RankOOD - Class Ranking-based Out-of-Distribution Detection

Supplementary Material

A. Introduction

In this supplementary material, we present additional details and results that were excluded from the main content due to space limitations.

A.1. Comparison with baselines

In the main text (Tab. 2 and Tab. 3), we report results only for baselines that placed in the top five in at least one setting. For completeness, we provide the full results for all 35 baselines in Tab. 2 (near-OOD) and Tab. 3 (far-OOD), with key observations summarized in the respective captions.

A.2. Detailed per-dataset OOD detection results

In Sec. 5.1, we reported average AUROC and FPR95 across all OOD datasets for each ID dataset (Tab. 2 and Tab. 3). For completeness, we provide the per-dataset results for each OOD benchmark. TinyImageNet results in Tab. 4 and Tab. 5; CIFAR-100 results in Tab. 6 and Tab. 7; and CIFAR-10 results are shown in Tab. 8 and Tab. 9.

A.3. Effectiveness of RankOOD-T

Fig. 1 shows standard logit-based OOD scores (MSP, MaxLogit, EBO (Energy based), and GEN) under RankOOD-T. It can be seen that RankOOD-T performs better across all scoring methods, including RankOOD-S defined in Eq. 6, compared to cross-entropy (CE) training. *This indicates that the observed performance gains primarily stem from the RankOOD-T objective rather than the RankOOD-S function itself.* In particular, ListMLE training improve s AUROC by at least 1.5% and reduces FPR95 by over 7% compared to CE training. Nonetheless, as shown in Fig 1, logit-based OOD scores can be used in open-world settings to obtain on-par performance after RankOOD-T.

A.4. Scalability and Computational Cost

GPU Time: Tab. 1 reports per-epoch GPU training time (seconds). As the number of classes increases, all methods exhibit comparable training costs. RankOOD incurs approximately 34% higher GPU time than LogitNorm in TinyImageNet due to the ListMLE objective, which requires computing probabilities over full permutations. *We clarify that RankOOD-T follows the same pre-training paradigm as LogitNorm with a total of 300/500 epochs and doesn't need extra fine-tuning epochs like CRAFT as canonical classes can be derived from a SOTA pre-trained model.*

Table 1. Per-epoch GPU time (sec.) and per-class ILP runtime (sec.) across different datasets (C-CIFAR, IN-ImageNet).

Method	**C-10**	**C-100**	**IN-200**	**IN-1k**	**Epochs**
LogitNorm	8.33	5.32	1045.50	6614.0	~ 200
CRAFT	17.45	25.44	1354.00	-	~ 100
RankOOD	13.80	14.34	1402.12	6855.7	$\sim 300-500$
ILP Runtime	0.0058	0.1718	2.4933	1413.6	-
ILP - Greedy	0.0006	0.0023	0.0177	0.2147	-

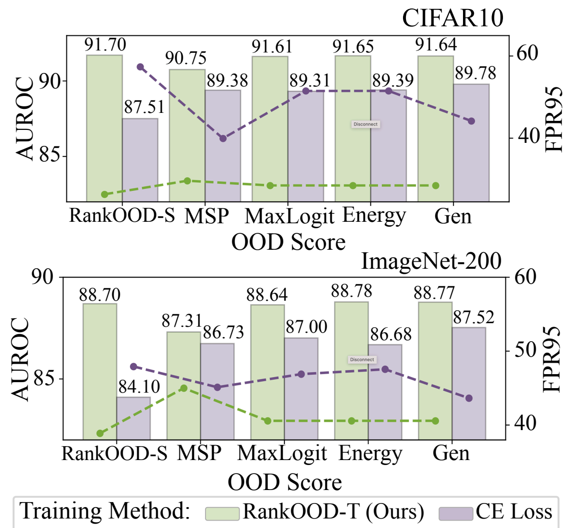

Figure 1. Avg. AUROC (bars, left y-axis) and FPR95 (lines, right y-axis) across multiple OOD datasets for other scoring functions.

ILP Runtime: Tab. 1 reports ILP rank execution time. ILP time grows exponentially with the number of ranks and classes, with worst-case complexity $O(2^{C\times K})$ where C-classes and K-ranks. However, this can be addressed using a greedy approach with complexity $O(CK.log(K))$, significantly reducing computational cost.

A.5. RankOOD Score Example

We provide a step-by-step examples of RankOOD-S computation in Fig. 2.

A.6. Conditional Probability Matrices

In Sec. 5.3, we reported class-conditional probability (CP) matrices for only five CIFAR-10 classes. Here, we provide the complete CP matrices for all ten classes under both CE and RankOOD training. For RankOOD, we additionally report CP matrices for all OOD datasets as well as the ID dataset (CIFAR-10). As shown in Fig. 3, each matrix reflects the model's ability to preserve the class-wise canonical rank order. Under conventional CE training, a sample predicted as Airplane achieves a probability of 0.29 for correctly identifying the rank-4 label, conditioned on all preceding ranks (rank-1 through rank-3) being accurately

Table 2. Performance comparison in *near-OOD detection*. For each column, the top five methods are marked in **bold**. Note that N/A indicates missing results in OpenOOD. *OE obtains the lowest average FPR95 (34.29), while RankOOD ranks second with 44.79. However, OE's performance is known to be biased toward seen outliers used during training, making its evaluation less reliable [13]. Among methods that do not rely on outliers, RankOOD achieves the best performance for both AUROC and FPR95. Notably, RankOOD achieves state-of-the-art performance on TinyImageNet near-OOD, reducing FPR95 by 4.3% relative to the strongest baseline OE.*

Method	CIFAR-10		CIFAR-100		TinyImageNet		Average	
	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓
Post-hoc inference methods								
OpenMax [1]	87.62 ± 0.29	43.62 ± 2.27	76.41 ± 0.25	56.58 ± 0.73	80.27 ± 0.10	63.48 ± 0.25	81.43	54.56
MSP [6]	88.03 ± 0.25	48.17 ± 3.92	80.27 ± 0.11	**54.80 ± 0.33**	83.34 ± 0.06	54.82 ± 0.35	83.88	52.60
TempScale [5]	88.09 ± 0.31	50.96 ± 4.32	80.90 ± 0.07	**54.49 ± 0.48**	**83.69 ± 0.04**	54.82 ± 0.23	84.23	53.42
ODIN [16]	82.87 ± 1.85	76.19 ± 6.08	79.90 ± 0.11	57.91 ± 0.51	80.27 ± 0.08	66.76 ± 0.26	81.01	66.95
MDS [15]	84.20 ± 2.40	49.90 ± 3.98	58.69 ± 0.09	83.53 ± 0.60	61.93 ± 0.51	79.11 ± 0.31	68.27	70.85
MDSEns [15]	60.43 ± 0.26	92.26 ± 0.20	46.31 ± 0.24	95.88 ± 0.04	54.32 ± 0.24	91.75 ± 0.10	53.69	93.30
RMDS [20]	89.80 ± 0.28	38.89 ± 2.39	80.15 ± 0.11	55.46 ± 0.41	82.57 ± 0.25	**54.02 ± 0.58**	84.17	**49.46**
Gram [21]	58.66 ± 4.83	90.87 ± 1.91	51.66 ± 0.77	92.28 ± 0.29	67.67 ± 1.07	86.40 ± 1.21	59.33	89.85
EBO [17]	87.58 ± 0.46	61.34 ± 4.63	**80.91 ± 0.08**	55.62 ± 0.61	82.50 ± 0.05	60.24 ± 0.57	83.66	59.07
OpenGAN [14]	53.71 ± 7.68	94.48 ± 4.01	65.98 ± 1.26	76.52 ± 2.59	59.79 ± 3.39	84.15 ± 3.85	59.83	85.05
GradNorm [10]	54.90 ± 0.98	94.72 ± 0.82	70.13 ± 0.47	85.58 ± 0.46	72.75 ± 0.48	82.67 ± 0.30	65.93	87.66
ReAct [24]	87.11 ± 0.61	63.56 ± 7.33	80.77 ± 0.05	56.39 ± 0.34	81.87 ± 0.98	62.49 ± 2.19	83.25	60.81
MLS [8]	87.52 ± 0.47	61.32 ± 4.62	**81.05 ± 0.07**	55.47 ± 0.66	82.90 ± 0.04	59.76 ± 0.59	83.82	58.85
KLM [8]	79.19 ± 0.80	87.86 ± 6.37	76.56 ± 0.25	77.92 ± 1.31	80.76 ± 0.08	70.26 ± 0.64	78.84	78.68
VIM [27]	88.68 ± 0.28	44.84 ± 2.31	74.98 ± 0.13	62.63 ± 0.27	78.68 ± 0.24	59.19 ± 0.71	80.78	55.55
KNN [23]	90.64 ± 0.20	34.01 ± 0.38	80.18 ± 0.15	61.22 ± 0.14	81.57 ± 0.17	60.18 ± 0.52	84.13	51.80
DICE [25]	78.34 ± 0.79	70.04 ± 7.64	79.38 ± 0.23	57.95 ± 0.53	81.78 ± 0.14	61.88 ± 0.67	79.83	63.29
RankFeat [22]	79.46 ± 2.52	60.88 ± 4.60	61.88 ± 1.28	80.59 ± 1.10	56.92 ± 1.59	92.06 ± 0.23	66.09	77.84
ASH [4]	75.27 ± 1.04	86.78 ± 1.82	78.20 ± 0.15	65.71 ± 0.24	82.38 ± 0.19	64.89 ± 0.90	78.62	72.46
SHE [31]	81.54 ± 0.51	79.65 ± 3.47	78.95 ± 0.18	59.07 ± 0.25	80.18 ± 0.25	66.80 ± 0.74	80.22	68.51
GEN [18]	88.20 ± 0.30	53.67 ± 3.14	**81.31 ± 0.08**	**54.42 ± 0.33**	**83.68 ± 0.06**	55.20 ± 0.20	**84.40**	54.43
ExCeL [12]	86.89 ± 0.23	66.55 ± 0.43	80.70 ± 0.06	55.21 ± 0.56	82.40 ± 0.04	57.90 ± 0.40	83.33	59.89
Training methods without outliers								
RankOOD (ours)	90.21 ± 0.41	**31.72 ± 0.67**	80.67 ± 0.40	**52.59 ± 0.75**	**85.30 ± 0.18**	**50.05 ± 0.16**	**85.39**	**44.79**
CRAFT[13]	**91.11 ± 0.04**	31.94 ± 1.41	80.90 ± 0.33	53.73 ± 0.62	**83.65 ± 0.41**	**54.62 ± 0.57**	**85.22**	46.76
ConfBranch [3]	89.84 ± 0.24	**31.28 ± 0.66**	71.60 ± 0.62	70.21 ± 0.83	79.10 ± 0.24	61.44 ± 0.34	80.18	54.31
G-ODIN [9]	89.12 ± 0.57	45.54 ± 2.52	77.15 ± 0.28	67.58 ± 0.98	77.28 ± 0.10	69.87 ± 0.46	81.18	61.00
CSI [26]	89.51 ± 0.19	33.66 ± 0.64	71.45 ± 0.27	70.26 ± 0.56	N/A	N/A	80.48	51.96
ARPL [2]	87.44 ± 0.15	40.33 ± 0.70	74.94 ± 0.93	61.56 ± 1.81	82.02 ± 0.10	55.74 ± 0.70	81.47	52.54
MOS [11]	71.45 ± 3.09	78.72 ± 5.86	80.40 ± 0.18	56.05 ± 1.01	69.84 ± 0.46	71.60 ± 0.48	73.90	68.79
LogitNorm [28]	**92.33 ± 0.08**	**29.34 ± 0.81**	78.47 ± 0.31	62.89 ± 0.57	82.66 ± 0.15	56.46 ± 0.37	**84.49**	**49.56**
CIDER [19]	**90.71 ± 0.16**	32.11 ± 0.94	73.10 ± 0.39	72.02 ± 0.31	80.58 ± 1.75	60.10 ± 0.73	81.46	54.74
Training methods with outliers								
OE [7]	**94.82 ± 0.21**	**19.84 ± 0.95**	**88.30 ± 0.10**	**30.73 ± 0.11**	**84.84 ± 0.16**	**52.30 ± 0.67**	**89.32**	**34.29**
MCD [30]	**91.03 ± 0.12**	**30.17 ± 0.06**	77.07 ± 0.32	55.88 ± 0.85	83.62 ± 0.09	**54.71 ± 0.83**	83.91	**46.92**
UDG [29]	89.91 ± 0.25	35.34 ± 0.95	78.02 ± 0.10	61.42 ± 0.48	74.30 ± 1.63	68.89 ± 1.72	80.74	55.22
MixOE [32]	88.73 ± 0.82	51.45 ± 7.78	**80.95 ± 0.20**	55.22 ± 0.49	82.62 ± 0.03	57.97 ± 0.40	84.10	54.88

Example 01

Given that $\pi^0=[0,1,3,2]$ and $Ref^0=[25,15,-15,-30]$

When predicted logit $x=[15,7,-1,-3]$

Therefore; $\hat{c}=0,\ \pi^{\hat{c}}=\pi^0, \bar{\pi}=[0,1,2,3]$,

Assume $w_i=1$ and $\gamma=1.1$

$\delta_{\pi^{\hat{0}}_0}=\delta_0=1.1^2,\quad \delta_{\pi^{\hat{0}}_1}=\delta_1=1.1^2$

$\delta_{\pi^{\hat{0}}_2}=\delta_3=1.1^2,\quad \delta_{\pi^{\hat{0}}_3}=\delta_2=1.1$

$U=[(\frac{15}{(1.1)^2}-25),(\frac{7}{(1.1)^2}-15),(\frac{-3}{(1.1)^2}+15),(\frac{-1}{1.1}+30)]$

$\mathit{RankOOD\ Score}=-96.57$

Example 02

Given that $\pi^0=[0,1,3,2]$ and $Ref^0=[25,15,-15,-30]$

When predicted logit $x=[27,7,-10,-1]$

Therefore; $\hat{c}=0,\ \pi^{\hat{c}}=\pi^0, \bar{\pi}=[0,1,3,2]$,

Assume $w_i=1$ and $\gamma=1.1$

Since $\bar{\pi}=\pi^0,\quad \delta_i=\gamma^0=1 \forall i\in[0,\ldots,3]$

$U=[(27-25),(7-15),(-1+15),(-10+30)]$

$\mathit{RankOOD\ Score}=28$

Figure 2. Toy RankOOD-S computation for a four-class problem.

predicted. In contrast, RankOOD training substantially improves the model's ability to maintain deeper ranking consistency on in-distribution (ID) data: for samples classified as Airplane, the conditional probability of correctly predicting the rank-9 label increases to 0.92. However, when evaluated on an OOD dataset such as CIFAR-100, this probability decreases to 0.55. This pronounced drop highlights that OOD examples are significantly less likely to preserve the learned ranking structure, thereby providing an effective signal for distinguishing ID from OOD samples.

Table 3. Performance comparison in *far-OOD detection*. For each column, the top five methods are marked in **bold**. Note that N/A indicates missing results in OpenOOD. *RankOOD ranks third in both AUROC (89.65) and FPR95 (32.04) on far-OOD detection. LogitNorm and G-ODIN achieve the strongest results, outperforming RankOOD by roughly 2.7% in FPR95. Nonetheless, RankOOD surpasses G-ODIN on CIFAR-10 and TinyImageNet and exceeds LogitNorm on CIFAR-100. Moreover, RankOOD outperforms OE on CIFAR-100 and TinyImageNet, while achieving an overall AUROC within 1% of the best-performing outlier based methods.*

Method	CIFAR-10		CIFAR-100		TinyImageNet		Average	
	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓
Post-hoc inference methods								
OpenMax [1]	89.62 ± 0.19	29.69 ± 1.21	79.48 ± 0.41	54.50 ± 0.68	90.20 ± 0.17	33.12 ± 0.66	86.43	39.10
MSP [6]	90.73 ± 0.43	31.72 ± 1.84	77.76 ± 0.44	58.70 ± 1.06	90.13 ± 0.09	35.43 ± 0.38	86.21	41.95
TempScale [5]	90.97 ± 0.52	33.48 ± 2.39	78.74 ± 0.51	57.94 ± 1.14	90.82 ± 0.09	34.00 ± 0.37	86.84	41.81
ODIN [16]	87.96 ± 0.61	57.62 ± 4.24	79.28 ± 0.21	58.86 ± 0.79	91.71 ± 0.19	34.23 ± 1.05	86.32	50.24
MDS [15]	89.72 ± 1.36	32.22 ± 3.40	69.39 ± 1.39	72.26 ± 1.56	74.72 ± 0.26	61.66 ± 0.27	77.94	55.38
MDSEns [15]	73.90 ± 0.27	61.47 ± 0.48	66.00 ± 0.69	66.74 ± 1.04	69.27 ± 0.57	80.96 ± 0.38	69.72	69.72
RMDS [20]	92.20 ± 0.21	25.35 ± 0.73	**82.92 ± 0.42**	52.81 ± 0.63	88.06 ± 0.34	32.45 ± 0.79	87.73	36.87
Gram [20]	71.73 ± 3.20	72.34 ± 6.73	73.36 ± 1.08	64.44 ± 2.37	71.19 ± 0.24	84.36 ± 0.78	72.09	73.71
EBO [17]	91.21 ± 0.92	41.69 ± 5.32	79.77 ± 0.61	56.59 ± 1.38	90.86 ± 0.21	34.86 ± 1.30	87.28	44.38
OpenGAN [14]	54.61 ± 15.5	83.52 ± 11.63	67.88 ± 7.16	70.49 ± 7.38	73.15 ± 4.07	64.16 ± 9.33	65.21	72.72
GradNorm [10]	57.55 ± 3.22	91.90 ± 2.23	69.14 ± 1.05	83.68 ± 1.92	84.26 ± 0.87	66.45 ± 0.22	70.32	80.68
ReAct [24]	90.42 ± 1.41	44.90 ± 8.37	80.39 ± 0.49	54.20 ± 1.56	**92.31 ± 0.56**	28.50 ± 0.95	87.71	42.53
MLS [8]	91.10 ± 0.89	41.68 ± 5.27	79.67 ± 0.57	56.73 ± 1.33	91.11 ± 0.19	34.03 ± 1.21	87.29	44.15
KLM [8]	82.68 ± 0.21	78.31 ± 4.84	76.24 ± 0.52	71.65 ± 2.01	88.53 ± 0.11	40.90 ± 1.08	82.48	63.62
VIM [27]	93.48 ± 0.24	25.05 ± 0.52	81.70 ± 0.62	**50.74 ± 1.00**	91.26 ± 0.19	**27.20 ± 0.30**	**88.81**	**34.33**
KNN [23]	92.96 ± 0.14	24.27 ± 0.40	**82.40 ± 0.17**	53.65 ± 0.28	**93.16 ± 0.22**	**27.27 ± 0.75**	**89.51**	35.06
DICE [25]	84.23 ± 1.89	51.76 ± 4.42	80.01 ± 0.18	56.25 ± 0.60	90.80 ± 0.31	36.51 ± 1.18	85.01	48.17
RankFeat [22]	75.87 ± 5.06	57.44 ± 7.99	67.10 ± 1.42	69.45 ± 1.01	38.22 ± 3.85	97.72 ± 0.75	60.40	74.87
ASH [4]	78.49 ± 2.58	79.03 ± 4.22	80.58 ± 0.66	59.20 ± 2.46	**93.90 ± 0.27**	**27.29 ± 1.12**	84.32	55.17
SHE [31]	85.32 ± 1.43	66.48 ± 5.98	76.92 ± 1.16	64.12 ± 2.70	89.81 ± 0.61	42.17 ± 1.24	84.02	57.59
GEN [18]	91.35 ± 0.69	34.73 ± 1.58	79.68 ± 0.75	56.71 ± 1.59	91.36 ± 0.10	32.10 ± 0.59	87.46	41.18
ExCeL [12]	91.69 ± 0.18	40.03 ± 0.84	**82.04 ± 0.90**	**52.24 ± 1.90**	91.97 ± 0.27	28.45 ± 0.80	88.57	40.24
Training methods without outliers								
RankOOD (ours)	93.19 ± 0.84	20.96 ± 2.55	**83.63 ± 1.06**	**47.44 ± 0.80**	92.14 ± 0.20	**27.73 ± 0.33**	**89.65**	**32.04**
CRAFT [13]	93.94 ± 0.20	**19.40 ± 0.88**	82.03 ± 0.34	**51.86 ± 0.49**	90.88 ± 0.89	32.67 ± 1.13	**88.95**	34.64
ConfBranch [3]	92.85 ± 0.29	21.48 ± 0.94	68.90 ± 1.83	71.82 ± 3.39	90.43 ± 0.18	34.75 ± 0.63	84.06	42.68
G-ODIN [9]	**95.51 ± 0.31**	21.45 ± 1.91	**85.67 ± 1.58**	**42.68 ± 3.19**	**92.33 ± 0.11**	30.18 ± 0.49	**91.17**	**31.44**
CSI [26]	92.00 ± 0.30	26.42 ± 0.29	66.31 ± 1.21	76.92 ± 1.29	N/A	N/A	79.16	51.67
ARPL [2]	89.31 ± 0.32	32.39 ± 0.74	73.69 ± 1.80	63.14 ± 2.53	89.23 ± 0.11	36.46 ± 0.08	84.08	44.00
MOS [11]	76.41 ± 5.93	62.90 ± 6.62	80.17 ± 1.21	57.28 ± 3.29	80.46 ± 0.92	51.56 ± 0.42	79.01	57.25
LogitNorm [28]	**96.74 ± 0.06**	**13.81 ± 0.20**	81.53 ± 1.26	53.61 ± 3.45	**93.04 ± 0.21**	**26.11 ± 0.52**	**90.44**	**31.18**
CIDER [19]	**94.71 ± 0.36**	**20.72 ± 0.85**	80.49 ± 0.68	54.22 ± 1.24	90.66 ± 1.68	30.17 ± 2.75	88.62	35.04
Training methods with outliers								
OE [7]	**96.00 ± 0.13**	**13.13 ± 0.53**	81.41 ± 1.49	54.82 ± 2.79	89.02 ± 0.18	34.17 ± 0.56	88.81	**34.04**
MCD [30]	91.00 ± 1.10	32.03 ± 4.21	74.72 ± 0.78	54.39 ± 1.34	88.94 ± 0.10	29.93 ± 0.30	84.89	38.78
UDG [29]	**94.06 ± 0.90**	**20.35 ± 2.41**	79.59 ± 1.77	59.00 ± 3.35	82.09 ± 2.78	62.04 ± 5.99	85.25	47.13
MixOE [32]	91.93 ± 0.69	33.84 ± 4.77	76.40 ± 1.44	63.88 ± 2.48	88.27 ± 0.41	40.93 ± 0.29	85.53	46.22

Table 4. FPR95 (% ↓) of various methods for different OOD datasets when TinyImageNet is ID. For each column, the top five methods are marked in **bold**. Note that N/A indicates that results are not reported in OpenOOD. *RankOOD achieves SOTA performance in near-OOD and ranks within the top three methods in two out of three far-OOD datasets.*

Method	Near OOD			Far OOD			
	SSB-hard	NINCO	Average	iNaturalist	Textures	OpenImage-O	Average
Post-hoc inference methods							
OpenMax [1]	72.37 ± 0.11	54.59 ± 0.54	63.48 ± 0.25	24.53 ± 0.96	36.80 ± 0.55	38.03 ± 0.49	33.12 ± 0.66
MSP [6]	66.00 ± 0.10	43.65 ± 0.75	54.82 ± 0.35	26.48 ± 0.73	44.58 ± 0.68	35.23 ± 0.18	35.43 ± 0.38
TempScale [5]	66.43 ± 0.26	**43.21 ± 0.70**	54.82 ± 0.23	24.39 ± 0.79	43.57 ± 0.77	34.04 ± 0.31	34.00 ± 0.37
ODIN [16]	73.51 ± 0.38	60.00 ± 0.80	66.76 ± 0.26	22.39 ± 1.87	42.99 ± 1.56	37.30 ± 0.59	34.23 ± 1.05
MDS [15]	83.65 ± 0.47	74.57 ± 0.15	79.11 ± 0.31	58.53 ± 0.75	58.16 ± 0.84	68.29 ± 0.28	61.66 ± 0.27
MDSEns [15]	92.13 ± 0.05	91.36 ± 0.16	91.75 ± 0.10	83.37 ± 0.70	72.27 ± 0.48	87.26 ± 0.10	80.96 ± 0.38
RMDS [20]	**65.91 ± 0.27**	**42.13 ± 1.04**	**54.02 ± 0.58**	24.70 ± 0.90	37.80 ± 1.32	34.85 ± 0.31	32.45 ± 0.79
Gram [21]	85.68 ± 0.85	87.13 ± 1.89	86.40 ± 1.21	85.54 ± 0.40	80.87 ± 1.20	86.66 ± 1.27	84.36 ± 0.78
EBO [17]	69.77 ± 0.32	50.70 ± 0.89	60.24 ± 0.57	26.41 ± 2.29	41.43 ± 1.85	36.74 ± 1.14	34.86 ± 1.30
OpenGAN [14]	88.07 ± 2.23	80.23 ± 5.71	84.15 ± 3.85	60.13 ± 9.79	66.00 ± 9.97	66.34 ± 8.44	64.16 ± 9.33
GradNorm [10]	82.17 ± 0.62	83.17 ± 0.21	82.67 ± 0.30	61.31 ± 2.86	66.88 ± 3.59	71.16 ± 0.23	66.45 ± 0.22
ReAct [24]	71.51 ± 1.92	53.47 ± 2.46	62.49 ± 2.19	22.97 ± 2.25	**29.67 ± 1.35**	32.86 ± 0.74	28.50 ± 0.95
MLS [8]	69.64 ± 0.37	49.87 ± 0.94	59.76 ± 0.59	25.09 ± 2.04	41.25 ± 1.86	35.76 ± 0.74	34.03 ± 1.21
KLM [8]	78.19 ± 2.30	62.33 ± 2.66	70.26 ± 0.64	26.66 ± 1.61	50.24 ± 1.26	45.81 ± 0.59	40.90 ± 1.08
VIM [27]	71.28 ± 0.49	47.10 ± 1.10	59.19 ± 0.71	27.34 ± 0.38	**20.39 ± 0.17**	33.86 ± 0.63	**27.20 ± 0.30**
KNN [23]	73.71 ± 0.31	46.64 ± 0.73	60.18 ± 0.52	24.46 ± 1.06	**24.45 ± 0.29**	32.90 ± 1.12	**27.27 ± 0.75**
DICE [25]	70.84 ± 0.30	52.91 ± 1.20	61.88 ± 0.67	29.66 ± 2.62	40.96 ± 1.87	38.91 ± 1.16	36.51 ± 1.18
RankFeat [22]	90.79 ± 0.37	93.32 ± 0.11	92.06 ± 0.23	98.00 ± 0.80	99.40 ± 0.68	95.77 ± 0.85	97.72 ± 0.75
ASH [4]	72.14 ± 0.97	57.63 ± 0.98	64.89 ± 0.90	22.49 ± 2.24	**25.65 ± 0.80**	33.72 ± 0.97	**27.29 ± 1.12**
SHE [31]	72.64 ± 0.30	60.96 ± 1.33	66.80 ± 0.74	34.38 ± 3.48	45.58 ± 2.42	46.54 ± 1.34	42.17 ± 1.24
GEN [18]	66.79 ± 0.26	43.61 ± 0.61	55.20 ± 0.20	**22.03 ± 0.98**	42.01 ± 0.92	**32.25 ± 0.31**	32.10 ± 0.59
ExCeL [12]	69.28 ± 0.60	46.51 ± 0.20	57.90 ± 0.40	**22.29 ± 1.00**	30.14 ± 0.64	32.91 ± 0.76	28.45 ± 0.80
Training methods without outliers							
RankOOD (Ours)	**60.68 ± 0.61**	**39.43 ± 0.99**	**50.05 ± 0.16**	**19.42 ± 0.71**	33.81 ± 0.95	**29.97 ± 0.56**	**27.73 ± 0.33**
CRAFT [13]	69.07 ± 0.39	**40.40 ± 1.13**	**54.62 ± 0.57**	24.89 ± 1.33	38.73 ± 4.55	33.48 ± 1.74	32.67 ± 1.13
ConfBranch [3]	72.24 ± 0.37	50.63 ± 0.60	61.44 ± 0.34	23.84 ± 0.40	42.42 ± 2.27	37.99 ± 0.09	34.75 ± 0.63
G-ODIN [9]	78.23 ± 0.70	61.52 ± 0.64	69.87 ± 0.46	26.13 ± 0.77	**28.98 ± 1.15**	35.43 ± 0.43	30.18 ± 0.49
CSI [26]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ARPL [2]	**65.73 ± 0.51**	45.75 ± 0.89	55.74 ± 0.70	29.32 ± 0.64	42.87 ± 1.09	37.20 ± 0.69	36.46 ± 0.08
MOS [11]	74.35 ± 0.32	68.85 ± 0.68	71.60 ± 0.48	49.55 ± 0.73	51.27 ± 1.02	53.86 ± 0.30	51.56 ± 0.42
LogitNorm [28]	67.46 ± 0.21	45.46 ± 0.69	56.46 ± 0.37	**15.70 ± 0.61**	32.13 ± 0.67	**30.49 ± 0.62**	**26.11 ± 0.52**
CIDER [19]	75.50 ± 0.68	44.69 ± 0.88	60.10 ± 0.73	26.54 ± 2.27	31.51 ± 3.68	**32.47 ± 2.40**	30.17 ± 2.75
Training methods with outliers							
OE [7]	**64.67 ± 0.25**	**39.93 ± 1.13**	**52.30 ± 0.67**	27.03 ± 0.47	41.92 ± 1.69	33.56 ± 0.46	34.17 ± 0.56
MCD [30]	**65.69 ± 0.36**	43.74 ± 1.32	**54.71 ± 0.83**	**21.74 ± 0.31**	38.11 ± 0.93	**29.93 ± 0.19**	29.93 ± 0.30
UDG [29]	75.84 ± 1.86	61.94 ± 1.61	68.89 ± 1.72	49.26 ± 7.88	71.94 ± 5.25	64.92 ± 5.07	62.04 ± 5.99
MixOE [32]	68.26 ± 0.48	47.69 ± 0.95	57.97 ± 0.40	30.84 ± 0.45	51.44 ± 0.61	40.51 ± 1.06	40.93 ± 0.29

Table 5. AUROC (% ↑) of various methods for different OOD datasets when TinyImageNet is ID. For each column, the top five methods are marked in **bold**. Note that N/A indicates that results are not reported in OpenOOD. *RankOOD ranks within the top three methods in two out of three far-OOD datasets while achieving SOTA performance in near-OOD setting.*

Method	Near OOD			Far OOD			
	SSB-hard	NINCO	Average	iNaturalist	Textures	OpenImage-O	Average
Post-hoc inference methods							
OpenMax [1]	77.53 ± 0.08	83.01 ± 0.17	80.27 ± 0.10	92.32 ± 0.32	90.21 ± 0.07	88.07 ± 0.14	90.20 ± 0.17
MSP [6]	80.38 ± 0.03	86.29 ± 0.11	83.34 ± 0.06	92.80 ± 0.25	88.36 ± 0.13	89.24 ± 0.02	90.13 ± 0.09
TempScale [5]	**80.71 ± 0.02**	**86.67 ± 0.08**	**83.69 ± 0.04**	93.39 ± 0.25	89.24 ± 0.11	89.84 ± 0.02	90.82 ± 0.09
ODIN [16]	77.19 ± 0.06	83.34 ± 0.12	80.27 ± 0.08	**94.37 ± 0.41**	90.65 ± 0.20	90.11 ± 0.15	91.71 ± 0.19
MDS [15]	58.38 ± 0.58	65.48 ± 0.46	61.93 ± 0.51	75.03 ± 0.76	79.25 ± 0.33	69.87 ± 0.14	74.72 ± 0.26
MDSEns [15]	50.46 ± 0.36	58.18 ± 0.42	54.32 ± 0.24	62.16 ± 0.73	80.70 ± 0.48	64.96 ± 0.51	69.27 ± 0.57
RMDS [20]	80.20 ± 0.23	84.94 ± 0.28	82.57 ± 0.25	90.64 ± 0.46	86.77 ± 0.38	86.77 ± 0.22	88.06 ± 0.34
Gram [21]	65.95 ± 1.08	69.40 ± 1.07	67.67 ± 1.07	65.30 ± 0.20	80.53 ± 0.37	67.72 ± 0.58	71.19 ± 0.24
EBO [17]	79.83 ± 0.02	85.17 ± 0.11	82.50 ± 0.05	92.55 ± 0.50	90.79 ± 0.16	89.23 ± 0.26	90.86 ± 0.21
OpenGAN [14]	55.08 ± 1.84	64.49 ± 4.98	59.79 ± 3.39	75.32 ± 3.32	70.58 ± 4.66	73.54 ± 4.48	73.15 ± 4.07
GradNorm [10]	72.12 ± 0.43	73.39 ± 0.63	72.75 ± 0.48	86.06 ± 1.90	86.07 ± 0.36	80.66 ± 1.09	84.26 ± 0.87
ReAct [24]	78.97 ± 1.33	84.76 ± 0.64	81.87 ± 0.98	93.65 ± 0.88	**92.86 ± 0.47**	**90.40 ± 0.35**	**92.31 ± 0.56**
MLS [8]	80.15 ± 0.01	85.65 ± 0.09	82.90 ± 0.04	93.12 ± 0.45	90.60 ± 0.16	89.62 ± 0.21	91.11 ± 0.19
KLM [8]	77.56 ± 0.18	83.96 ± 0.12	80.76 ± 0.08	91.80 ± 0.21	86.13 ± 0.12	87.66 ± 0.17	88.53 ± 0.11
VIM [27]	74.04 ± 0.31	83.32 ± 0.19	78.68 ± 0.24	90.96 ± 0.36	**94.61 ± 0.12**	88.20 ± 0.18	91.26 ± 0.19
KNN [23]	77.03 ± 0.23	86.10 ± 0.12	81.57 ± 0.17	**93.99 ± 0.36**	**95.29 ± 0.02**	**90.19 ± 0.32**	**93.16 ± 0.22**
DICE [25]	79.06 ± 0.05	84.49 ± 0.24	81.78 ± 0.14	91.81 ± 0.79	91.53 ± 0.21	89.06 ± 0.34	90.80 ± 0.31
RankFeat [22]	58.74 ± 0.94	55.10 ± 2.52	56.92 ± 1.59	33.08 ± 4.68	29.10 ± 2.57	52.48 ± 4.44	38.22 ± 3.85
ASH [4]	79.52 ± 0.37	85.24 ± 0.08	82.38 ± 0.19	**95.10 ± 0.47**	**94.77 ± 0.19**	**91.82 ± 0.25**	**93.90 ± 0.27**
SHE [31]	78.30 ± 0.20	82.07 ± 0.33	80.18 ± 0.25	91.43 ± 1.28	90.51 ± 0.19	87.49 ± 0.70	89.81 ± 0.61
GEN [18]	**80.75 ± 0.03**	86.60 ± 0.08	**83.68 ± 0.06**	93.70 ± 0.18	90.25 ± 0.10	90.13 ± 0.06	91.36 ± 0.10
ExCeL [12]	79.39 ± 0.03	85.40 ± 0.04	82.40 ± 0.04	93.76 ± 0.43	92.40 ± 0.05	89.75 ± 0.32	91.97 ± 0.27
Training methods without outliers							
RankOOD (Ours)	**82.57 ± 0.28**	**88.04 ± 0.26**	**85.30 ± 0.18**	**94.69 ± 0.17**	90.85 ± 0.14	**90.89 ± 0.25**	92.14 ± 0.20
CRAFT [13]	80.70 ± 0.18	**86.74 ± 0.84**	**83.65 ± 0.41**	92.85 ± 0.66	89.94 ± 0.49	89.85 ± 0.46	90.88 ± 0.89
ConfBranch [3]	75.01 ± 0.35	83.19 ± 0.14	79.10 ± 0.24	93.40 ± 0.09	89.64 ± 0.52	88.26 ± 0.07	90.43 ± 0.18
G-ODIN [9]	72.94 ± 0.05	81.63 ± 0.21	77.28 ± 0.10	93.12 ± 0.21	**93.67 ± 0.21**	90.18 ± 0.15	**92.33 ± 0.11**
CSI [26]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ARPL [2]	79.24 ± 0.14	84.81 ± 0.07	82.02 ± 0.10	91.54 ± 0.05	88.11 ± 0.34	88.04 ± 0.20	89.23 ± 0.11
MOS [11]	66.54 ± 0.49	73.14 ± 0.47	69.84 ± 0.46	79.69 ± 1.38	81.38 ± 0.75	80.29 ± 0.68	80.46 ± 0.92
LogitNorm [28]	78.42 ± 0.23	**86.90 ± 0.07**	82.66 ± 0.15	**96.26 ± 0.20**	91.85 ± 0.21	**91.01 ± 0.27**	**93.04 ± 0.21**
CIDER [19]	76.04 ± 2.37	85.13 ± 1.13	80.58 ± 1.75	90.69 ± 2.13	92.38 ± 1.35	88.92 ± 1.58	90.66 ± 1.68
Training methods with outliers							
OE [7]	**82.34 ± 0.16**	**87.35 ± 0.23**	**84.84 ± 0.16**	90.30 ± 0.16	87.76 ± 0.32	89.01 ± 0.24	89.02 ± 0.18
MCD [30]	**81.51 ± 0.14**	85.74 ± 0.07	83.62 ± 0.09	90.83 ± 0.10	86.87 ± 0.12	89.12 ± 0.18	88.94 ± 0.10
UDG [29]	70.73 ± 1.74	77.88 ± 1.56	74.30 ± 1.63	85.95 ± 2.97	81.79 ± 2.57	78.54 ± 2.98	82.09 ± 2.78
MixOE [32]	80.23 ± 0.15	85.01 ± 0.10	82.62 ± 0.03	90.64 ± 0.36	86.80 ± 0.45	87.36 ± 0.49	88.27 ± 0.41

Table 6. FPR95 (% ↓) of various methods for different OOD datasets when CIFAR-100 is ID. For each column, the top five methods are marked in **bold**. *RankOOD ranks within the top four methods in four out of six OOD datasets while achieving second best average FPR95 for both near- and far-OOD detection.*

Method	Near OOD			Far OOD				
	CIFAR-10	TIN	Average	MNIST	SVHN	Textures	Places365	Average
Post-hoc inference methods								
OpenMax [1]	60.17 ± 0.97	52.99 ± 0.51	56.58 ± 0.73	53.82 ± 4.74	53.20 ± 1.78	56.12 ± 1.91	**54.85 ± 1.42**	54.50 ± 0.68
MSP [6]	**58.91 ± 0.93**	50.70 ± 0.34	**54.80 ± 0.33**	57.23 ± 4.68	59.07 ± 2.53	61.88 ± 1.28	56.62 ± 0.87	58.70 ± 1.06
TempScale [5]	**58.72 ± 0.81**	50.26 ± 0.16	**54.49 ± 0.48**	56.05 ± 4.61	57.71 ± 2.68	61.56 ± 1.43	56.46 ± 0.94	57.94 ± 1.14
ODIN [16]	60.64 ± 0.56	55.19 ± 0.57	57.91 ± 0.51	45.94 ± 3.29	67.41 ± 3.88	62.37 ± 2.96	59.71 ± 0.92	58.86 ± 0.79
MDS [15]	88.00 ± 0.49	79.05 ± 1.22	83.53 ± 0.60	71.72 ± 2.94	67.21 ± 6.09	70.49 ± 2.48	79.61 ± 0.34	72.26 ± 1.56
MDSEns [15]	95.94 ± 0.16	95.82 ± 0.12	95.88 ± 0.04	**2.83 ± 0.86**	82.57 ± 2.58	84.94 ± 0.83	96.61 ± 0.17	66.74 ± 1.04
RMDS [20]	61.37 ± 0.24	49.56 ± 0.90	55.46 ± 0.41	52.05 ± 6.28	51.65 ± 3.68	53.99 ± 1.06	**53.57 ± 0.43**	52.81 ± 0.63
Gram [21]	92.71 ± 0.64	91.85 ± 0.86	92.28 ± 0.29	53.53 ± 7.45	**20.06 ± 1.96**	89.51 ± 2.54	94.67 ± 0.60	64.44 ± 2.37
EBO [17]	59.21 ± 0.75	52.03 ± 0.50	55.62 ± 0.61	52.62 ± 3.83	53.62 ± 3.14	62.35 ± 2.06	57.75 ± 0.86	56.59 ± 1.38
OpenGAN [14]	78.83 ± 3.94	74.21 ± 1.25	76.52 ± 2.59	63.09 ± 23.3	70.35 ± 2.06	74.77 ± 1.78	73.75 ± 8.32	70.49 ± 7.38
GradNorm [10]	84.30 ± 0.36	86.85 ± 0.62	85.58 ± 0.46	86.97 ± 1.44	69.90 ± 7.94	92.51 ± 0.61	85.32 ± 0.44	83.68 ± 1.92
ReAct [24]	61.30 ± 0.43	51.47 ± 0.47	56.39 ± 0.34	56.04 ± 5.66	50.41 ± 2.02	55.04 ± 0.82	**55.30 ± 0.41**	54.20 ± 1.56
MLS [8]	**59.11 ± 0.64**	51.83 ± 0.70	55.47 ± 0.66	52.95 ± 3.82	53.90 ± 3.04	62.39 ± 2.13	57.68 ± 0.91	56.73 ± 1.33
KLM [8]	84.77 ± 2.95	71.07 ± 0.59	77.92 ± 1.31	73.09 ± 6.67	50.30 ± 7.04	81.80 ± 5.80	81.40 ± 1.58	71.65 ± 2.01
VIM [27]	70.59 ± 0.43	54.66 ± 0.42	62.63 ± 0.27	48.32 ± 1.07	46.22 ± 5.46	**46.86 ± 2.29**	61.57 ± 0.77	**50.74 ± 1.00**
KNN [23]	72.80 ± 0.44	49.65 ± 0.37	61.22 ± 0.14	48.58 ± 4.67	51.75 ± 3.12	**53.56 ± 2.32**	60.70 ± 1.03	53.65 ± 0.28
DICE [25]	60.98 ± 1.10	54.93 ± 0.53	57.95 ± 0.53	51.79 ± 3.67	49.58 ± 3.32	64.23 ± 1.65	59.39 ± 1.25	56.25 ± 0.60
RankFeat [22]	82.78 ± 1.56	78.40 ± 0.95	80.59 ± 1.10	75.01 ± 5.83	58.49 ± 2.30	66.87 ± 3.80	77.42 ± 1.96	69.45 ± 1.01
ASH [4]	68.06 ± 0.44	63.35 ± 0.90	65.71 ± 0.24	66.58 ± 3.88	46.00 ± 2.67	61.27 ± 2.74	62.95 ± 0.99	59.20 ± 2.46
SHE [31]	60.41 ± 0.51	57.74 ± 0.73	59.07 ± 0.25	58.78 ± 2.70	59.15 ± 7.61	73.29 ± 3.22	65.24 ± 0.98	64.12 ± 2.70
GEN [18]	**58.87 ± 0.69**	49.98 ± 0.05	**54.42 ± 0.33**	53.92 ± 5.71	55.45 ± 2.76	61.23 ± 1.40	56.25 ± 1.01	56.71 ± 1.59
ExCeL [12]	61.07 ± 0.81	**49.35 ± 0.31**	55.21 ± 0.56	54.67 ± 5.86	**45.13 ± 0.33**	**51.14 ± 0.14**	58.02 ± 1.28	**52.24 ± 1.90**
Training methods without outliers								
RankOOD (Ours)	**55.71 ± 1.92**	49.47 ± 0.80	**52.59 ± 0.75**	**42.25 ± 2.92**	**39.21 ± 1.59**	**52.88 ± 1.89**	55.40 ± 0.47	**47.44 ± 0.80**
CRAFT [13]	59.19 ± 0.64	**48.26 ± 1.21**	53.73 ± 0.62	48.95 ± 1.90	47.50 ± 5.22	56.97 ± 1.77	**54.02 ± 0.30**	**51.86 ± 0.49**
ConfBranch [3]	74.56 ± 1.22	65.86 ± 0.56	70.21 ± 0.83	55.95 ± 6.15	76.01 ± 12.3	85.43 ± 1.17	69.90 ± 0.28	71.82 ± 3.39
G-ODIN [9]	78.82 ± 1.86	56.34 ± 0.45	67.58 ± 0.98	**27.19 ± 6.24**	**42.68 ± 5.74**	**35.83 ± 1.15**	65.03 ± 1.16	**42.68 ± 3.19**
CSI [26]	72.62 ± 0.49	67.90 ± 0.64	70.26 ± 0.56	80.54 ± 4.87	67.21 ± 3.35	90.51 ± 1.47	69.41 ± 0.58	76.92 ± 1.29
ARPL [2]	64.84 ± 1.25	58.27 ± 2.40	61.56 ± 1.81	59.12 ± 8.04	59.76 ± 1.58	71.66 ± 1.81	62.01 ± 0.89	63.14 ± 2.53
MOS [11]	60.60 ± 1.47	51.49 ± 0.69	56.05 ± 1.01	52.70 ± 3.81	56.33 ± 8.46	61.24 ± 2.06	58.86 ± 0.41	57.28 ± 3.29
LogitNorm [28]	73.88 ± 1.21	51.89 ± 0.10	62.89 ± 0.57	**34.12 ± 8.32**	47.52 ± 8.02	77.38 ± 2.99	55.44 ± 1.45	53.61 ± 3.45
CIDER [19]	82.71 ± 1.25	61.33 ± 0.64	72.02 ± 0.31	75.32 ± 4.21	**17.82 ± 2.80**	54.43 ± 2.56	69.30 ± 1.81	54.22 ± 1.24
Training methods with outliers								
OE [7]	61.26 ± 0.22	**0.21 ± 0.01**	**30.73 ± 0.11**	53.31 ± 9.91	51.84 ± 3.45	55.83 ± 1.82	58.30 ± 0.72	54.82 ± 2.79
MCD [30]	62.65 ± 0.54	**49.10 ± 1.29**	55.88 ± 0.85	62.78 ± 2.91	**43.71 ± 3.73**	56.89 ± 0.64	**54.17 ± 1.13**	54.39 ± 1.34
UDG [29]	66.40 ± 0.51	56.43 ± 0.68	61.42 ± 0.48	**45.14 ± 12.8**	59.67 ± 5.62	71.33 ± 3.59	59.85 ± 0.57	59.00 ± 3.35
MixOE [32]	61.12 ± 1.08	**49.32 ± 0.36**	55.22 ± 0.49	59.49 ± 7.74	73.09 ± 4.00	66.04 ± 0.98	56.93 ± 0.78	63.88 ± 2.48

Table 7. AUROC (% ↑) of various methods for different OOD datasets when CIFAR-100 is ID. For each column, the top five methods are marked in **bold**. *RankOOD ranks within the top four methods in three out of five far-OOD detection scenarios.*

Method	Near OOD			Far OOD				
	CIFAR-10	TIN	Average	MNIST	SVHN	Textures	Places365	Average
Post-hoc inference methods								
OpenMax [1]	74.38 ± 0.37	78.44 ± 0.14	76.41 ± 0.25	76.01 ± 1.39	82.07 ± 1.53	80.56 ± 0.09	79.29 ± 0.40	79.48 ± 0.41
MSP [6]	78.47 ± 0.07	82.07 ± 0.17	80.27 ± 0.11	76.08 ± 1.86	78.42 ± 0.89	77.32 ± 0.71	79.22 ± 0.29	77.76 ± 0.44
TempScale [5]	**79.02 ± 0.06**	82.79 ± 0.09	80.90 ± 0.07	77.27 ± 1.85	79.79 ± 1.05	78.11 ± 0.72	79.80 ± 0.25	78.74 ± 0.51
ODIN [16]	78.18 ± 0.14	81.63 ± 0.08	79.90 ± 0.11	83.79 ± 1.31	74.54 ± 0.76	79.33 ± 1.08	79.45 ± 0.26	79.28 ± 0.21
MDS [15]	55.87 ± 0.22	61.50 ± 0.28	58.69 ± 0.09	67.47 ± 0.81	70.68 ± 6.40	76.26 ± 0.69	63.15 ± 0.49	69.39 ± 1.39
MDSEns [15]	43.85 ± 0.31	48.78 ± 0.19	46.31 ± 0.24	**98.21 ± 0.78**	53.76 ± 1.63	69.75 ± 1.14	42.27 ± 0.73	66.00 ± 0.69
RMDS [20]	77.75 ± 0.19	82.55 ± 0.02	80.15 ± 0.11	79.74 ± 2.49	84.89 ± 1.10	**83.65 ± 0.51**	**83.40 ± 0.46**	**82.92 ± 0.42**
Gram [21]	49.41 ± 0.58	53.91 ± 1.58	51.66 ± 0.77	80.71 ± 4.15	**95.55 ± 0.60**	70.79 ± 1.32	46.38 ± 1.21	73.36 ± 1.08
EBO [17]	**79.05 ± 0.11**	82.76 ± 0.08	**80.91 ± 0.08**	79.18 ± 1.37	82.03 ± 1.74	78.35 ± 0.83	79.52 ± 0.23	79.77 ± 0.61
OpenGAN [14]	63.23 ± 2.44	68.74 ± 2.29	65.98 ± 1.26	68.14 ± 18.8	68.40 ± 2.15	65.84 ± 3.43	69.13 ± 7.08	67.88 ± 7.16
GradNorm [10]	70.32 ± 0.20	69.95 ± 0.79	70.13 ± 0.47	65.35 ± 1.12	76.95 ± 4.73	64.58 ± 0.13	69.69 ± 0.17	69.14 ± 1.05
ReAct [24]	78.65 ± 0.05	82.88 ± 0.08	80.77 ± 0.05	78.37 ± 1.59	83.01 ± 0.97	80.15 ± 0.46	**80.03 ± 0.11**	80.39 ± 0.49
MLS [8]	**79.21 ± 0.10**	82.90 ± 0.05	**81.05 ± 0.07**	78.91 ± 1.47	81.65 ± 1.49	78.39 ± 0.84	79.75 ± 0.24	79.67 ± 0.57
KLM [8]	73.91 ± 0.25	79.22 ± 0.28	76.56 ± 0.25	74.15 ± 2.59	79.34 ± 0.44	75.77 ± 0.45	75.70 ± 0.24	76.24 ± 0.52
VIM [27]	72.21 ± 0.41	77.76 ± 0.16	74.98 ± 0.13	81.89 ± 1.02	83.14 ± 3.71	**85.91 ± 0.78**	75.85 ± 0.37	81.70 ± 0.62
KNN [23]	77.02 ± 0.25	**83.34 ± 0.16**	80.18 ± 0.15	82.36 ± 1.52	84.15 ± 1.09	**83.66 ± 0.83**	79.43 ± 0.47	**82.40 ± 0.17**
DICE [25]	78.04 ± 0.32	80.72 ± 0.30	79.38 ± 0.23	79.86 ± 1.89	84.22 ± 2.00	77.63 ± 0.34	78.33 ± 0.66	80.01 ± 0.18
RankFeat [22]	58.04 ± 2.36	65.72 ± 0.22	61.88 ± 1.28	63.03 ± 3.86	72.14 ± 1.39	69.40 ± 3.08	63.82 ± 1.83	67.10 ± 1.42
ASH [4]	76.48 ± 0.30	79.92 ± 0.20	78.20 ± 0.15	77.23 ± 0.46	**85.60 ± 1.40**	80.72 ± 0.70	78.76 ± 0.16	80.58 ± 0.66
SHE [31]	78.15 ± 0.03	79.74 ± 0.36	78.95 ± 0.18	76.76 ± 1.07	80.97 ± 3.98	73.64 ± 1.28	76.30 ± 0.51	76.92 ± 1.16
GEN [18]	**79.38 ± 0.04**	**83.25 ± 0.13**	**81.31 ± 0.08**	78.29 ± 2.05	81.41 ± 1.50	78.74 ± 0.81	**80.28 ± 0.27**	79.68 ± 0.75
ExCeL [12]	78.14 ± 0.09	**83.26 ± 0.03**	80.70 ± 0.06	78.99 ± 1.73	**85.91 ± 0.73**	**83.28 ± 0.58**	79.98 ± 0.57	**82.04 ± 0.90**
Training methods without outliers								
RankOOD (Ours)	78.84 ± 0.79	82.50 ± 0.39	80.67 ± 0.40	**84.00 ± 2.15**	**87.75 ± 1.25**	82.04 ± 0.70	80.73 ± 0.46	**83.63 ± 1.06**
CRAFT [13]	**78.67 ± 0.21**	83.14 ± 0.73	80.90 ± 0.33	80.34 ± 1.84	85.16 ± 1.15	80.91 ± 0.45	**81.71 ± 0.12**	82.03 ± 0.34
ConfBranch [3]	68.80 ± 0.73	74.41 ± 0.54	71.60 ± 0.62	74.29 ± 4.44	65.51 ± 8.07	65.39 ± 0.16	70.42 ± 0.26	68.90 ± 1.83
G-ODIN [9]	73.04 ± 0.39	81.26 ± 0.29	77.15 ± 0.28	**91.15 ± 2.86**	83.74 ± 3.10	**89.62 ± 0.36**	78.17 ± 0.62	**85.67 ± 1.58**
CSI [26]	69.50 ± 0.43	73.40 ± 0.13	71.45 ± 0.27	51.79 ± 6.77	80.24 ± 1.80	62.22 ± 0.98	70.99 ± 0.54	66.31 ± 1.21
ARPL [2]	73.38 ± 0.78	76.50 ± 1.11	74.94 ± 0.93	73.77 ± 5.89	76.45 ± 1.00	69.93 ± 1.33	74.62 ± 0.57	73.69 ± 1.80
MOS [11]	78.54 ± 0.13	82.26 ± 0.25	80.40 ± 0.18	80.68 ± 1.65	81.59 ± 3.81	79.92 ± 0.57	78.50 ± 0.34	80.17 ± 1.21
LogitNorm [28]	74.57 ± 0.39	82.37 ± 0.24	78.47 ± 0.31	**90.69 ± 1.38**	82.80 ± 4.57	72.37 ± 0.67	**80.25 ± 0.61**	81.53 ± 1.26
CIDER [19]	67.55 ± 0.60	78.65 ± 0.35	73.10 ± 0.39	68.14 ± 3.98	**97.17 ± 0.34**	82.21 ± 1.93	74.43 ± 0.64	80.49 ± 0.68
Training methods with outliers								
OE [7]	76.70 ± 0.19	**99.89 ± 0.02**	**88.30 ± 0.10**	80.68 ± 5.82	84.37 ± 1.34	82.18 ± 0.68	78.39 ± 0.41	81.41 ± 1.49
MCD [30]	75.40 ± 0.46	78.75 ± 0.21	77.07 ± 0.32	68.25 ± 1.99	75.92 ± 0.37	77.07 ± 0.76	77.65 ± 0.09	74.72 ± 0.78
UDG [29]	75.15 ± 0.15	80.90 ± 0.21	78.02 ± 0.10	**83.88 ± 5.98**	79.80 ± 1.61	75.57 ± 0.80	79.11 ± 0.17	79.59 ± 1.77
MixOE [32]	78.17 ± 0.29	**83.73 ± 0.12**	**80.95 ± 0.20**	76.06 ± 5.52	72.28 ± 0.81	77.34 ± 0.91	79.92 ± 0.30	76.40 ± 1.44

Table 8. FPR95 (% ↓) of various methods for different OOD datasets when CIFAR-10 is ID. For each column, the top five methods are marked in **bold**. *RankOOD ranks within the top five methods in two out of three near-OOD detection scenarios.*

Method	Near OOD			Far OOD				
	CIFAR-100	TIN	Average	MNIST	SVHN	Textures	Places365	Average
Post-hoc inference methods								
OpenMax [1]	48.06 ± 3.25	39.18 ± 1.44	43.62 ± 2.27	23.33 ± 4.67	25.40 ± 1.47	31.50 ± 4.05	38.52 ± 2.27	29.69 ± 1.21
MSP [6]	53.08 ± 4.86	43.27 ± 3.00	48.17 ± 3.92	23.64 ± 5.81	25.82 ± 1.64	34.96 ± 4.64	42.47 ± 3.81	31.72 ± 1.84
TempScale [5]	55.81 ± 5.07	46.11 ± 3.63	50.96 ± 4.32	23.53 ± 7.05	26.97 ± 2.65	38.16 ± 5.89	45.27 ± 4.50	33.48 ± 2.39
ODIN [16]	77.00 ± 5.74	75.38 ± 6.42	76.19 ± 6.08	23.83 ± 12.3	68.61 ± 0.52	67.70 ± 11.1	70.36 ± 6.96	57.62 ± 4.24
MDS [15]	52.81 ± 3.62	46.99 ± 4.36	49.90 ± 3.98	27.30 ± 3.55	25.96 ± 2.52	27.94 ± 4.20	47.67 ± 4.54	32.22 ± 3.40
MDSEns [15]	91.87 ± 0.10	92.66 ± 0.42	92.26 ± 0.20	**1.30 ± 0.51**	74.34 ± 1.04	76.07 ± 0.17	94.16 ± 0.33	61.47 ± 0.48
RMDS [20]	43.86 ± 3.49	33.91 ± 1.39	38.89 ± 2.39	21.49 ± 2.32	23.46 ± 1.48	25.25 ± 0.53	31.20 ± 0.28	25.35 ± 0.73
Gram [21]	91.68 ± 2.24	90.06 ± 1.59	90.87 ± 1.91	70.30 ± 8.96	33.91 ± 17.4	94.64 ± 2.71	90.49 ± 1.93	72.34 ± 6.73
EBO [17]	66.60 ± 4.46	56.08 ± 4.83	61.34 ± 4.63	24.99 ± 12.9	35.12 ± 6.11	51.82 ± 6.11	54.85 ± 6.52	41.69 ± 5.32
OpenGAN [14]	94.84 ± 3.83	94.11 ± 4.21	94.48 ± 4.01	79.54 ± 19.7	75.27 ± 26.9	83.95 ± 14.9	95.32 ± 4.45	83.52 ± 11.6
GradNorm [10]	94.54 ± 1.11	94.89 ± 0.60	94.72 ± 0.82	85.41 ± 4.85	91.65 ± 2.42	98.09 ± 0.49	92.46 ± 2.28	91.90 ± 2.23
ReAct [24]	67.40 ± 7.34	59.71 ± 7.31	63.56 ± 7.33	33.77 ± 18.0	50.23 ± 15.9	51.42 ± 11.4	44.20 ± 3.35	44.90 ± 8.37
MLS [8]	66.59 ± 4.44	56.06 ± 4.82	61.32 ± 4.62	25.06 ± 12.9	35.09 ± 6.09	51.73 ± 6.13	54.84 ± 6.51	41.68 ± 5.27
KLM [8]	90.55 ± 5.83	85.18 ± 7.60	87.86 ± 6.37	76.22 ± 12.1	59.47 ± 7.06	81.95 ± 9.95	95.58 ± 2.12	78.31 ± 4.84
VIM [27]	49.19 ± 3.15	40.49 ± 1.55	44.84 ± 2.31	18.36 ± 1.42	19.29 ± 0.41	**21.14 ± 1.83**	41.43 ± 2.17	25.05 ± 0.52
KNN [23]	37.64 ± 0.31	30.37 ± 0.65	34.01 ± 0.38	20.05 ± 1.36	22.60 ± 1.26	24.06 ± 0.55	30.38 ± 0.63	24.27 ± 0.40
DICE [25]	73.71 ± 7.67	66.37 ± 7.68	70.04 ± 7.64	30.83 ± 10.5	36.61 ± 4.74	62.42 ± 4.79	77.19 ± 12.6	51.76 ± 4.42
RankFeat [22]	65.32 ± 3.48	56.44 ± 5.76	60.88 ± 4.60	61.86 ± 12.8	64.49 ± 7.38	59.71 ± 9.79	43.70 ± 7.39	57.44 ± 7.99
ASH [4]	87.31 ± 2.06	86.25 ± 1.58	86.78 ± 1.82	70.00 ± 10.6	83.64 ± 6.48	84.59 ± 1.74	77.89 ± 7.28	79.03 ± 4.22
SHE [31]	81.00 ± 3.42	78.30 ± 3.52	79.65 ± 3.47	42.22 ± 20.6	62.74 ± 4.01	84.60 ± 5.30	76.36 ± 5.32	66.48 ± 5.98
GEN [18]	58.75 ± 3.97	48.59 ± 2.34	53.67 ± 3.14	23.00 ± 7.75	28.14 ± 2.59	40.74 ± 6.61	47.03 ± 3.22	34.73 ± 1.58
ExCeL [12]	71.16 ± 1.34	61.42 ± 0.26	66.55 ± 0.43	**15.46 ± 1.89**	31.78 ± 3.65	53.67 ± 2.19	55.09 ± 1.12	40.03 ± 0.84
Training methods without outliers								
RankOOD (Ours)	**35.42 ± 0.65**	28.01 ± 1.41	**31.72 ± 0.67**	17.18 ± 4.09	15.75 ± 3.20	22.66 ± 5.39	28.27 ± 0.83	20.96 ± 2.55
CRAFT [13]	36.61 ± 2.93	**27.28 ± 0.09**	31.94 ± 1.41	17.13 ± 0.99	14.58 ± 4.62	**20.78 ± 0.04**	**25.12 ± 0.15**	**19.40 ± 0.88**
ConfBranch [3]	**34.44 ± 0.81**	**28.11 ± 0.61**	**31.28 ± 0.66**	**15.79 ± 2.00**	**14.06 ± 0.84**	27.24 ± 1.32	28.85 ± 1.03	21.48 ± 0.94
G-ODIN [9]	48.86 ± 2.91	42.21 ± 2.18	45.54 ± 2.52	**4.53 ± 2.08**	**10.72 ± 0.88**	27.27 ± 6.73	43.30 ± 3.57	21.45 ± 1.91
CSI [26]	37.57 ± 0.89	29.74 ± 0.42	33.66 ± 0.64	24.41 ± 1.57	17.56 ± 0.12	28.95 ± 1.33	34.76 ± 1.52	26.42 ± 0.29
ARPL [2]	43.38 ± 0.37	37.28 ± 1.21	40.33 ± 0.70	21.49 ± 2.03	35.68 ± 3.48	35.19 ± 1.79	37.21 ± 0.80	32.39 ± 0.74
MOS [11]	79.38 ± 5.06	78.05 ± 6.69	78.72 ± 5.86	65.95 ± 17.5	57.79 ± 5.79	76.78 ± 3.86	51.09 ± 1.33	62.90 ± 6.62
LogitNorm [28]	**34.37 ± 1.30**	**24.30 ± 0.54**	**29.34 ± 0.81**	**3.93 ± 1.99**	**8.33 ± 1.78**	**21.94 ± 0.85**	**21.04 ± 0.71**	**13.81 ± 0.20**
CIDER [19]	**35.60 ± 0.78**	28.61 ± 1.10	32.11 ± 0.94	24.76 ± 2.82	**8.04 ± 0.43**	25.05 ± 3.29	**25.03 ± 1.36**	**20.72 ± 0.85**
Training methods with outliers								
OE [7]	36.71 ± 2.06	**2.97 ± 1.17**	**19.84 ± 0.95**	24.67 ± 2.55	**1.25 ± 0.36**	**12.07 ± 2.14**	**14.53 ± 2.80**	**13.13 ± 0.53**
MCD [30]	**34.36 ± 0.37**	**25.98 ± 0.44**	**30.17 ± 0.06**	62.11 ± 11.8	19.43 ± 5.93	22.51 ± 5.16	**24.10 ± 1.58**	32.03 ± 4.21
UDG [29]	40.75 ± 0.69	29.93 ± 1.27	35.34 ± 0.95	16.61 ± 5.14	17.39 ± 7.87	**19.70 ± 1.89**	27.70 ± 1.80	**20.35 ± 2.41**
MixOE [32]	58.29 ± 8.25	44.62 ± 7.57	51.45 ± 7.78	38.28 ± 13.4	20.36 ± 3.99	33.19 ± 4.28	43.54 ± 4.95	33.84 ± 4.77

Table 9. AUROC (% ↑) of various methods for different OOD datasets when CIFAR-10 is ID. For each column, the top five methods are marked in **bold**. *RankOOD achieves on par performance compared to CRAFT, a class rank based method.*

Method	Near OOD			Far OOD				
	CIFAR-100	TIN	Average	MNIST	SVHN	Textures	Places365	Average
Post-hoc inference methods								
OpenMax [1]	86.91 ± 0.31	88.32 ± 0.28	87.62 ± 0.29	90.50 ± 0.44	89.77 ± 0.45	89.58 ± 0.60	88.63 ± 0.28	89.62 ± 0.19
MSP [6]	87.19 ± 0.33	88.87 ± 0.19	88.03 ± 0.25	92.63 ± 1.57	91.46 ± 0.40	89.89 ± 0.71	88.92 ± 0.47	90.73 ± 0.43
TempScale [5]	87.17 ± 0.40	89.00 ± 0.23	88.09 ± 0.31	93.11 ± 1.77	91.66 ± 0.52	90.01 ± 0.74	89.11 ± 0.52	90.97 ± 0.52
ODIN [16]	82.18 ± 1.87	83.55 ± 1.84	82.87 ± 1.85	95.24 ± 1.96	84.58 ± 0.77	86.94 ± 2.26	85.07 ± 1.24	87.96 ± 0.61
MDS [15]	83.59 ± 2.27	84.81 ± 2.53	84.20 ± 2.40	90.10 ± 2.41	91.18 ± 0.47	92.69 ± 1.06	84.90 ± 2.54	89.72 ± 1.36
MDSEns [15]	61.29 ± 0.23	59.57 ± 0.53	60.43 ± 0.26	**99.17 ± 0.41**	66.56 ± 0.58	77.40 ± 0.28	52.47 ± 0.15	73.90 ± 0.27
RMDS [20]	88.83 ± 0.35	90.76 ± 0.27	89.80 ± 0.28	93.22 ± 0.80	91.84 ± 0.26	92.23 ± 0.23	91.51 ± 0.11	92.20 ± 0.21
Gram [21]	58.33 ± 4.49	58.98 ± 5.19	58.66 ± 4.83	72.64 ± 2.34	91.52 ± 4.45	62.34 ± 8.27	60.44 ± 3.41	71.73 ± 3.20
EBO [17]	86.36 ± 0.58	88.80 ± 0.36	87.58 ± 0.46	94.32 ± 2.53	91.79 ± 0.98	89.47 ± 0.70	89.25 ± 0.78	91.21 ± 0.92
OpenGAN [14]	52.81 ± 7.69	54.62 ± 7.68	53.71 ± 7.68	56.14 ± 24.1	52.81 ± 27.6	56.14 ± 18.3	53.34 ± 5.79	54.61 ± 15.5
GradNorm [10]	54.43 ± 1.59	55.37 ± 0.41	54.90 ± 0.98	63.72 ± 7.37	53.91 ± 6.36	52.07 ± 4.09	60.50 ± 5.33	57.55 ± 3.22
ReAct [24]	85.93 ± 0.83	88.29 ± 0.44	87.11 ± 0.61	92.81 ± 3.03	89.12 ± 3.19	89.38 ± 1.49	90.35 ± 0.78	90.42 ± 1.41
MLS [8]	86.31 ± 0.59	88.72 ± 0.36	87.52 ± 0.47	94.15 ± 2.48	91.69 ± 0.94	89.41 ± 0.71	89.14 ± 0.76	91.10 ± 0.89
KLM [8]	77.89 ± 0.75	80.49 ± 0.85	79.19 ± 0.80	85.00 ± 2.04	84.99 ± 1.18	82.35 ± 0.33	78.37 ± 0.33	82.68 ± 0.21
VIM [27]	87.75 ± 0.28	89.62 ± 0.33	88.68 ± 0.28	94.76 ± 0.38	94.50 ± 0.48	**95.15 ± 0.34**	89.49 ± 0.39	93.48 ± 0.24
KNN [23]	**89.73 ± 0.14**	91.56 ± 0.26	90.64 ± 0.20	94.26 ± 0.38	92.67 ± 0.30	93.16 ± 0.24	**91.77 ± 0.23**	92.96 ± 0.14
DICE [25]	77.01 ± 0.88	79.67 ± 0.87	78.34 ± 0.79	90.37 ± 5.97	90.02 ± 1.77	81.86 ± 2.35	74.67 ± 4.98	84.23 ± 1.89
RankFeat [22]	77.98 ± 2.24	80.94 ± 2.80	79.46 ± 2.52	75.87 ± 5.22	68.15 ± 7.44	73.46 ± 6.49	85.99 ± 3.04	75.87 ± 5.06
ASH [4]	74.11 ± 1.55	76.44 ± 0.61	75.27 ± 1.04	83.16 ± 4.66	73.46 ± 6.41	77.45 ± 2.39	79.89 ± 3.69	78.49 ± 2.58
SHE [31]	80.31 ± 0.69	82.76 ± 0.43	81.54 ± 0.51	90.43 ± 4.76	86.38 ± 1.32	81.57 ± 1.21	82.89 ± 1.22	85.32 ± 1.43
GEN [18]	87.21 ± 0.36	89.20 ± 0.25	88.20 ± 0.30	93.83 ± 2.14	91.97 ± 0.66	90.14 ± 0.76	89.46 ± 0.65	91.35 ± 0.69
ExCeL [12]	85.31 ± 0.26	88.48 ± 0.19	86.89 ± 0.23	**95.87 ± 0.45**	91.40 ± 1.43	89.66 ± 0.64	89.84 ± 0.41	91.69 ± 0.18
Training methods without outliers								
RankOOD (Ours)	89.11 ± 0.24	91.32 ± 0.58	90.21 ± 0.41	93.95 ± 2.23	94.70 ± 1.03	92.63 ± 1.53	91.51 ± 0.36	93.19 ± 0.84
CRAFT [13]	**90.18 ± 0.14**	**92.04 ± 0.06**	**91.11 ± 0.04**	94.59 ± 0.02	94.94 ± 1.10	93.46 ± 0.29	**92.77 ± 0.10**	93.94 ± 0.20
ConfBranch [3]	88.91 ± 0.25	90.77 ± 0.25	89.84 ± 0.24	94.49 ± 0.77	**95.42 ± 0.35**	91.10 ± 0.41	90.39 ± 0.40	92.85 ± 0.29
G-ODIN [9]	88.14 ± 0.60	90.09 ± 0.54	89.12 ± 0.57	**98.95 ± 0.53**	**97.76 ± 0.14**	**95.02 ± 1.10**	90.31 ± 0.65	**95.51 ± 0.31**
CSI [26]	88.16 ± 0.16	90.87 ± 0.23	89.51 ± 0.19	92.55 ± 1.15	95.18 ± 0.45	90.71 ± 0.44	89.56 ± 0.51	92.00 ± 0.30
ARPL [2]	86.76 ± 0.16	88.12 ± 0.14	87.44 ± 0.15	92.62 ± 0.88	87.69 ± 0.97	88.57 ± 0.43	88.39 ± 0.16	89.31 ± 0.32
MOS [11]	70.57 ± 3.04	72.34 ± 3.16	71.45 ± 3.09	74.81 ± 10.1	73.66 ± 9.14	70.35 ± 3.11	86.81 ± 1.85	76.41 ± 5.93
LogitNorm [28]	**90.95 ± 0.22**	**93.70 ± 0.06**	**92.33 ± 0.08**	**99.14 ± 0.45**	**98.25 ± 0.41**	**94.77 ± 0.43**	**94.79 ± 0.16**	**96.74 ± 0.06**
CIDER [19]	89.47 ± 0.19	**91.94 ± 0.19**	**90.71 ± 0.16**	93.30 ± 1.08	**98.06 ± 0.07**	93.71 ± 0.39	**93.77 ± 0.68**	**94.71 ± 0.36**
Training methods with outliers								
OE [7]	**90.54 ± 0.53**	**99.11 ± 0.34**	**94.82 ± 0.21**	90.22 ± 1.31	**99.60 ± 0.14**	**97.58 ± 0.27**	**96.58 ± 0.70**	**96.00 ± 0.13**
MCD [30]	**89.88 ± 0.07**	**92.18 ± 0.18**	**91.03 ± 0.12**	84.22 ± 2.10	93.76 ± 2.30	93.35 ± 1.30	92.66 ± 0.36	91.00 ± 1.10
UDG [29]	88.62 ± 0.32	91.20 ± 0.20	89.91 ± 0.25	**95.81 ± 1.52**	94.55 ± 2.27	**93.92 ± 0.44**	91.97 ± 0.41	**94.06 ± 0.90**
MixOE [32]	87.47 ± 0.97	90.00 ± 0.73	88.73 ± 0.82	91.66 ± 2.21	93.82 ± 1.27	91.84 ± 0.51	90.38 ± 0.55	91.93 ± 0.69

CE Loss Training

Predicted Class / Rank (i)	0	1	2	3	4	5	6	7	8	9
Airplane	1.00	0.29	0.11	0.03	0.01	0.00	0.00	0.00	0.00	0.00
Automobile	1.00	0.56	0.29	0.12	0.05	0.02	0.01	0.00	0.00	0.00
Bird	1.00	0.21	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Cat	1.00	0.42	0.10	0.03	0.01	0.00	0.00	0.00	0.00	0.00
Deer	1.00	0.18	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Dog	1.00	0.57	0.16	0.06	0.01	0.01	0.00	0.00	0.00	0.00
Frog	1.00	0.30	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Horse	1.00	0.36	0.13	0.03	0.01	0.00	0.00	0.00	0.00	0.00
Ship	1.00	0.40	0.14	0.08	0.02	0.01	0.00	0.00	0.00	0.00
Truck	1.00	0.51	0.23	0.10	0.04	0.02	0.02	0.01	0.00	0.00

RankOOD Loss Training

CIFAR-10 (ID)

Predicted Class / Rank (i)	0	1	2	3	4	5	6	7	8	9
Airplane	1.00	0.94	0.93	0.93	0.92	0.92	0.92	0.92	0.92	0.92
Automobile	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.98	0.98
Bird	1.00	0.95	0.93	0.92	0.90	0.90	0.90	0.90	0.90	0.90
Cat	1.00	0.91	0.86	0.84	0.84	0.84	0.84	0.84	0.84	0.84
Deer	1.00	0.96	0.94	0.94	0.93	0.93	0.93	0.93	0.93	0.93
Dog	1.00	0.98	0.96	0.94	0.94	0.94	0.93	0.93	0.93	0.93
Frog	1.00	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Horse	1.00	0.96	0.96	0.96	0.94	0.94	0.94	0.94	0.94	0.94
Ship	1.00	0.99	0.99	0.99	0.98	0.98	0.98	0.97	0.97	0.97
Truck	1.00	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

CIFAR-100

Predicted Class / Rank (i)	0	1	2	3	4	5	6	7	8	9
Airplane	1.00	0.68	0.64	0.60	0.56	0.56	0.55	0.55	0.55	0.55
Automobile	1.00	0.89	0.81	0.79	0.71	0.67	0.67	0.61	0.61	0.61
Bird	1.00	0.77	0.67	0.64	0.52	0.51	0.51	0.51	0.50	0.50
Cat	1.00	0.62	0.53	0.50	0.47	0.47	0.47	0.46	0.46	0.46
Deer	1.00	0.70	0.61	0.61	0.56	0.56	0.55	0.55	0.55	0.55
Dog	1.00	0.92	0.85	0.79	0.79	0.79	0.76	0.76	0.76	0.76
Frog	1.00	0.91	0.91	0.82	0.82	0.81	0.80	0.80	0.80	0.80
Horse	1.00	0.64	0.59	0.59	0.44	0.43	0.43	0.43	0.43	0.43
Ship	1.00	0.94	0.91	0.88	0.82	0.82	0.81	0.79	0.79	0.79
Truck	1.00	0.85	0.80	0.79	0.76	0.75	0.74	0.74	0.74	0.74

TIN

Predicted Class / Rank (i)	0	1	2	3	4	5	6	7	8	9
Airplane	1.00	0.60	0.54	0.52	0.46	0.46	0.46	0.46	0.46	0.46
Automobile	1.00	0.91	0.79	0.72	0.64	0.57	0.57	0.47	0.47	0.47
Bird	1.00	0.75	0.66	0.60	0.46	0.45	0.45	0.44	0.44	0.44
Cat	1.00	0.55	0.48	0.45	0.43	0.43	0.43	0.43	0.43	0.43
Deer	1.00	0.63	0.55	0.55	0.52	0.51	0.51	0.50	0.50	0.50
Dog	1.00	0.92	0.83	0.75	0.74	0.74	0.68	0.68	0.68	0.68
Frog	1.00	0.90	0.89	0.75	0.75	0.74	0.73	0.73	0.73	0.73
Horse	1.00	0.60	0.54	0.54	0.40	0.40	0.40	0.40	0.39	0.39
Ship	1.00	0.94	0.90	0.86	0.77	0.76	0.76	0.73	0.73	0.73
Truck	1.00	0.77	0.74	0.70	0.68	0.68	0.67	0.67	0.66	0.66

MNIST

Predicted Class / Rank (i)	0	1	2	3	4	5	6	7	8	9
Airplane	1.00	0.65	0.62	0.56	0.45	0.45	0.45	0.45	0.45	0.45
Automobile	1.00	0.52	0.47	0.41	0.40	0.32	0.32	0.28	0.28	0.28
Bird	1.00	0.93	0.90	0.82	0.64	0.63	0.62	0.62	0.62	0.62
Cat	1.00	0.49	0.41	0.34	0.27	0.27	0.26	0.26	0.26	0.26
Deer	1.00	0.16	0.10	0.10	0.08	0.08	0.08	0.08	0.08	0.08
Dog	1.00	0.95	0.83	0.79	0.78	0.78	0.75	0.75	0.75	0.75
Frog	1.00	0.11	0.11	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Horse	1.00	0.81	0.56	0.56	0.28	0.28	0.28	0.28	0.28	0.28
Ship	1.00	0.94	0.91	0.86	0.80	0.80	0.80	0.77	0.77	0.77
Truck	1.00	0.80	0.73	0.67	0.55	0.54	0.54	0.54	0.54	0.54

SVHN

Predicted Class / Rank (i)	0	1	2	3	4	5	6	7	8	9
Airplane	1.00	0.97	0.95	0.91	0.78	0.78	0.76	0.76	0.76	0.76
Automobile	1.00	0.64	0.64	0.45	0.18	0.09	0.09	0.09	0.09	0.09
Bird	1.00	0.79	0.58	0.52	0.34	0.33	0.33	0.33	0.33	0.33
Cat	1.00	0.56	0.54	0.45	0.43	0.43	0.43	0.42	0.42	0.42
Deer	1.00	0.33	0.25	0.25	0.23	0.22	0.22	0.22	0.22	0.22
Dog	1.00	0.98	0.83	0.74	0.74	0.73	0.72	0.72	0.72	0.72
Frog	1.00	0.85	0.85	0.80	0.80	0.73	0.73	0.67	0.65	0.65
Horse	1.00	0.31	0.22	0.22	0.03	0.03	0.03	0.03	0.03	0.03
Ship	1.00	0.83	0.63	0.47	0.34	0.34	0.34	0.34	0.34	0.34
Truck	1.00	0.86	0.71	0.67	0.62	0.62	0.52	0.48	0.48	0.48

Textures

Predicted Class / Rank (i)	0	1	2	3	4	5	6	7	8	9
Airplane	1.00	0.58	0.51	0.49	0.44	0.44	0.42	0.42	0.42	0.42
Automobile	1.00	0.90	0.77	0.73	0.62	0.55	0.55	0.47	0.47	0.47
Bird	1.00	0.67	0.53	0.49	0.28	0.27	0.26	0.26	0.26	0.26
Cat	1.00	0.56	0.52	0.48	0.45	0.45	0.44	0.44	0.44	0.44
Deer	1.00	0.50	0.47	0.47	0.46	0.45	0.45	0.45	0.45	0.45
Dog	1.00	0.94	0.88	0.77	0.73	0.71	0.69	0.69	0.69	0.69
Frog	1.00	0.91	0.89	0.70	0.70	0.68	0.67	0.67	0.67	0.67
Horse	1.00	0.64	0.53	0.53	0.34	0.32	0.32	0.32	0.32	0.32
Ship	1.00	0.87	0.74	0.67	0.61	0.61	0.60	0.59	0.59	0.59
Truck	1.00	0.80	0.74	0.73	0.65	0.65	0.64	0.64	0.64	0.64

Places365

Predicted Class / Rank (i)	0	1	2	3	4	5	6	7	8	9
Airplane	1.00	0.49	0.45	0.43	0.39	0.39	0.39	0.39	0.39	0.39
Automobile	1.00	0.88	0.72	0.70	0.59	0.55	0.55	0.41	0.41	0.41
Bird	1.00	0.71	0.63	0.56	0.43	0.41	0.41	0.41	0.41	0.41
Cat	1.00	0.48	0.40	0.38	0.36	0.36	0.35	0.35	0.35	0.35
Deer	1.00	0.62	0.52	0.51	0.46	0.43	0.43	0.42	0.42	0.42
Dog	1.00	0.84	0.77	0.71	0.70	0.69	0.62	0.62	0.62	0.62
Frog	1.00	0.85	0.84	0.62	0.62	0.61	0.61	0.61	0.60	0.60
Horse	1.00	0.59	0.53	0.53	0.44	0.43	0.43	0.43	0.42	0.42
Ship	1.00	0.95	0.94	0.92	0.84	0.83	0.83	0.80	0.80	0.80
Truck	1.00	0.78	0.76	0.73	0.70	0.69	0.68	0.68	0.68	0.68

Figure 3. CIFAR-10 Conditional probability matrix (CP) of rank position i given that all prior ranks have been correctly predicted

References

[1] Abhijit Bendale et al. Towards open set deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1563–1572, 2016. 2, 3, 4, 5, 6, 7, 8, 9

[2] Guangyao Chen, Peixi Peng, Xiangqian Wang, and Yonghong Tian. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8065–8081, 2021. 2, 3, 4, 5, 6, 7, 8, 9

[3] Terrance DeVries et al. Learning confidence for out-of-distribution detection in neural networks. *arXiv preprint arXiv:1802.04865*, 2018. 2, 3, 4, 5, 6, 7, 8, 9

[4] Andrija Djurisic, Nebojsa Bozanic, Arjun Ashok, and Rosanne Liu. Extremely simple activation shaping for out-of-distribution detection. In *The Eleventh International Conference on Learning Representations*, 2023. 2, 3, 4, 5, 6, 7, 8, 9

[5] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, page 1321–1330. JMLR.org, 2017. 2, 3, 4, 5, 6, 7, 8, 9

[6] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2017. 2, 3, 4, 5, 6, 7, 8, 9

[7] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *International Conference on Learning Representations*, 2019. 2, 3, 4, 5, 6, 7, 8, 9

[8] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joseph Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, pages 8759–8773. PMLR, 2022. 2, 3, 4, 5, 6, 7, 8, 9

[9] Yen-Chang Hsu, Yilin Shen, Hongxia Jin, and Zsolt Kira. Generalized odin: Detecting out-of-distribution image without learning from out-of-distribution data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10951–10960, 2020. 2, 3, 4, 5, 6, 7, 8, 9

[10] Rui Huang et al. On the importance of gradients for detecting distributional shifts in the wild. *Advances in Neural Information Processing Systems*, 34:677–689, 2021. 2, 3, 4, 5, 6, 7, 8, 9

[11] Rui Huang et al. Mos: Towards scaling out-of-distribution detection for large semantic space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8710–8719, 2021. 2, 3, 4, 5, 6, 7, 8, 9

[12] Naveen Karunanayake, Suranga Seneviratne, and Sanjay Chawla. Excel: Combined extreme and collective logit information for enhancing out-of-distribution detection. *arXiv preprint arXiv:2311.14754*, 2023. 2, 3, 4, 5, 6, 7, 8, 9

[13] Naveen Karunanayake, Suranga Seneviratne, and Sanjay Chawla. Craft: Class ranking aware fine-tuning for enhanced out-of-distribution detection. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4119–4128. IEEE, 2025. 2, 3, 4, 5, 6, 7, 8, 9

[14] Shu Kong et al. Opengan: Open-set recognition via open data generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 813–822, 2021. 2, 3, 4, 5, 6, 7, 8, 9

[15] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *Advances in Neural Information Processing Systems*, 2018. 2, 3, 4, 5, 6, 7, 8, 9

[16] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *International Conference on Learning Representations*, 2018. 2, 3, 4, 5, 6, 7, 8, 9

[17] Weitang Liu, Xiaoyun Wang, John D. Owens, and Yixuan Li. Energy-based out-of-distribution detection. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2020. 2, 3, 4, 5, 6, 7, 8, 9

[18] Xixi Liu et al. Gen: Pushing the limits of softmax-based out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23946–23955, 2023. 2, 3, 4, 5, 6, 7, 8, 9

[19] Yifei Ming, Yiyou Sun, Ousmane Dia, and Yixuan Li. How to exploit hyperspherical embeddings for out-of-distribution detection? In *The Eleventh International Conference on Learning Representations*, 2023. 2, 3, 4, 5, 6, 7, 8, 9

[20] Jie Ren, Stanislav Fort, Jeremiah Liu, Abhijit Guha Roy, Shreyas Padhy, and Balaji Lakshminarayanan. A simple fix to mahalanobis distance for improving near-ood detection. *arXiv preprint arXiv:2106.09022*, 2021. 2, 3, 4, 5, 6, 7, 8, 9

[21] Chandramouli Shama Sastry et al. Detecting out-of-distribution examples with gram matrices. In *International Conference on Machine Learning*, pages 8491–8501. PMLR, 2020. 2, 4, 5, 6, 7, 8, 9

[22] Yue Song et al. Rankfeat: Rank-1 feature removal for out-of-distribution detection. *Advances in Neural Information Processing Systems*, 35:17885–17898, 2022. 2, 3, 4, 5, 6, 7, 8, 9

[23] Yiyou Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In *International Conference on Machine Learning*, pages 20827–20840. PMLR, 2022. 2, 3, 4, 5, 6, 7, 8, 9

[24] Yiyou Sun et al. React: Out-of-distribution detection with rectified activations. *Advances in Neural Information Processing Systems*, 34:144–157, 2021. 2, 3, 4, 5, 6, 7, 8, 9

[25] Yiyou Sun et al. Dice: Leveraging sparsification for out-of-distribution detection. In *European Conference on Computer Vision*, pages 691–708. Springer, 2022. 2, 3, 4, 5, 6, 7, 8, 9

[26] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in Neural Information Processing Systems*, 33:11839–11852, 2020. 2, 3, 4, 5, 6, 7, 8, 9

[27] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. Vim: Out-of-distribution with virtual-logit matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4921–4930, 2022. 2, 3, 4, 5, 6, 7, 8, 9

[28] Hongxin Wei, Renchunzi Xie, Hao Cheng, Lei Feng, Bo An, and Yixuan Li. Mitigating neural network overconfidence with logit normalization. In *International Conference on Machine Learning*, pages 23631–23644. PMLR, 2022. 2, 3, 4, 5, 6, 7, 8, 9

[29] Jingkang Yang, Haoqi Wang, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang, and Ziwei Liu. Semantically coherent out-of-distribution detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8301–8309, 2021. 2, 3, 4, 5, 6, 7, 8, 9

[30] Qing Yu et al. Unsupervised out-of-distribution detection by maximum classifier discrepancy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9518–9526, 2019. 2, 3, 4, 5, 6, 7, 8, 9

[31] Jinsong Zhang, Qiang Fu, Xu Chen, Lun Du, Zelin Li, Gang Wang, Shi Han, Dongmei Zhang, et al. Out-of-distribution detection based on in-distribution data patterns memorization with modern hopfield energy. In *The Eleventh International Conference on Learning Representations*, 2022. 2, 3, 4, 5, 6, 7, 8, 9

[32] Jingyang Zhang, Nathan Inkawhich, Randolph Linderman, Yiran Chen, and Hai Li. Mixture outlier exposure: Towards out-of-distribution detection in fine-grained environments. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5531–5540, 2023. 2, 3, 4, 5, 6, 7, 8, 9