Supplementary Material

Random Word Data Augmentation with CLIP for Zero-Shot Anomaly Detection

Masato Tamura
masato.tamura@ieee.org

Big Data Analytics Solutions Lab
Hitachi America, Ltd.
2535 Augustine Dr, Santa Clara, California USA

A Detailed Quantitative Results

We report category-wise performance of CLIP [2], WinCLIP [1], and ours for future reference. Our performances are those with the score combination of CLIP and our trained feed-forward network. The reported values of our method are the mean and standard deviation of trials with 10 random seeds.

Table 1: Class-wise AUROC on the MVTec-AD dataset.

	0-shot (Object unknown)		0-shot (Object known)		1-shot		2-shot		4-shot	
Category	CLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours
Bottle	99.4	97.1±1.1	99.2	97.0±0.8	98.2±0.9	97.8±0.6	99.3±0.3	98.1±0.7	99.3±0.4	98.2±0.6
Cable	83.0	84.2±0.6	86.5	86.4±1.4	88.9±1.9	90.8±1.1	88.4±0.7	91.3±2.0	90.9±0.9	92.1±0.8
Capsule	85.3	87.3±2.0	72.9	88.7±2.8	72.3±6.8	76.8±6.4	77.3±8.8	83.6±7.1	82.3±8.9	87.8±7.9
Carpet	100	100±0.0	100	100±0.0	99.8±0.3	100±0.0	99.8±0.3	100±0.0	100±0.0	100±0.0
Grid	99.2	99.0±0.4	98.8	98.7±0.2	99.5±0.3	99.2±0.5	99.4±0.2	99.1±0.4	99.6±0.1	99.6±0.2
Hazelnut	92.0	94.3±1.1	93.9	94.0±0.5	97.5±1.4	94.9±0.6	98.3±0.7	95.0±0.5	98.4±0.4	95.1±0.5
Leather	100	100±0.0	100	100±0.0	99.9±0.0	100±0.0	99.9±0.0	99.9±0.0	100±0.0	100±0.0
Metal nut	94.4	96.1±1.3	97.1	94.7±1.7	98.7±0.8	96.1±1.3	99.4±0.2	96.1±1.4	99.5±0.2	96.3±1.3
Pill	88.3	88.6±1.6	79.1	90.2±1.0	91.2±2.1	92.5±1.3	92.3±0.7	92.6±1.2	92.8±1.0	92.6±1.3
Screw	76.1	76.6±1.5	83.3	75.5±2.1	86.4±0.9	77.4±1.8	86.0±2.1	77.5±2.1	87.9±1.2	77.6±2.1
Tile	99.4	99.5±0.2	100	99.5±0.2	99.9±0.0	99.6±0.1	99.9±0.2	99.6±0.1	99.9±0.1	99.6±0.1
Toothbrush	92.8	88.4±3.5	87.5	94.0±0.9	92.2±4.9	94.7±1.3	97.5±1.6	95.0±1.0	96.7±2.6	96.4±1.5
Transistor	79.7	86.6±1.4	88.0	88.9±1.3	83.4±3.8	89.4±1.3	85.3±1.7	89.6±1.4	85.7±2.5	89.6±1.3
Wood	97.8	97.6±0.6	99.4	98.1±0.5	99.9±0.1	98.7±0.6	99.9±0.1	98.8±0.5	99.8±0.3	98.8±0.6
Zipper	85.5	87.4±1.0	91.5	88.8±0.1	88.8±5.9	91.6±2.9	94.0±1.4	93.8±1.0	94.5±0.5	94.3±1.0

References

[1] Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. WinCLIP: Zero-/few-shot anomaly classification and segmentation. In *CVPR*, 2023.

Table 2: Class-wise AUPR on the MVTec-AD dataset.

	0-shot (Object unknown)		0-shot (Object known)		1-shot		2-shot		4-shot	
Category	CLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours
Bottle	99.8	99.2±0.3	99.8	99.2±0.2	99.4±0.3	99.4±0.2	99.8±0.1	99.4±0.2	99.8±0.1	99.5±0.2
Cable	88.3	89.3±0.5	91.2	91.6±1.0	93.2±1.1	94.6±0.8	92.9±0.6	94.9±1.2	94.4±0.3	95.4±0.5
Capsule	96.8	96.8±0.7	91.5	97.0±0.9	91.6±2.7	93.4±2.2	93.3±3.6	95.6±2.2	95.1±3.3	96.6±2.6
Carpet	100	100±0.0	100	100±0.0	99.9±0.1	100±0.0	99.9±0.1	100±0.0	100±0.0	100±0.0
Grid	99.7	99.7±0.1	99.6	99.6±0.0	99.9±0.1	99.7±0.2	99.8±0.1	99.7±0.1	99.9±0.0	99.9±0.1
Hazelnut	96.0	96.7±0.6	96.9	96.2±0.4	98.6±0.7	96.8±0.5	99.1±0.4	96.8±0.4	99.1±0.2	96.9±0.5
Leather	100	100±0.0	100	100±0.0	100±0.0	100±0.0	100±0.0	100±0.0	100±0.0	100±0.0
Metal nut	98.8	99.1±0.3	99.3	98.8±0.4	99.7±0.2	99.1±0.3	99.9±0.0	99.1±0.3	99.9±0.1	99.2±0.3
Pill	97.5	97.6±0.4	95.7	98.0±0.2	98.3±0.5	98.5±0.3	98.6±0.1	98.5±0.3	98.6±0.2	98.5±0.3
Screw	90.3	90.7±0.8	93.1	90.1±1.2	94.2±0.6	91.0±1.1	94.1±1.5	91.1±1.1	94.9±0.8	91.1±1.1
Tile	99.8	99.8±0.1	100	99.8±0.1	100±0.0	99.8±0.1	100±0.1	99.8±0.0	100±0.0	99.8±0.1
Toothbrush	97.7	95.4±1.7	95.6	97.6±0.4	96.7±2.0	97.9±0.6	99.0±0.6	98.0±0.4	98.7±1.1	98.6±0.6
Transistor	76.5	80.4±1.7	87.1	82.3±2.1	79.0±4.0	83.3±2.1	80.7±2.3	83.5±2.2	80.7±3.2	83.6±2.1
Wood	99.3	99.3±0.2	99.8	99.4±0.2	100±0.0	99.6±0.2	100±0.0	99.6±0.2	99.9±0.1	99.6±0.2
Zipper	95.6	96.1±0.3	97.5	96.7±0.0	96.8±1.8	97.5±0.9	98.3±0.4	98.1±0.4	98.5±0.2	98.3±0.3

Table 3: Class-wise F_1-max on the MVTec-AD dataset.

	0-shot (Object unknown)		0-shot (Object known)		1-shot		2-shot		4-shot	
Category	CLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours
Bottle	98.4	95.2±1.1	97.6	95.9±0.7	96.5±1.3	96.5±0.6	97.7±0.7	96.6±0.7	97.8±0.6	96.6±0.7
Cable	84.5	85.7±0.8	84.5	84.4±1.2	86.1±1.3	87.7±1.0	85.2±0.7	87.6±2.0	87.2±0.6	87.8±1.0
Capsule	90.7	92.6±0.8	91.4	93.3±0.8	91.6±0.7	93.1±0.7	92.1±0.7	94.0±0.8	92.5±0.5	94.7±0.8
Carpet	99.4	99.6±0.4	99.4	99.6±0.4	99.2±0.8	100±0.0	99.3±0.7	100±0.0	99.9±0.2	100±0.0
Grid	97.3	97.8±0.6	98.2	97.3±0.0	98.9±0.4	98.1±0.7	99.1±0.0	98.1±0.5	99.1±0.0	98.4±0.4
Hazelnut	88.9	91.9±1.7	89.7	91.8±1.0	94.7±2.3	92.7±0.9	95.6±1.6	92.9±0.8	96.2±1.0	93.0±0.8
Leather	100	100±0.0	100	100±0.0	99.5±0.0	99.7±0.3	99.7±0.2	99.5±0.3	99.8±0.2	99.9±0.2
Metal nut	93.3	95.0±1.1	96.3	94.0±1.2	97.7±1.0	95.0±1.0	98.4±0.5	94.9±1.2	98.5±0.6	95.1±1.0
Pill	93.7	93.4±0.6	91.6	93.9±0.3	93.8±0.7	95.1±0.7	94.3±0.4	95.2±0.7	94.1±0.4	95.1±0.8
Screw	86.9	86.9±0.0	87.4	86.9±0.2	88.5±0.3	86.9±0.3	89.0±0.6	87.0±0.4	89.6±0.7	87.0±0.4
Tile	97.7	97.8±0.6	99.4	97.8±0.6	98.9±0.2	97.9±0.5	99.2±0.3	97.9±0.5	99.2±0.3	98.0±0.5
Toothbrush	93.1	90.2±0.9	87.9	92.7±0.7	94.1±1.9	94.1±1.1	96.7±1.8	93.7±0.9	96.8±2.3	95.3±1.6
Transistor	69.7	79.6±1.8	79.5	81.7±1.5	75.1±3.1	81.6±1.7	75.9±2.4	81.7±1.7	76.6±2.8	81.8±1.5
Wood	95.7	95.8±0.6	98.3	96.0±0.4	99.4±0.3	97.2±0.6	99.5±0.4	97.2±0.6	99.2±0.9	97.1±0.7
Zipper	90.0	91.2±0.8	92.9	91.9±0.1	92.1±2.5	93.7±2.1	94.4±0.3	95.5±0.3	94.7±0.4	95.5±0.3

[2] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

Table 4: Class-wise AUROC on the VisA dataset.

	0-shot (Object unknown)		0-shot (Object known)		1-shot		2-shot		4-shot	
Category	CLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours
Candle	96.0	96.9±0.3	95.4	97.0±0.1	93.4±1.4	97.7±0.2	94.8±1.0	97.9±0.1	95.1±0.3	97.9±0.2
Capsules	74.9	78.1±2.3	85.0	75.6±0.3	85.0±3.1	82.1±0.6	84.9±0.8	82.5±0.6	86.8±1.7	82.6±0.6
Cashew	87.3	91.6±2.4	92.1	91.8±1.0	94.0±0.4	92.9±1.1	94.3±0.5	93.1±0.8	95.2±0.8	93.1±0.8
Chewing gum	89.8	94.3±1.3	96.5	93.0±1.9	97.6±0.8	95.0±0.7	97.3±0.8	95.1±0.7	97.7±0.3	95.3±0.6
Fryum	88.2	90.0±0.5	80.3	89.2±0.5	88.5±1.9	92.6±1.2	90.5±0.4	92.8±1.3	90.8±0.5	92.9±1.2
Macaroni1	82.0	82.5±2.0	76.2	89.4±1.4	82.9±1.5	91.5±1.7	83.3±1.9	91.5±1.7	85.2±0.9	91.6±1.6
Macaroni2	65.6	71.3±1.5	63.7	68.6±1.5	70.2±0.9	69.8±1.5	71.8±2.0	69.9±1.4	70.9±2.2	69.9±1.3
PCB1	58.2	49.7±7.0	73.6	69.4±3.8	75.6±23.0	70.7±16.9	76.7±5.2	80.2±10.4	88.3±1.7	83.1±2.5
PCB2	51.6	50.7±1.0	51.2	47.3±0.1	62.2±3.9	60.7±2.2	62.6±3.7	63.8±1.9	67.5±2.6	66.4±1.5
PCB3	66.0	67.9±1.3	73.4	66.3±0.4	74.1±1.1	74.0±8.4	78.8±1.9	80.3±3.4	83.3±1.7	83.4±1.8
PCB4	74.5	75.2±0.7	79.6	82.9±0.4	85.2±8.9	88.1±9.0	82.3±9.9	87.6±10.4	87.6±8.0	89.7±10.1
Pipe fryum	84.1	89.9±1.4	69.7	87.2±2.6	97.2±1.1	92.6±1.7	98.0±0.6	93.0±1.8	98.5±0.4	93.2±1.8

Table 5: Class-wise AUPR on the VisA dataset.

	0-shot (Object unknown)		0-shot (Object known)		1-shot		2-shot		4-shot	
Category	CLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours
Candle	96.3	96.9±0.3	95.8	96.9±0.2	93.6±1.5	97.6±0.2	95.1±1.1	97.7±0.1	95.3±0.4	97.8±0.1
Capsules	83.1	85.4±1.5	90.9	84.8±0.1	89.9±2.5	88.6±0.4	88.9±0.7	88.9±0.3	91.5±1.4	89.0±0.3
Cashew	94.3	96.2±1.0	96.4	96.3±0.4	97.2±0.2	96.9±0.4	97.3±0.2	96.9±0.3	97.7±0.4	96.9±0.3
Chewing gum	95.6	97.7±0.6	98.6	97.0±0.9	99.0±0.3	97.9±0.3	98.9±0.3	98.0±0.3	99.0±0.1	98.0±0.2
Fryum	94.8	95.7±0.2	90.1	95.4±0.3	94.7±1.0	97.0±0.5	95.8±0.2	97.0±0.6	96.0±0.3	97.1±0.6
Macaroni1	85.0	84.5±1.7	75.8	90.1±1.1	84.9±1.2	92.0±1.4	84.7±1.5	92.0±1.4	86.5±0.6	92.1±1.3
Macaroni2	60.8	67.4±1.6	60.3	66.3±1.9	68.4±1.8	68.5±1.6	70.4±1.8	68.9±1.7	69.6±2.8	68.9±1.7
PCB1	63.6	55.3±6.3	78.4	71.9±4.0	76.5±19.0	72.4±13.8	78.3±4.3	80.0±8.8	87.7±1.7	82.4±2.6
PCB2	54.9	54.0±0.7	49.2	49.6±0.1	64.9±3.3	64.7±2.1	65.8±4.0	68.0±1.5	71.3±3.4	70.1±0.7
PCB3	66.4	69.6±1.3	76.5	67.3±0.3	73.5±1.6	74.6±8.7	80.9±1.6	82.5±3.0	84.8±1.8	85.4±1.4
PCB4	79.6	79.9±0.5	77.7	84.4±0.4	78.5±15.5	83.2±13.7	72.5±16.2	83.6±14.4	85.6±8.9	86.0±14.6
Pipe fryum	91.6	95.1±0.8	82.3	93.6±2.1	98.6±0.5	96.2±1.1	99.0±0.3	96.4±1.1	99.2±0.2	96.5±1.1

Table 6: Class-wise F_1-max on the VisA dataset.

	0-shot (Object unknown)		0-shot (Object known)		1-shot		2-shot		4-shot	
Category	CLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours	WinCLIP	CLIP + ours
Candle	91.1	92.4±0.6	89.4	92.2±0.3	87.8±1.2	94.7±0.5	89.1±1.3	95.2±0.4	88.9±1.0	95.1±0.3
Capsules	79.8	81.4±1.3	83.9	79.3±0.2	84.9±2.0	82.8±0.6	85.4±0.6	83.0±0.6	86.0±0.9	83.1±0.5
Cashew	84.3	89.0±2.3	88.4	89.1±1.2	90.7±0.7	89.6±1.3	90.9±0.7	89.8±1.0	91.6±1.3	90.0±1.0
Chewing gum	87.4	92.9±1.7	94.8	90.2±2.4	95.6±0.9	92.7±1.0	95.4±0.6	93.0±0.9	95.7±0.5	93.4±0.6
Fryum	86.2	87.6±0.8	82.7	86.9±0.9	87.2±1.4	91.0±2.0	88.4±0.6	91.0±1.9	88.9±0.8	91.4±1.9
Macaroni1	74.4	76.7±1.1	74.2	83.1±1.2	76.2±1.4	84.9±1.5	76.7±2.0	84.8±1.5	78.2±1.2	84.9±1.6
Macaroni2	71.9	71.6±1.1	69.8	69.4±0.7	72.3±1.1	70.1±0.9	73.9±0.9	69.9±0.7	73.1±1.6	70.0±0.7
PCB1	66.7	66.7±0.0	71.0	69.2±1.1	81.3±6.6	75.6±6.9	73.2±3.7	78.7±4.8	83.1±2.2	78.5±2.3
PCB2	67.1	67.2±0.2	67.1	67.1±0.1	67.2±0.3	68.1±0.8	67.3±0.3	68.5±0.5	67.7±0.6	69.3±0.8
PCB3	70.5	69.9±0.9	71.0	68.9±0.3	73.5±1.5	72.7±3.4	73.9±1.3	75.3±2.0	77.0±1.4	76.9±1.8
PCB4	73.3	74.3±0.6	74.9	77.0±0.5	86.1±2.1	89.6±3.9	86.8±3.8	88.4±6.6	84.6±7.0	89.6±5.8
Pipe fryum	85.1	88.6±1.6	80.7	86.9±1.3	94.4±0.7	90.9±1.8	95.4±0.8	91.3±1.6	95.6±0.7	91.5±1.4