MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Supplementary Material – Towards Zero-shot 3D Anomaly Localization

Wang, Yizhou; Peng, Kuan-Chuan; Fu, Raymond

TR2025-019 March 01, 2025

Abstract

We summarize the detailed class-specific performance results in the supplement. The complete image-level AUROC results of 3DSR [55] are provided in Tab. 7, the complete image-level AUROC results of BTF [24] are provided in Tab. 8, and the complete pixel-level AUPRO results of BTF [24] and 3DSR [55] are provided in Tab. 9 and 10 respectively

IEEE Winter Conference on Applications of Computer Vision (WACV) 2025

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Supplementary Material – Towards Zero-shot 3D Anomaly Localization

A. Detailed baseline results

We summarize the detailed class-specific performance results in the supplement. The complete image-level AUROC results of 3DSR [55] are provided in Tab. 7, the complete image-level AUROC results of BTF [24] are provided in Tab. 8, and the complete pixel-level AUPRO results of BTF [24] and 3DSR [55] are provided in Tab. 9 and 10 respectively.

B. Detailed ablation study results

For completeness, we summarize the detailed class-specific performance results of the ablation studies on 3 training classes bagel, carrot and peach in the supplement. The complete pixel-level AUPRO results of 3DzAL without L_{rd}, C_w or input perturbation are provided in Tab. 11; the complete pixel-level AUPRO results of 3DzAL without C_w or input perturbation are provided in Tab. 12; the complete pixel-level AUPRO results of 3DzAL without input perturbation are provided in Tab. 13. In addition, the complete pixel-level AUPRO results of 3DzAL without "removing-point" type pseudo anomalies are provided in Tab. 14 and the complete pixel-level AUPRO results of 3DzAL without "adding-point" type pseudo anomalies are provided in Tab. 15.

To have a more intuitive and explicit view of our pseudo anomaly generation module, we visualize the normal sample patches (positive samples in contrastive learning) and the pseudo abnormal patches (negative samples in contrastive learning) in Fig. 5, where our generated pseudo abnormal 3D point cloud patch is able to mimic unseen anomaly types including contamination, bent, combined, hole, crack *etc*. In contrast, the positive patch samples are relatively smoother and have surface-like shapes. This partially explains why 3DzAL can work under the zero-shot 3D localization setting as 3DzAL aims to learn the relative difference between normal and abnormal 3D point cloud data locally and is not particularly affected by the category prior or class information.

C. Visualization

D. Reproducibility

For all the experimental results reported, the number of algorithm runs used to compute each reported result is 1. In our implementation code, the random seed of the main body part is set as 0 for reproducibility. In 3D pseudo anomaly generation part for-loop, since we need to guarantee the diversity and variety of the generated samples, we can not use the same random seed for random sample selection, so we assign the for-loop index numbers as the random seed. Even if the random seed vary in the for-loops, the sequence of the random seeds is fixed for each run. In this way, the reproducibility of our paper can still be guaranteed.

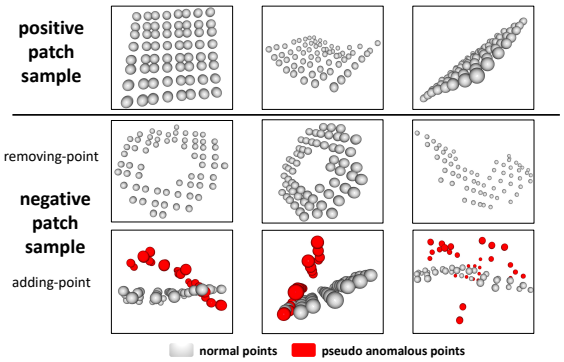


Figure 5. **Visualization of the positive/negative examples.** The positive patch samples are point parts (in gray) selected from the normal training sample and the adding-point-type negative patch samples are generated by attaching the places of interest (in red) from task-irrelevant data. Please zoom in for details.

We search the percentage τ in value set $\{0.01\%, 0.1\%, 1\%\}$ and we search w_{rd} in value set $\{0.1, 1, 10, 100, 1000, 10000\}$. We set $w_d = w_c = 1$. We run experiments on a machine running CentOS Linux 7 (Core) with 503 GiB RAM, Nvidia Tesla V100-SXM2-16GB GPUs and Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz CPUs. The python packages and the corresponding versions are listed here: Bottleneck 1.3.5, certifi 2022.6.15, charset-normalizer 2.1.1, click 8.1.3, cloudpickle 2.0.0, ConfigArgParse 1.5.3, cycler 0.11.0, cytoolz 0.11.0, dask 2021.10.0, docker-pycreds 0.4.0, fonttools 4.37.1, fsspec 2022.3.0, gitdb 4.0.9, GitPython 3.1.27, idna 3.3, imageio 2.9.0, importlib-metadata 4.12.0, joblib 1.1.0, kiwisolver 1.4.4, lmdb 1.3.0, locket 1.0.0, matplotlib 3.5.3, mkl-fft 1.3.1, mkl-random 1.2.2, mkl-service 2.4.0, networkx 2.6.3, numexpr 2.8.3, numpy 1.21.5, opencv-python 4.6.0.66, packaging 21.3, pandas 1.3.5, partd 1.2.0, pathtools 0.1.2, Pillow 9.0.1, pip 21.2.2, promise 2.3, protobuf 3.20.1, psutil 5.9.2, pyparsing 3.0.4, python-dateutil 2.8.2, pytz 2022.1, PyWavelets 1.3.0, PyYAML 6.0, requests 2.28.1, scikit-image 0.19.2, scikit-learn 1.0.2, scipy 1.6.2, sentry-sdk 1.9.8, setproctitle 1.3.2, setuptools 61.2.0, shortuuid 1.0.9, six 1.16.0, smilelogging 0.2.11, smmap 5.0.0, tabulate 0.8.10, threadpoolctl 2.2.0, tifffile 2020.10.1, timm 0.6.7, toolz 0.11.2, torch 1.8.1, torchaudio 0.8.0a0+e4e171a, torchsummaryX 1.3.0, torchvision 0.9.1, tqdm 4.40.0, typing-extensions 4.1.1, urllib3 1.26.12, wandb 0.13.3, wheel 0.37.1, zipp 3.8.1.

E. Computational efficiency

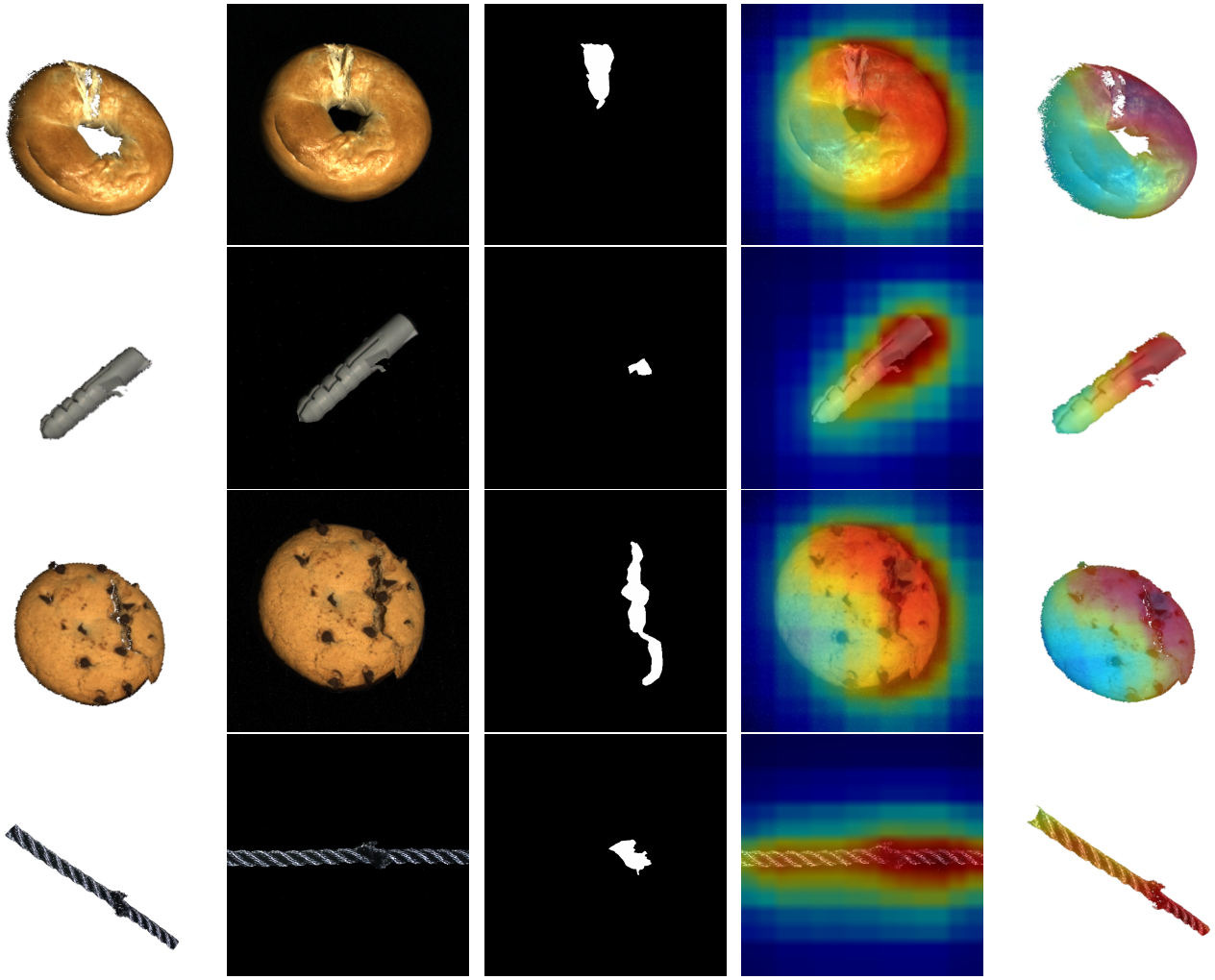
We measure the inference time when testing on the cable gland class. The average inference time of BTF/3DzAL is 212.8/327.8 (s) for 159 test samples. 3DzAL 's inference time is a little larger than that of BTF, mainly due to the additional learnable 3D feature extraction and gradient computing for perturbation. However, we believe that our significant anomaly detection and localization performance gain outweighs the additional inference time.

F. 3D attention overlay of the inductive bias

We provide the 3D inductive bias overlay corresponding to Figure 3 in the main paper. Wer include the visualizations of input 3D point cloud xyz data, the 2D RGB image, the anomaly ground truth map, the inductive bias-based attention 2D visualization and 3D visualization in 5 columns respectively. As illustrated in Fig. 6, we can see more concretely that the inductive bias of a random network is to encompass the areas of interest, which includes the locations identified in the ground truth anomaly maps.

G. Ablation study of the random network architecture for inductive bias generation

We perform additional ablation study of the different network architectures for inductive bias generation. For our main paper and all the reported result of our method, we use ResNet-50 [21] for the random network architecture. We also show the inductive bias visualizations generated using ResNeXt-50-32×4d [51] and Wide-ResNet-50-2 [53]. As indicated in Fig. 7, inductive bias exists across different network architectures for 3D point cloud data and the attention maps typically cover the places of interest despite not being exactly the same. This demonstrates the robustness of our proposed `3DzAL` framework for 3D anomaly detection and localization and further verifies our finding that **an untrained CNN that is initialized randomly has an inherent tendency to identify and locate points of interest on 3D point cloud data, based on its inductive bias.**



(a) 3D input (b) 2D input (c) Ground truth (d) 2D attention (e) 3D attention

Figure 6. Attention overlay visualizations of four testing samples from the classes bagel, dowel, cookie, and rope.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	40.4	47.8	21.6	54.9	47.9	62.3	45.0	51.3	44.1	46.1
cable	41.8	-	39.7	39.0	44.7	38.9	54.2	45.1	76.1	49.9	47.7
carrot	41.3	49.8	-	38.5	53.1	44.6	47.5	44.5	41.3	53.5	46.0
cookie	51.9	48.4	51.2	-	54.7	36.6	54.0	53.2	55.8	49.5	50.6
dowel	33.1	45.6	48.1	46.5	-	54.5	44.2	35.9	66.8	50.9	47.3
foam	50.3	63.5	47.4	67.3	51.0	-	49.3	55.0	83.3	32.9	55.6
peach	39.6	47.9	54.0	30.0	54.3	45.4	-	43.2	58.4	32.2	45.0
potato	48.5	36.9	60.1	38.5	59.1	61.2	51.7	-	48.2	42.9	49.7
rope	53.8	55.4	51.5	32.6	56.1	56.8	59.9	35.6	-	49.0	50.1
tire	50.7	55.2	49.3	48.9	55.4	57.8	42.8	50.4	52.2	-	51.4

Table 7. The detailed image-level AUROC (%) of baseline 3DSR under the zero-shot setting.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	44.7	59.8	67.5	55.7	55.6	46.0	57.9	37.5	55.9	53.4
cable	55.3	-	52.9	54.3	61.9	51.5	50.3	48.5	38.9	37.1	50.1
carrot	51.3	45.3	-	50.6	46.0	57.3	50.5	48.0	59.2	49.2	50.8
cookie	36.6	48.3	49.5	-	53.8	47.2	41.7	48.9	66.5	53.1	49.5
dowel	53.9	49.5	40.0	51.7	-	61.2	47.9	62.8	39.6	54.7	51.3
foam	61.3	46.0	44.6	38.9	54.0	-	47.4	53.8	46.5	56.2	49.9
peach	53.0	51.5	59.0	47.5	54.9	48.4	-	79.7	53.3	53.0	55.6
potato	47.7	50.7	67.2	49.1	53.1	46.0	48.0	-	48.3	58.6	52.1
rope	48.7	48.3	53.1	42.9	41.5	54.5	46.6	44.6	-	61.8	49.1
tire	48.1	54.4	47.7	44.3	53.1	56.3	51.2	50.5	49.2	-	50.5

Table 8. The detailed image-level AUROC (%) of baseline BTF under the zero-shot setting.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	76.1	91.7	88.4	81.8	48.8	90.7	96.4	81.5	69.2	80.5
cable	32.7	-	86.9	46.4	80.9	55.6	65.0	89.1	80.4	77.2	68.2
carrot	44.6	75.2	-	56.7	83.7	35.9	66.7	91.2	82.9	80.7	68.6
cookie	63.2	72.8	90.9	-	80.2	30.6	82.5	94.1	78.5	72.0	73.9
dowel	14.6	71.3	90.0	20.3	-	31.7	46.9	82.1	78.2	81.4	57.4
foam	12.7	72.4	86.1	8.2	78.4	-	52.3	79.5	78.5	81.1	61.0
peach	81.4	78.8	92.5	84.4	82.6	42.1	-	97.9	81.8	68.8	78.9
potato	77.3	75.9	96.6	86.2	82.9	42.3	88.5	-	82.0	72.8	78.3
rope	7.2	66.3	84.1	3.4	79.2	30.7	38.7	75.0	-	83.6	52.0
tire	7.1	69.3	86.3	8.2	79.7	40.6	38.0	75.1	79.5	-	53.8

Table 9. The detailed pixel-level AUPRO (%) of baseline BTF under the zero-shot setting.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	2.3	4.7	10.5	16.1	11.0	0.4	0.0	21.6	4.5	7.9
cable	8.0	-	0.3	28.6	4.8	3.9	9.4	0.0	2.9	1.1	6.6
carrot	4.1	14.4	-	23.5	33.8	1.6	6.6	51.4	58.0	1.9	21.7
cookie	27.2	0.2	1.8	-	16.7	12.1	0.0	0.0	12.0	4.9	8.3
dowel	36.9	71.3	11.9	15.4	-	27.0	41.7	38.6	37.0	38.7	35.4
foam	4.6	0.0	0.0	0.2	0.0	-	0.0	0.0	0.0	0.0	0.5
peach	12.8	40.4	2.2	14.8	12.6	14.0	-	1.0	19.1	11.1	14.2
potato	14.2	1.4	0.7	8.7	6.8	19.7	2.4	-	28.4	36.5	13.2
rope	1.9	50.9	49.5	0.8	27.2	0.8	10.6	12.7	-	19.7	19.3
tire	65.5	21.6	10.2	46.5	6.0	15.9	33.3	7.2	7.6	-	23.8

Table 10. The detailed pixel-level AUPRO (%) of baseline 3DSR under the zero-shot setting.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	77.1	91.9	89.4	81.6	48.3	91.2	96.5	81.8	67.4	80.6
dowel	17.6	71.1	89.8	21.6	-	31.6	46.1	83.6	80.6	77.0	57.7
foam	17.1	73.9	86.0	3.6	79.4	-	54.6	80.2	79.2	83.8	62.0

Table 11. The detailed pixel-level AUPRO (%) of 3DzAL without L_{rd},C_w or input perturbation.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	78.6	92.1	89.1	81.3	48.9	91.2	96.5	82.0	68.1	80.9
dowel	12.6	70.8	89.8	20.9	-	33.0	46.7	83.0	81.6	82.2	57.8
foam	18.1	74.6	86.0	6.6	79.2	-	57.1	79.4	79.4	82.0	62.5

Table 12. The detailed pixel-level AUPRO (%) of 3DzAL without C_w or input perturbation.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	77.2	92.5	89.3	81.6	48.6	91.2	96.5	81.8	85.9	82.7
dowel	16.9	70.6	89.8	20.9	-	46.2	49.3	86.2	82.3	89.1	61.3
foam	18.7	77.0	86.0	6.6	80.6	-	57.0	79.3	80.0	88.8	63.8

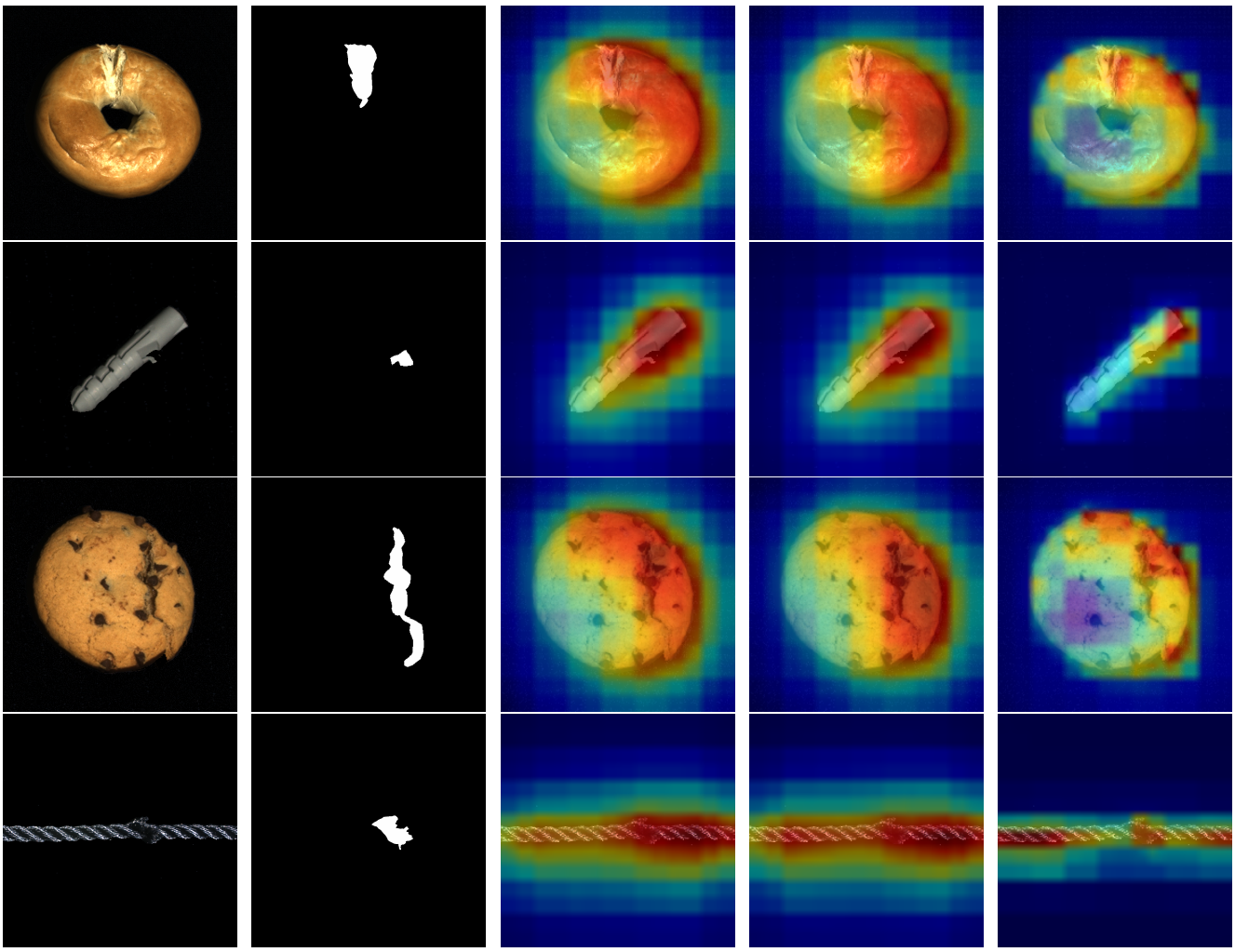
Table 13. The detailed pixel-level AUPRO (%) of 3DzAL without input perturbation.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	76.1	91.6	86.3	81.5	49.3	90.8	96.4	82.0	69.1	80.3
potato	77.5	75.6	96.5	86.2	82.1	42.8	-	87.9	82.8	82.2	79.3
rope	13.4	76.3	87.7	6.0	80.5	46.1	47.5	82.7	-	89.4	58.8

Table 14. The detailed pixel-level AUPRO (%) of 3DzAL with only "adding-point" pseudo anomalies.

train\test	bagel	cable	carrot	cookie	dowel	foam	peach	potato	rope	tire	mean
bagel	-	76.1	89.0	88.6	82.2	49.0	90.7	96.4	81.6	75.4	81.0
potato	77.3	76.4	96.5	86.2	82.8	41.7	-	88.6	82.2	84.3	79.6
rope	13.3	76.2	87.7	6.4	80.5	46.1	47.5	82.8	-	89.4	58.9

Table 15. The detailed pixel-level AUPRO (%) of 3DzAL with only "removing-point" pseudo anomalies.



(a) 2D input (b) Ground Truth (c) ResNet-50 (d) ResNeXt-50 (e) Wide-ResNet-50

Figure 7. Attention overlay visualization comparison of different random network architectures for 3D inductive bias generation.