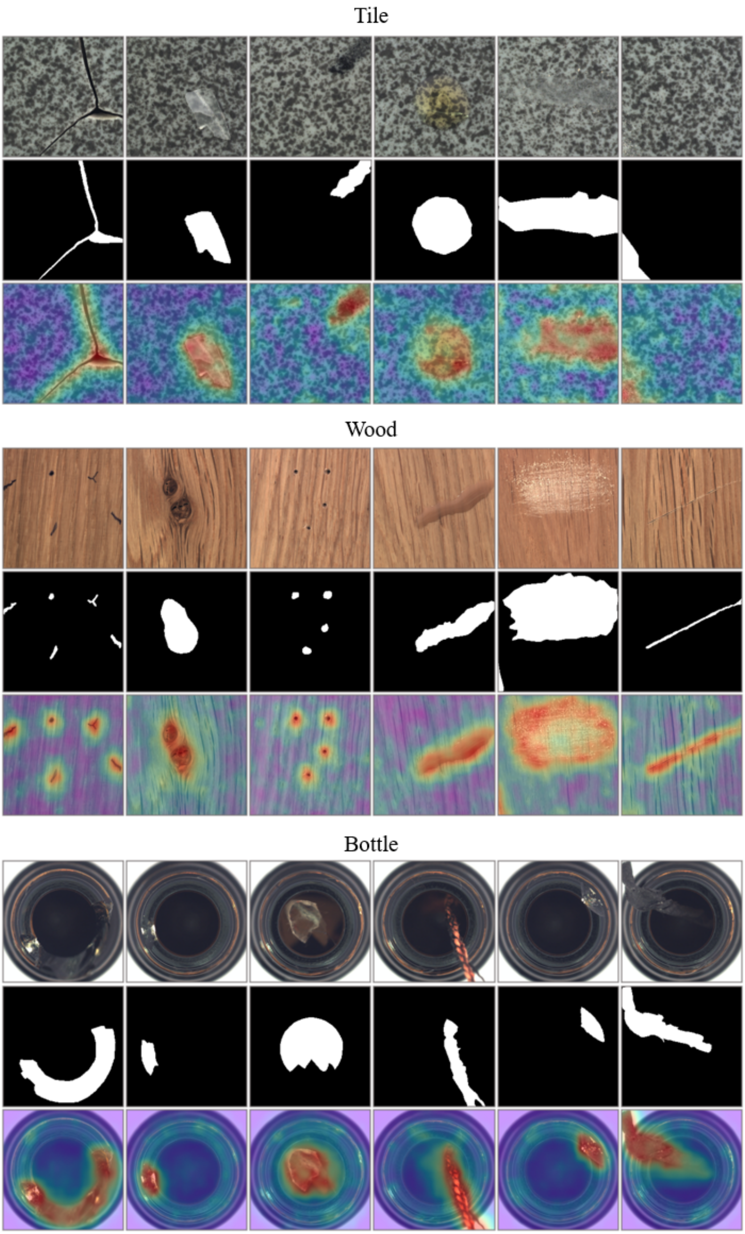

Fig. 2 Anomaly localization on tile, wood and bottle class of MVTec AD dataset. From top to bottom are input images, ground-truth localization masks, and ProtoAD prediction.

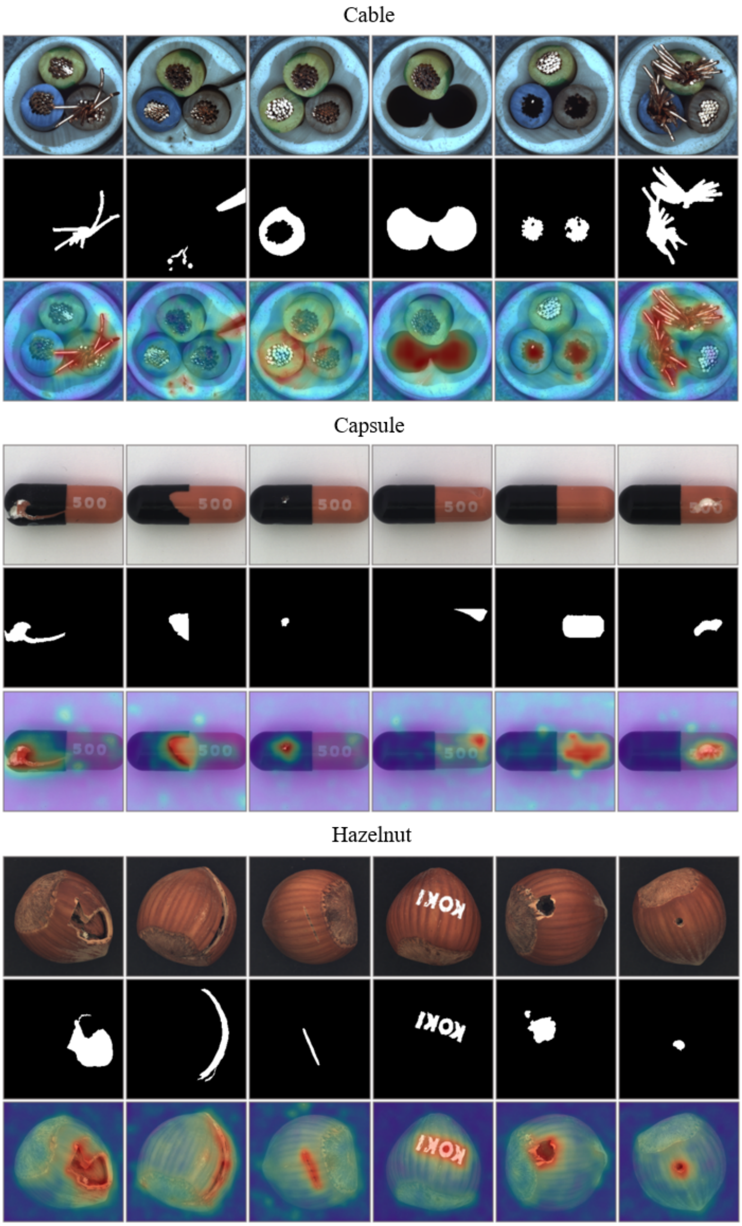

Fig. 3 Anomaly localization on cable, capsule and hazelnut class of MVTec AD dataset. From top to bottom are input images, ground-truth localization masks, and ProtoAD prediction.

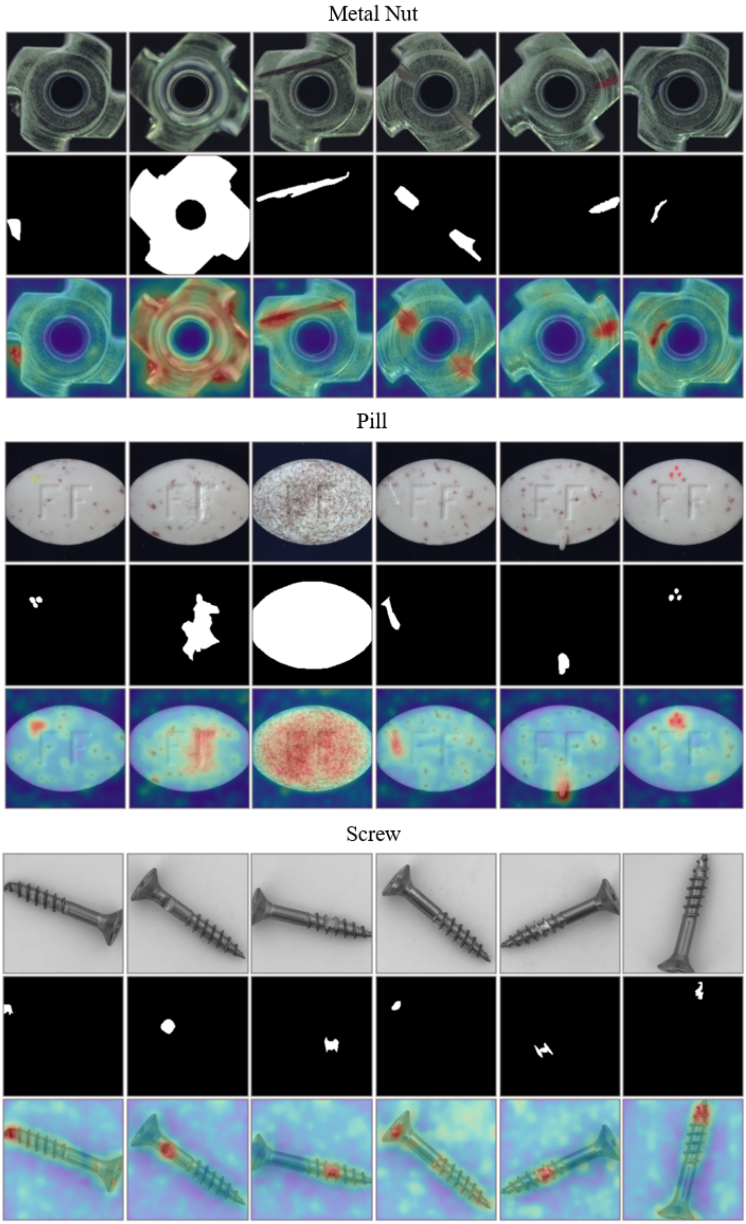

Fig. 4 Anomaly localization on metal nut, pill and screw class of MVTec AD dataset. From top to bottom are input images, ground-truth localization masks, and ProtoAD prediction.

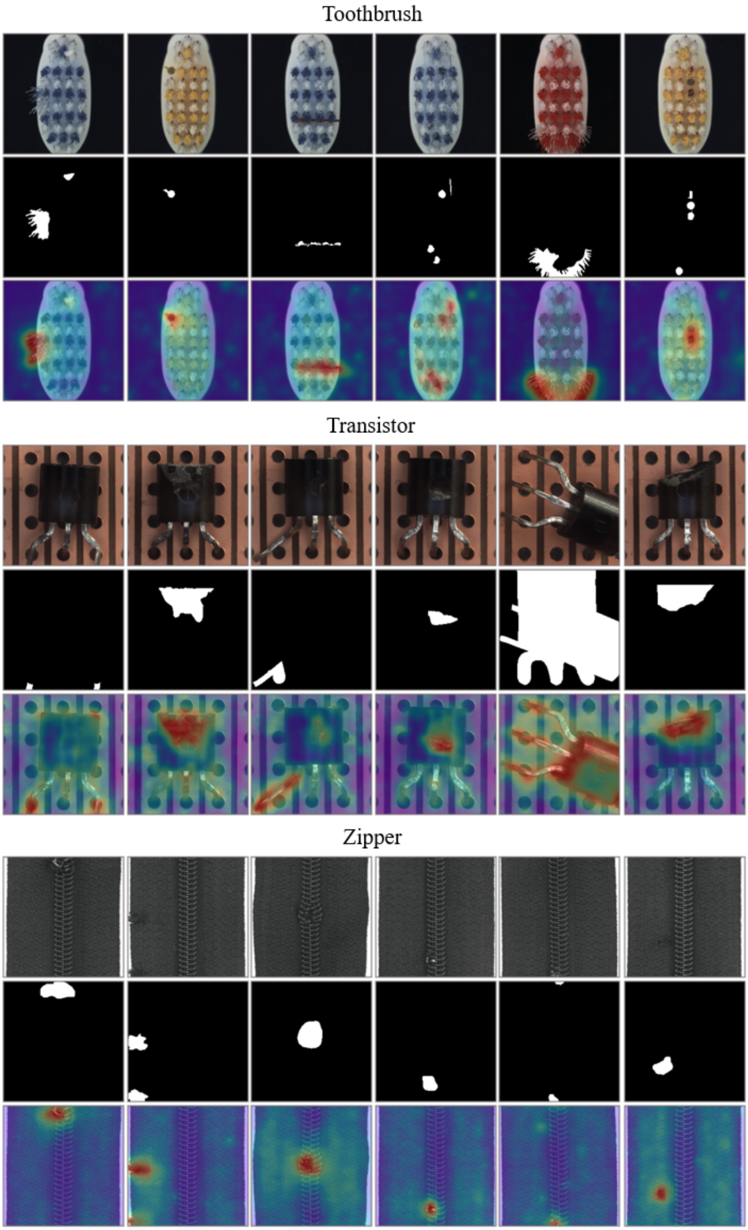

Fig. 5 Anomaly localization on toothbrush, transistor and zipper class of MVTec AD dataset. From top to bottom are input images, ground-truth localization masks, and ProtoAD prediction.