Supplementary Material for *Unilaterally Aggregated Contrastive Learning with Hierarchical Augmentation for Anomaly Detection*

Guodong Wang[1,2], Yunhong Wang[2], Jie Qin[3], Dongming Zhang[4], Xiuguo Bao[4], Di Huang[1,2*]
[1]State Key Laboratory of Software Development Environment, Beihang University, Beijing, China
[2]School of Computer Science and Engineering, Beihang University, Beijing, China
[3]College of Computer Science and Technology, NUAA, Nanjing, China
[4]Natl. Comp. Net. Emer. Resp. Tech. Team/Coord. Ctr. of China, Beijing, China
{wanggd,yhwang,dhuang}@buaa.edu.cn, qinjiebuaa@gmail.com, zhdm@cert.org.cn, baoxiuguo@139.com

In this supplementary material, we provide more experimental details and more experimental results (*i.e.* per-class performance) on one-class CIFAR-100 (20 super-classes) and ImageNet-30, as well as more analysis on distribution-shifting/identity-preserving augmentations.

1. Experimental Details

Hierarchical Augmentation. We employ HA along the network to further prompt a higher concentration of inliers, in which deeper residual stages address stronger data augmentations. Following [2], we use the combination of random resized crop, color jittering, gray-scale and horizontal flip with increasing augmentation strengths for T_i $(i = 1, 2, 3, 4)$ to generate positive views. Table A shows the detailed augmentation configurations.

Table A: Augmentation configurations for T_i $(i = 1, 2, 3, 4)$. RRC, CJ, GS, HF are short for random resized crop, color jittering, gray-scale and horizontal flip, respectively. $RRC(i, j)$ specifies the range of the cropped area and $CJ(b, c, s, h)$ specifies the range of brightness, contrast, saturation and hue.

T_1	$RRC(0.75, 1), CJ(0.1, 0.1, 0.1, 0.025), GS, HF$
T_2	$RRC(0.54, 1), CJ(0.2, 0.2, 0.2, 0.050), GS, HF$
T_3	$RRC(0.30, 1), CJ(0.3, 0.3, 0.3, 0.075), GS, HF$
T_4	$RRC(0.08, 1), CJ(0.4, 0.4, 0.4, 0.100), GS, HF$

An extra projection head g_i is additionally attached at the end of res_i to down-sample and project the feature maps with the same shape as in the last stage res_4. Similar to [10], each g_i consists of a series of down-sampling blocks and projection blocks. Table E shows the detailed network structure.

*Corresponding author.

Soft Aggregation. In Fig. 1, we display two rows of the augmented views of inliers induced by standard data augmentation $\mathcal{T}$ as in CSI [8]. Notably, some views capture the main body of planes, whereas others are distracted by the background. It indicates that the generated views probably suffer from the semantic shift, and imposing such noisy inliers to be close reduces the purity of the inlier distribution.

Outlier Exposure (OE) [4]. OE leverages an auxiliary dataset as outliers and enables anomaly detectors to generalize well to unseen anomalies. In this paper, we investigate the 80 Million Tiny Images dataset [9] as the OE dataset with images from CIFAR-10 removed to make sure that the OE dataset and CIFAR-10 are disjoint. In practice, we use 300K random images[1] and observe that only a small fraction of this dataset is sufficiently effective for AD. Meanwhile, from Table B, we observe the increasing performance with more outliers exposed. Additionally, in the case of no OE applied, we vary $|\mathcal{D}_{\text{out}}|$ by randomly keeping some inliers not being rotated. Tab. B shows that we can benefit more from a larger size of $\mathcal{D}_{\text{out}}$.

Table B: Ablation w.r.t. OE and $|\mathcal{D}_{\text{out}}|$ ratios on CIFAR-10.

	0%	25%	50%	75%	100%
$\|\mathcal{D}_{\text{out}}\|$	/	91.3	92.6	94.5	95.4
OE	95.4	95.9	96.2	96.6	**96.9**

2. Per-class Results on One-class Settings

Tables C and D present the AD results of our UniCon-HA on one-class CIFAR-100 (20 super-classes) and ImageNet-30, respectively. Clearly, our method outperforms the other state-of-the-art methods [8, 3, 5, 1, 7], which also utilize transformations to create virtual outliers on most

[1]https://github.com/hendrycks/outlier-exposure

Table C: Per-class AUROC scores on one-class CIFAR-100 (20 super-classes). Numbers in the first column indicate the super-class IDs. * denotes the results directly adopted from [8] and bold numbers denote the best results.

	OC-SVM* [6]	Geom [3]	Rot*[5]	Rot+Trans*[5]	GOAD*[1]	DROC [7]	CSI [8]	**UniCon-HA** (Ours)
0	68.4	74.7	78.6	79.6	73.9	82.9	86.3	**89.8**
1	63.6	68.5	73.4	73.3	69.2	84.3	84.8	**90.2**
2	52.0	74.0	70.1	71.3	67.6	88.6	88.9	**94.4**
3	64.7	81.0	68.6	73.9	71.8	86.4	85.7	**89.5**
4	58.2	78.4	78.7	79.7	72.7	92.6	93.7	**96.3**
5	54.9	59.1	69.7	72.6	67.0	84.5	81.9	**87.6**
6	57.2	81.8	78.8	85.1	80.0	73.4	91.8	**93.0**
7	62.9	65.0	62.5	66.8	59.1	84.2	83.9	**87.8**
8	65.6	85.5	84.2	86.0	79.5	87.7	91.6	**94.0**
9	74.1	90.6	86.3	87.3	83.7	94.1	95.0	**97.1**
10	84.1	87.6	87.1	88.6	84.0	85.2	**94.0**	92.2
11	58.0	83.9	76.2	77.1	68.7	87.8	90.1	**90.5**
12	68.5	83.2	83.3	84.6	75.1	82.0	90.3	**93.4**
13	64.6	58.0	60.7	62.1	56.6	82.7	81.5	**86.9**
14	51.2	92.1	87.1	88.0	83.8	93.4	94.4	**97.2**
15	62.8	68.3	69.0	71.9	66.9	75.8	**85.6**	84.2
16	66.6	73.5	71.7	75.6	67.5	80.3	83.0	**90.8**
17	73.7	93.8	92.2	93.5	91.6	97.5	97.5	**98.1**
18	52.8	90.7	90.4	91.5	88.0	94.4	95.9	**98.0**
19	58.4	85.0	86.5	88.1	82.6	92.4	95.2	**96.7**
Mean	63.1	78.7	77.7	79.8	74.5	86.5	89.6	**92.4**

Table D: Per-class AUROC scores on one-class ImageNet-30. Numbers in the first and fourth rows indicate the class IDs. Bold numbers denote the best results.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
CSI [8]	85.9	**99.0**	**99.8**	90.5	95.8	99.2	96.6	83.5	92.2	84.3	**99.0**	94.5	97.1	87.7	96.4
UniCon-HA	**87.3**	98.7	**99.8**	**93.1**	**96.4**	**99.3**	**97.5**	**88.4**	**94.3**	**89.2**	98.9	**95.3**	**97.4**	**90.0**	**96.7**
	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
CSI [8]	84.7	**99.7**	75.6	95.2	73.8	**94.7**	95.2	**99.2**	98.5	82.5	**89.7**	82.1	97.2	82.1	97.6
UniCon-HA	**85.8**	99.5	**83.9**	**95.3**	**79.8**	94.5	**95.4**	98.8	**98.7**	**84.8**	89.2	**87.1**	**97.4**	**86.8**	**97.9**

Figure 1: Illustration of augmented samples for the Plane class in CIFAR-10. Figures are from the same mini-batch during training and ranked according to the descent order of their ω_x. Our soft mechanism enables us to identify the most likely inliers while suppress the potential outliers for a purified inlier concentration.

Table E: The structure of the projection head g_i.

<table>
<tr><th>#</th><th colspan="2">Down-sampling blocks</th><th>Projection blocks</th></tr>
<tr><td rowspan="2">g_1</td><td>SepConv</td><td>Conv, Conv, BN, ReLU
Conv, Conv, BN, ReLU</td><td rowspan="2">Linear, ReLU, Linear</td></tr>
<tr><td colspan="2">SepConv
SepConv
Conv, BN, ReLU, AvgPool</td></tr>
<tr><td>g_2</td><td colspan="2">SepConv
SepConv
Conv, BN, ReLU, AvgPool</td><td>Linear, ReLU, Linear</td></tr>
<tr><td>g_3</td><td colspan="2">SepConv
Conv, BN, ReLU, AvgPool</td><td>Linear, ReLU, Linear</td></tr>
<tr><td>g_4</td><td colspan="2">AvgPool</td><td>Linear, ReLU, Linear</td></tr>
</table>

classes.

Though sharing the same spirit of creating virtual outliers, we develop a completely different way of exploiting those outliers. Recall that a good representation distribution for AD requires: (a) a compact distribution for inliers and

(b) a dispersive distribution for (virtual) outliers. Both the requirements are only partially considered in the previous literature [8, 3, 5, 1, 7] with sub-optimal results obtained, while we explicitly encourage the concentration of inliers and the dispersion of outliers as our training objective. Interestingly, our method is free from any auxiliary branches to differentiate the specific types of transformations, outside of the commonly adopted transformation (*e.g.* rotation) prediction based on a classifier for AD.

3. Analysis on Augmentations

Following CSI [8], we try to remove or convert-to-shift identity-preserving augmentations $\mathcal{T}$, including random resized crop, color jittering, horizontal flip and gray-scale. Table F confirms the observations from CSI: (1) treating $\mathcal{T}$ as distribution-shifting augmentations leads to a sub-optimal solution as these augmentations shift the original distribution less than rotation does, increasing false negative samples; (2) removing any augmentations from $\mathcal{T}$ degrades performance, showing the importance of identity-preserving augmentations to generating diverse positive views, where random crop is the most influential.

Table F: Ablation study w.r.t. augmentations on CIFAR-10.

	Base		Crop	Color	Flip	Gray
CSI [8]	94.3	+shift	85.4	87.3	86.2	88.7
		-remove	88.0	90.2	93.6	93.7
Ours	**95.4**	+shift	84.6	90.4	87.4	92.0
		-remove	90.8	91.5	94.2	94.9

References

[1] Liron Bergman and Yedid Hoshen. Classification-based anomaly detection for general data. In *ICLR*, 2020. 1, 2, 3

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020. 1

[3] Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. In *NeurIPS*, 2018. 1, 2, 3

[4] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *ICLR*, 2019. 1

[5] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. In *NeurIPS*, 2019. 1, 2, 3

[6] Bernhard Schölkopf, Robert C Williamson, Alex Smola, John Shawe-Taylor, and John Platt. Support vector method for novelty detection. In *NeurIPS*, 2000. 2

[7] Kihyuk Sohn, Chun-Liang Li, Jinsung Yoon, Minho Jin, and Tomas Pfister. Learning and evaluating representations for deep one-class classification. In *ICLR*, 2020. 1, 2, 3

[8] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. In *NeurIPS*, 2020. 1, 2, 3

[9] Antonio Torralba, Rob Fergus, and William T Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE TPAMI*, 30(11):1958–1970, 2008. 1

[10] Junbo Zhang and Kaisheng Ma. Rethinking the augmentation module in contrastive learning: Learning hierarchical augmentation invariance with expanded views. In *CVPR*, 2022. 1