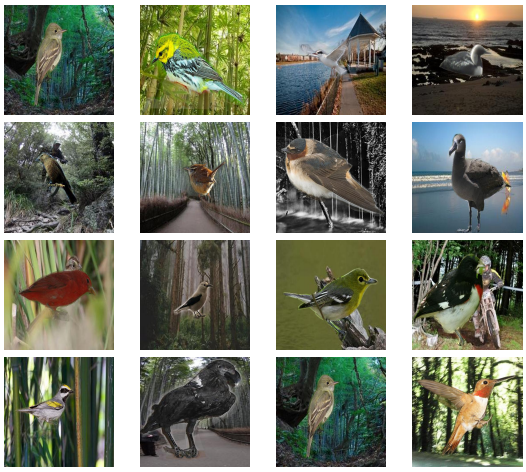
Image-based Outlier Synthesis With Training Data

Supplementary Material

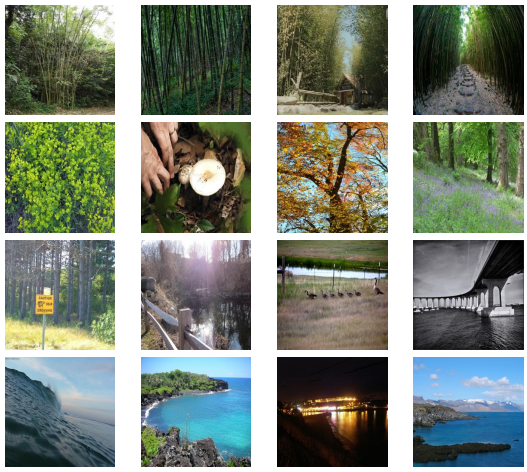
A. Datasets

A.1. Spurious setting

We use Waterbirds and CelebA as ID datasets for the OOD evaluation in the spurious setting.



(a) ID images of Waterbirds datasets.



(b) Spurious OOD images of Waterbirds datasets.

Figure 6. Examples of ID and spurious OOD images of Waterbirds datasets.

Waterbirds: Waterbirds $\mathbb{D}_{\text{in}}$ [69] is constructed from the combination of photographs from the Caltech-UCSD Birds-200-2011 (CUB) datasets [82] $\mathcal{X}_{\text{inv}}$ with image backgrounds $\mathcal{E}$ from the Places datasets [108]. Each bird is labeled as $\mathcal{Y} = \{\texttt{waterbird}, \texttt{landbird}\}$ and placed on an environment $\mathcal{E} = \{\texttt{water background}, \texttt{land background}\}$, i.e., $\mathbf{x} = \psi(\mathbf{x}_{\text{inv}}, \mathbf{e})$. Images of land and water alone are considered spurious OOD, i.e., $\mathbf{x}' = \psi(\mathbf{x}'_{\text{inv}}, \mathbf{e})$, where $\mathbf{x}'_{\text{inv}} = \emptyset$. The correlation between land (water) and landbird (waterbird) in the training set is set to ~ 0.9. The examples of ID and spurious OOD images of the Waterbirds datasets are presented in Figure 6.

(a) ID images of CelebA datasets.

(b) Spurious OOD images of CelebA datasets.

Figure 7. Examples of ID and spurious OOD images of CelebA datasets.

CelebA: CelebA $\mathbb{D}_{\text{in}}$ [51] is a large-scale face attributes datasets containing hair color attributes $\{\texttt{grey}, \texttt{non-grey}\}$. We consider the label space as $\mathcal{Y} = \{\texttt{grey hair}, \texttt{nongrey hair}\}$. The environments $\mathcal{E} = \{\texttt{male}, \texttt{female}\}$ denote the gender of the person. The correlation between `grey hair` and `male` gender (environmental feature) is ~ 0.8 in the training set. Spurious OOD inputs consist of `bald male`, which contain environmental feature (gender) without invariant feature (hair).

The examples of ID and spurious OOD images of CelebA datasets are presented in Figure 7. For a detailed formalization of ID and spurious OOD of Waterbirds and CelebA datasets, we refer readers to Ming *et al*. [56].

A.2. Fine-grained setting

In this setting, ID/OOD splits are established through a holdout class approach. Specifically, a subset of categories is designated as ID, while the remaining categories are excluded from the training set and treated as OOD during testing.

(a) ID images of Aircraft datasets.

(b) Fine-grained OOD images of Aircraft datasets.

Figure 8. Examples of ID and fine-grained OOD images of Aircraft datasets.

Aircraft: We consider 100 classes of the Aircraft datasets. Out of these, 10 classes are randomly selected to serve as fine-grained OOD data, while the remaining classes constitute the ID set. The ID and OOD classes are:

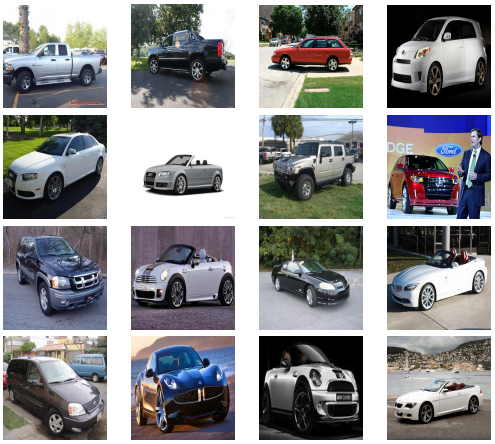
- ID classes: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 39, 41, 42, 43, 44, 45, 46, 47, 48, 51, 52, 53, 54, 55, 56, 57, 59, 60, 61, 62, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 97, 98, 99

- OOD classes: 15, 17, 32, 38, 40, 49, 50, 58, 63, 96

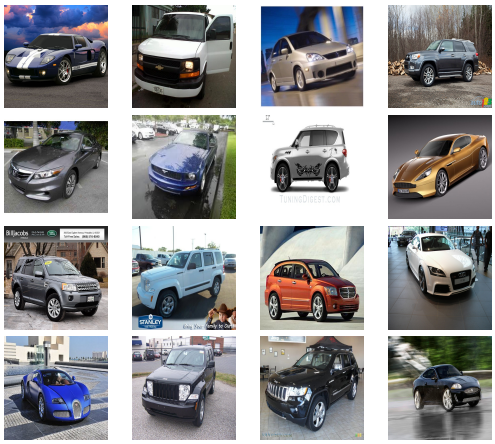
The examples of ID and fine-grained OOD images of the Aircraft datasets are presented in Figure 8.

Car: The Car datasets include 196 classes. Among these, 46 classes are randomly assigned as fine-grained OOD data, with the rest retained as ID classes. The ID and OOD classes are:

- ID classes: 0, 1, 2, 3, 4, 5, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 41, 42, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 58, 59, 60, 61, 62, 63, 64, 65, 66, 68, 69, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 83, 84, 86, 88, 89, 90, 91, 92, 94, 95, 96, 97, 98, 100, 102, 104, 105, 107, 108, 109, 110, 112, 113, 114, 115, 116, 119, 120, 122, 123, 124, 125, 126, 128, 130, 131, 132, 133, 134, 135, 137, 138, 139, 142, 144, 145, 149, 150, 151, 152, 153, 155, 156, 157, 158, 159, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 174, 175, 177, 178, 179, 181, 183, 184, 185, 187, 189, 190, 191, 192, 193, 194, 195



(a) ID images of Car datasets.



(b) Fine-grained OOD images of Car datasets.

Figure 9. Examples of ID and fine-grained OOD images of Car datasets.

- OOD classes: 87, 67, 82, 8, 106, 121, 43, 182, 93, 180, 101, 7, 70, 85, 34, 117, 186, 10, 136, 148, 118, 176, 57, 13, 25, 154, 45, 140, 111, 143, 99, 9, 141, 160, 24, 188, 161, 129, 103, 147, 6, 44, 127, 146, 50, 173

The examples of ID and fine-grained OOD images of the Car datasets are presented in Figure 9.

B. i-ODIN: invariant-ODIN

The softmax probability for class i incorporating temperature hyperparameter $T > 0$ is:

$$S_i(\mathbf{x}; T) = \frac{\exp(f_i(\phi_\gamma(\mathbf{x}))/T)}{\sum_{j=1}^{C} \exp(f_j(\phi_\gamma(\mathbf{x}))/T)} \tag{6}$$

The confidence score is given by $S(\mathbf{x}; T) = \max_i S_i(\mathbf{x}; T)$. ODIN improves OOD detection via two components:

1. Temperature Scaling: Adjusting T sharpens the probability distribution, enhancing ID-OOD separability.
2. Input Preprocessing: Small perturbations refine the confidence gap:

$$\tilde{\mathbf{x}} = \mathbf{x} - \varepsilon \cdot \text{sign}(-\nabla_{\mathbf{x}} \log S(\mathbf{x}; T)) \tag{7}$$

where ε controls perturbation magnitude, increasing softmax scores more for ID than OOD samples.

We propose i-ODIN, an enhanced variant of ODIN that improves OOD detection performance through selective perturbation application. While ODIN applies perturbations across the entire input, i-ODIN introduces a masking mechanism that focuses perturbations on the most discriminative regions. i-ODIN follows the same overall framework as ODIN but introduces a critical modification in the input preprocessing stage. Given input $\mathbf{x} \in \mathcal{X}$, gradients $\nabla_{\mathbf{x}} \log S(\mathbf{x}; T)$ are computed, and the top-k influential features are selected:

1. Generate mask $\mathcal{M}$ for the top-$p_{\text{inv}}\%$ gradient magnitudes:

$$\mathcal{M} = \text{TopK}(|\nabla_{\mathbf{x}} \log S(\mathbf{x}; T)|,\ p_{\text{inv}}\%) \tag{8}$$

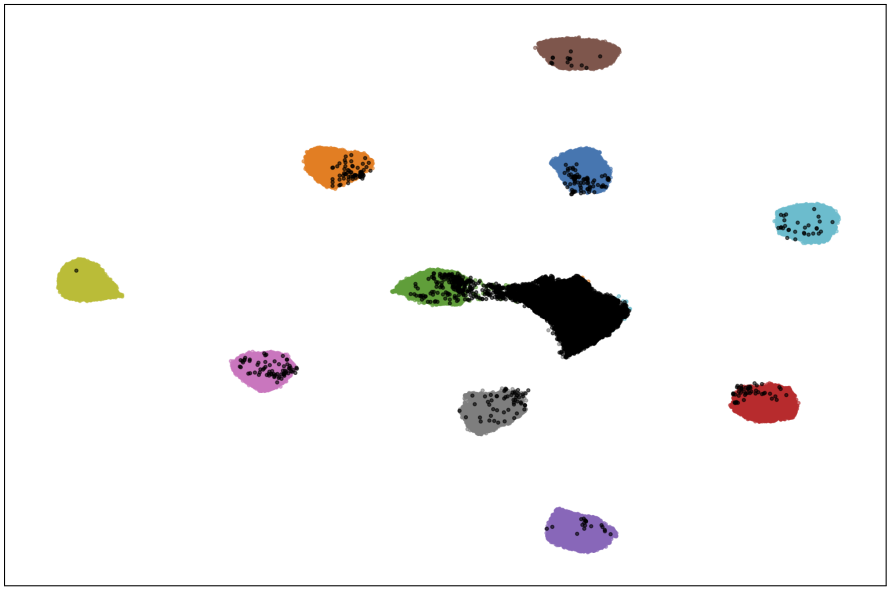
2. Apply perturbation selectively:

$$\tilde{\mathbf{x}} = \mathbf{x} - \varepsilon \cdot \text{sign}(-\nabla_{\mathbf{x}} \log S(\mathbf{x}; T)) \odot \mathcal{M} \tag{9}$$

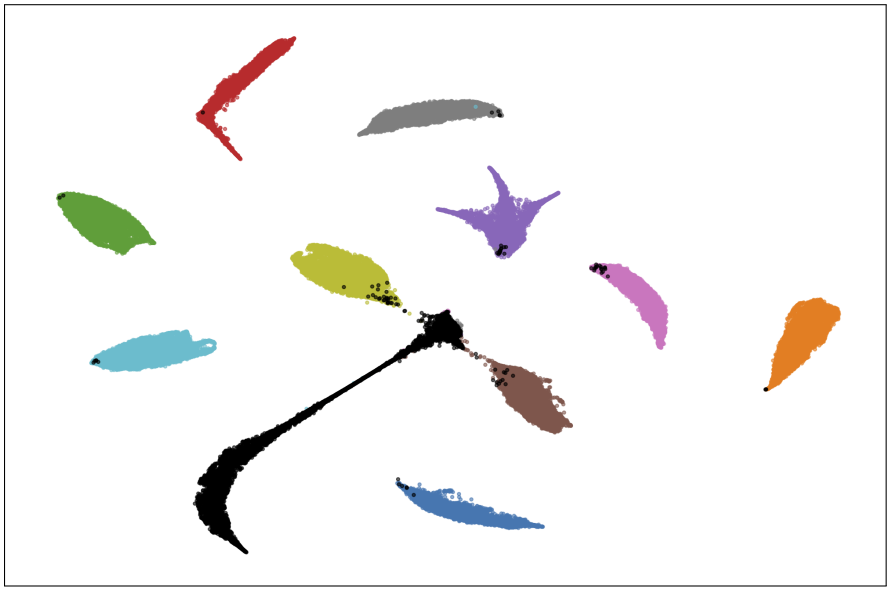
This targeted approach enhances ID-OOD separation by focusing on the most relevant features.

C. UMAP Visualization of feature space CIFAR-10 datasets and Conventional OOD (iSUN)

The UMAP visualizations of the feature space of a ResNet-18 model trained in the CIFAR-10 datasets using (a) the cross-entropy baseline and (b) ASCOOD (with random pixel shuffle perturbation for outlier synthesis) are presented in the Figure 10a and Figure 10b respectively. Samples from the Conventional OOD datasets (iSUN) are visualized in black color within each plot while the rest of the colors denote samples from various classes of ID datasets. The visualizations indicate a reduced intersection between OOD and ID samples in the case of ASCOOD compared to the cross-entropy baseline. It suggests that ASCOOD effectively maps OOD samples further from the ID distribution. This separation ultimately contributes to an enhancement in OOD detection performance.



(a) Cross-entropy baseline



(b) ASCOOD

Figure 10. Comparison of UMAP visualizations: (a) Cross-entropy baseline and (b) ASCOOD. OOD samples are in black color, while other colors represent ID samples. The overlap between ID and OOD is reduced in (b) ASCOOD compared to (a) Cross-entropy baseline.

D. Additional Details and Experiments

We use `PyTorch` deep learning library and `torchvision` package to conduct all the experiments. For image preprocessing in spurious and fine-grained settings, we use `RandomResizedCrop` and `RandomHorizontalFlip` transforms. For conventional setting, we use `RandomHorizontalFlip` and `RandomCrop` transforms. We deal with the spatial dimensions of $(32, 32)$, $(224, 224)$, and $(448, 448)$ in conventional, spurious, and fine-grained settings respectively. Additionally, we also deal with ImageNet-100 datasets in (224, 224) spatial dimension.

OOD detection postprocessor. Unless otherwise noted, we use ODIN and SCALE postprocessors for spurious and conventional settings, and NNGuide and Relation postprocessors for Aircraft and Car benchmarks in ASCOOD experiments. In the case of the ODIN postprocessor, we perform hyperparameter tuning using noise values $\{0.0014, 0.0028, 0.0042, 0.0056, 0.0070, 0.0084, 0.0098\}$ different from just $\{0.0014, 0.0028\}$ specified in OpenOOD for spurious setting. Consistent with OpenOOD, hyperparameters are optimized with respect to the validation OOD datasets (AUROC metric), and model selection is done with respect to validation accuracy.

`Shuffle` perturbation: In this work, we utilize pixel shuffle perturbation for outlier synthesis. Generally, pixel shuffle perturbation requires only the 2D pixel locations to perform shuffling. However, when applying pixel attribution method (using $\mathbf{G}$) to an input image with C' channels, the resulting attribution output contains C' channels containing both positive and negative values. To derive the required 2D pixel locations, we transform $\mathbf{G}$ into a 2D map by computing $\sum_{C'} \exp(\mathbf{G})$. We then select the top $p_{\text{inv}}\%$ of pixels with the highest magnitudes from the reduced 2D map for shuffling.

ASCOOD hyperparameters: The initial learning rate of FC layer is set to 0.005 for the CIFAR-100 datasets. The final hyperparameters/setup for each datasets is provided in Table 7. Additional ablation studies are presented in subsequent sections.

Hyperparameter/Setup	Waterbirds	CelebA	Car	Aircraft	CIFAR-10	CIFAR-100	ImageNet-100
σ	0.5	0.5	0.5	0.5	0.5	0.5	0.5
λ	0.1	1.0	1.0	1.0	1.0	5.0	1.0
p_{inv}	10%	5%	10%	10%	20%	10%	10%
α ($\mathbf{G}_{\text{inv}}$)	300→30	50→30	0.1	0.1	10	10	10
α (Gaussian noise)	0.1	0.1	N/A	N/A	0.1	0.01	N/A
Outlier synthesis	$\mathbf{G}_{\text{inv}}$ addition	Gaussian noise addition	$\mathbf{G}_{\text{inv}}$ addition	$\mathbf{G}_{\text{inv}}$ addition	Random pixel shuffle	Invariant pixel shuffle	$\mathbf{G}_{\text{inv}}$ addition

Table 7. Final hyperparameter settings for each datasets. (N/A indicates that the corresponding experiment was not conducted/needed.)

D.1. ASCOOD vs. OE methods

We further evaluate the OOD detection performance of our method ASCOOD against established outlier exposure techniques (OE and MixOE) on the CIFAR-10 and CIFAR-100 benchmarks. Consistent with OpenOOD v1.5, we employ Tiny ImageNet as an external $\mathbb{D}_{\text{out}}$ for OE and MixOE to promote predictive uncertainty towards OOD samples. We train both OE and MixOE from scratch, unlike the approach taken in OpenOOD v1.5. The results presented in Table 8 and Table 9 demonstrate that our method ASCOOD maintains superior performance on these CIFAR benchmarks. While OE achieves superior performance on some individual datasets within the CIFAR-10 benchmark, our method (ASCOOD) consistently outperforms both OE and MixOE across all datasets in the more challenging CIFAR-100 benchmark.

Method	CIFAR-10					
	MNIST	SVHN	iSUN	Texture	Places365	Average
OE	33.29$_{\pm 8.75}$ / 86.43$_{\pm 3.94}$	**2.69**$_{\pm 1.54}$ / **99.13**$_{\pm 0.47}$	37.95$_{\pm 11.98}$ / 83.81$_{\pm 5.45}$	**10.74**$_{\pm 2.65}$ / 97.70$_{\pm 0.51}$	**12.82**$_{\pm 2.46}$ / **97.01**$_{\pm 0.73}$	19.50$_{\pm 3.72}$ / 92.82$_{\pm 1.80}$
MixOE	39.29$_{\pm 17.26}$ / 92.86$_{\pm 2.18}$	34.42$_{\pm 14.02}$ / 93.37$_{\pm 2.41}$	32.99$_{\pm 6.91}$ / 93.74$_{\pm 0.60}$	82.73$_{\pm 2.85}$ / 83.72$_{\pm 0.57}$	71.03$_{\pm 1.24}$ / 86.73$_{\pm 0.93}$	52.09$_{\pm 6.20}$ / 90.09$_{\pm 0.64}$
ASCOOD	**1.37**$_{\pm 0.35}$ / **99.68**$_{\pm 0.11}$	6.73$_{\pm 1.37}$ / 98.61$_{\pm 0.34}$	**0.94**$_{\pm 0.19}$ / **99.79**$_{\pm 0.05}$	10.93$_{\pm 0.52}$ / **97.95**$_{\pm 0.06}$	18.47$_{\pm 1.12}$ / 95.56$_{\pm 0.34}$	**7.69**$_{\pm 0.29}$ / **98.32**$_{\pm 0.07}$

Table 8. ASCOOD vs OE methods in FPR@95↓ / AUROC↑ on CIFAR-10 benchmark.

Method	CIFAR-100					
	MNIST	SVHN	iSUN	Texture	Places365	Average
OE	20.21$_{\pm 4.43}$ / 95.04$_{\pm 0.74}$	40.80$_{\pm 4.65}$ / 88.99$_{\pm 1.09}$	50.89$_{\pm 19.81}$ / 87.35$_{\pm 5.83}$	56.14$_{\pm 4.07}$ / 85.21$_{\pm 0.96}$	66.05$_{\pm 5.77}$ / 77.60$_{\pm 1.88}$	46.82$_{\pm 4.65}$ / 86.84$_{\pm 1.44}$
MixOE	68.92$_{\pm 6.85}$ / 74.86$_{\pm 3.88}$	61.62$_{\pm 16.37}$ / 77.06$_{\pm 7.63}$	68.69$_{\pm 6.71}$ / 72.69$_{\pm 5.06}$	72.58$_{\pm 3.37}$ / 74.69$_{\pm 2.19}$	71.97$_{\pm 1.01}$ / 74.98$_{\pm 0.30}$	68.76$_{\pm 3.98}$ / 74.86$_{\pm 2.42}$
ASCOOD	**6.76**$_{\pm 1.50}$ / **98.58**$_{\pm 0.27}$	**33.50**$_{\pm 3.76}$ / **89.93**$_{\pm 1.24}$	**1.88**$_{\pm 0.56}$ / **99.37**$_{\pm 0.09}$	**53.06**$_{\pm 0.81}$ / **86.98**$_{\pm 0.81}$	**54.30**$_{\pm 1.00}$ / **81.91**$_{\pm 0.77}$	**29.90**$_{\pm 0.76}$ / **91.35**$_{\pm 0.13}$

Table 9. ASCOOD vs OE methods in FPR@95↓ / AUROC↑ on CIFAR-100 benchmark.

D.2. Ablation study of outlier synthesis in fine-grained settings

Outlier synthesis	p_{inv}	Aircraft				Car			
		Fine-grained OOD		Conventional OOD		Fine-grained OOD		Conventional OOD	
		FPR@95↓	AUROC↑	FPR@95↓	AUROC↑	FPR@95↓	AUROC↑	FPR@95↓	AUROC↑
Invariant pixel shuffle	5%	48.63$_{\pm 7.37}$	83.75$_{\pm 1.71}$	1.34$_{\pm 0.09}$	99.68$_{\pm 0.03}$	65.23$_{\pm 0.82}$	85.51$_{\pm 0.30}$	**1.33**$_{\pm 0.18}$	**99.71**$_{\pm 0.02}$
	10%	47.98$_{\pm 2.04}$	83.87$_{\pm 0.38}$	1.19$_{\pm 0.23}$	99.69$_{\pm 0.07}$	63.84$_{\pm 3.21}$	85.38$_{\pm 0.29}$	2.09$_{\pm 0.04}$	99.59$_{\pm 0.02}$
	20%	**47.35**$_{\pm 4.82}$	83.40$_{\pm 0.74}$	1.24$_{\pm 0.09}$	99.69$_{\pm 0.02}$	62.53$_{\pm 2.97}$	85.48$_{\pm 0.68}$	1.90$_{\pm 0.42}$	99.62$_{\pm 0.06}$
$\mathbf{G}_{\text{inv}}$ addition	1%	59.40$_{\pm 2.24}$	88.25$_{\pm 0.42}$	0.60$_{\pm 0.10}$	99.81$_{\pm 0.02}$	46.24$_{\pm 2.54}$	91.02$_{\pm 0.27}$	4.02$_{\pm 0.31}$	98.85$_{\pm 0.11}$
	5%	56.34$_{\pm 4.31}$	88.32$_{\pm 0.74}$	0.58$_{\pm 0.07}$	99.84$_{\pm 0.02}$	42.96$_{\pm 3.10}$	91.56$_{\pm 0.37}$	3.38$_{\pm 0.23}$	98.92$_{\pm 0.06}$
	10%	47.94$_{\pm 5.38}$	**89.75**$_{\pm 1.01}$	**0.55**$_{\pm 0.07}$	**99.84**$_{\pm 0.02}$	**40.76**$_{\pm 1.13}$	**91.86**$_{\pm 0.20}$	4.28$_{\pm 0.53}$	98.79$_{\pm 0.12}$
	20%	53.13$_{\pm 4.00}$	88.81$_{\pm 0.91}$	0.59$_{\pm 0.08}$	99.83$_{\pm 0.01}$	44.55$_{\pm 1.01}$	91.57$_{\pm 0.06}$	4.59$_{\pm 0.55}$	98.73$_{\pm 0.08}$
	100%	48.45$_{\pm 5.04}$	89.14$_{\pm 0.78}$	0.60$_{\pm 0.10}$	99.81$_{\pm 0.01}$	43.00$_{\pm 3.33}$	91.32$_{\pm 0.40}$	4.20$_{\pm 0.21}$	98.81$_{\pm 0.02}$

Table 10. Ablation study of outlier synthesis methods on Aircraft and Car datasets.

We evaluate two effective outlier synthesis methods: invariant pixel shuffle and $\mathbf{G}_{\text{inv}}$ addition in fine-grained settings (Aircraft and Car datasets). The results summarized in Table 10 present OOD detection performance for Aircraft and Car datasets. The results show that $\mathbf{G}_{\text{inv}}$ addition demonstrates overall superior performance. Since the satisfactorily high accuracy of 89.59% and 94.14% is achieved in Aircraft and Car datasets, the gradient $\mathbf{G}$ becomes highly reliable for outlier synthesis as soon as satisfactory accuracy is achieved. So, it leads to effective perturbation of invariant features. The results also indicate that using an invariant pixel shuffle for outlier synthesis may be effective for conventional OOD detection. Still, its performance falls quite short in fine-grained OOD detection.

D.3. Gradient of the logit versus gradient of softmax probability for outlier synthesis

We investigate how the choice between using gradient of the logit and gradient of softmax probability for outlier synthesis affects OOD detection performance. Table 11 reports results obtained in fine-grained settings using $\mathbf{G}_{\text{inv}}$ addition for outlier synthesis. For gradient of the logit, optimal α is set to 1, while for the gradient of the softmax probability, α is set to 0.1. The results demonstrate that both methods achieve comparable performance with a gradient of the logit slightly performing better in overall performance. Consequently, we simply adopt gradient of the logit for outlier synthesis.

$\mathbf{G}$	p_{inv}	Aircraft					Car				
		Fine-grained OOD		Conventional OOD		Acc	Fine-grained OOD		Conventional OOD		Acc
		FPR@95↓	AUROC↑	FPR@95↓	AUROC↑		FPR@95↓	AUROC↑	FPR@95↓	AUROC↑	
$\mathbf{G} = \frac{\partial \mathbf{z}_c}{\partial \mathbf{x}}$	5%	56.34$_{\pm 4.31}$	88.32$_{\pm 0.74}$	0.58$_{\pm 0.07}$	99.84$_{\pm 0.02}$	88.42$_{\pm 0.13}$%	42.96$_{\pm 3.10}$	91.56$_{\pm 0.37}$	3.38$_{\pm 0.23}$	98.92$_{\pm 0.06}$	93.51$_{\pm 0.08}$%
	10%	**47.94**$_{\pm 5.38}$	**89.75**$_{\pm 1.01}$	**0.55**$_{\pm 0.07}$	**99.84**$_{\pm 0.02}$	**89.61**$_{\pm 0.19}$%	**40.76**$_{\pm 1.13}$	**91.86**$_{\pm 0.20}$	4.28$_{\pm 0.53}$	98.79$_{\pm 0.12}$	**94.20**$_{\pm 0.11}$%
	100%	48.45$_{\pm 5.04}$	89.14$_{\pm 0.78}$	0.60$_{\pm 0.10}$	99.81$_{\pm 0.01}$	88.93$_{\pm 0.07}$%	43.00$_{\pm 3.33}$	91.32$_{\pm 0.40}$	4.20$_{\pm 0.21}$	98.81$_{\pm 0.02}$	94.00$_{\pm 0.12}$%
$\mathbf{G} = \frac{\partial \mathbf{p}_c}{\partial \mathbf{x}}$	5%	51.55$_{\pm 4.12}$	89.07$_{\pm 1.13}$	0.56$_{\pm 0.06}$	99.82$_{\pm 0.01}$	89.46$_{\pm 0.29}$%	44.59$_{\pm 2.47}$	91.37$_{\pm 0.29}$	4.27$_{\pm 0.40}$	98.79$_{\pm 0.09}$	94.15$_{\pm 0.19}$%
	10%	50.07$_{\pm 4.55}$	88.96$_{\pm 0.56}$	0.60$_{\pm 0.04}$	99.82$_{\pm 0.00}$	89.16$_{\pm 0.13}$%	42.11$_{\pm 0.83}$	91.56$_{\pm 0.12}$	4.34$_{\pm 0.69}$	98.00$_{\pm 0.13}$	94.00$_{\pm 0.02}$%
	100%	48.73$_{\pm 3.02}$	89.37$_{\pm 0.44}$	0.59$_{\pm 0.06}$	99.83$_{\pm 0.02}$	89.26$_{\pm 0.25}$%	40.94$_{\pm 2.22}$	91.77$_{\pm 0.23}$	**4.02**$_{\pm 0.37}$	**98.85**$_{\pm 0.14}$	94.14$_{\pm 0.10}$%

Table 11. Gradient of logit vs gradient of softmax probability for outlier synthesis.

D.4. Sensitivity of λ

The sensitivity analysis of λ in Aircraft and Car datasets is shown in Table 12. We use $\mathbf{x}' = \mathbf{x} + \alpha \cdot \mathbf{G}_{\text{inv}}$ for outlier synthesis and report the OOD detection results. The results indicate that improved classification accuracy correlates with enhanced OOD detection performance. Setting $\lambda = 1.0$ leads to better accuracy and enhanced OOD detection simultaneously.

λ	Aircraft					Car				
	Fine-grained OOD		Conventional OOD		Acc	Fine-grained OOD		Conventional OOD		Acc
	FPR@95↓	AUROC↑	FPR@95↓	AUROC↑		FPR@95↓	AUROC↑	FPR@95↓	AUROC↑	
0.1	53.56$_{\pm 4.89}$	88.22$_{\pm 0.98}$	1.13$_{\pm 0.22}$	99.69$_{\pm 0.04}$	88.42$_{\pm 0.13}$%	45.46$_{\pm 2.08}$	90.89$_{\pm 0.28}$	**1.97**$_{\pm 0.25}$	**99.61**$_{\pm 0.04}$	93.51$_{\pm 0.08}$%
1.0	**47.94**$_{\pm 5.38}$	**89.75**$_{\pm 1.01}$	**0.55**$_{\pm 0.07}$	**99.84**$_{\pm 0.02}$	**89.61**$_{\pm 0.19}$%	**40.76**$_{\pm 1.13}$	**91.86**$_{\pm 0.20}$	4.28$_{\pm 0.53}$	98.79$_{\pm 0.12}$	**94.20**$_{\pm 0.11}$%
10.0	58.10$_{\pm 4.82}$	78.79$_{\pm 0.88}$	0.63$_{\pm 0.09}$	99.83$_{\pm 0.01}$	88.93$_{\pm 0.07}$%	99.99$_{\pm 0.01}$	50.00$_{\pm 0.00}$	99.99$_{\pm 0.01}$	50.0$_{\pm 0.00}$	94.00$_{\pm 0.12}$%

Table 12. Ablation study of outlier synthesis methods on Aircraft and Car datasets.

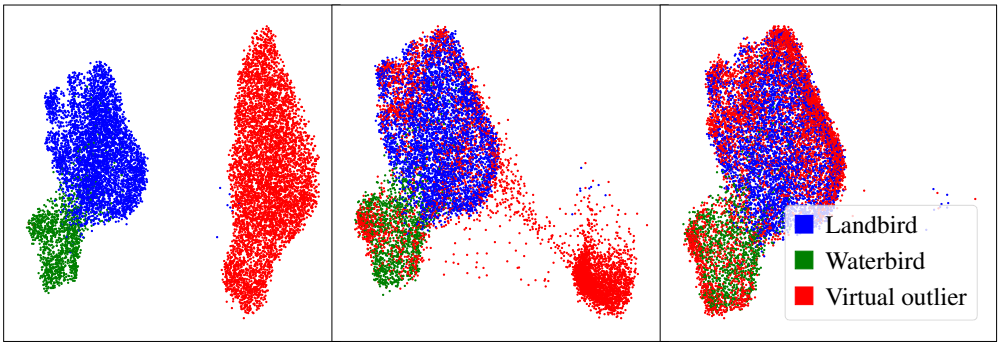
D.5. Gradient addition vs. subtraction for outlier synthesis

The complete result of Table 5 is provided in Tab. 13. We set the hyperparameters as: $\lambda = 1$, $p_{\text{inv}} = 10\%$, and $\alpha = 0.1$. Interestingly, gradient addition for outlier synthesis not only leads to significantly better fine-grained OOD detection performance but also better classification accuracy.

$\mathbf{x}'$	Aircraft					Car				
	Fine-grained OOD		Conventional OOD		Acc	Fine-grained OOD		Conventional OOD		Acc
	FPR@95↓	AUROC↑	FPR@95↓	AUROC↑		FPR@95↓	AUROC↑	FPR@95↓	AUROC↑	
$\mathbf{x}' = \mathbf{x} - \alpha \cdot \mathbf{G}_{\text{inv}}$	50.15$_{\pm 4.80}$	83.64$_{\pm 0.82}$	1.30$_{\pm 0.19}$	99.68$_{\pm 0.04}$	87.43$_{\pm 0.20}$%	60.20$_{\pm 1.91}$	86.27$_{\pm 0.34}$	**2.00**$_{\pm 0.45}$	**99.57**$_{\pm 0.08}$	93.59$_{\pm 0.13}$%
$\mathbf{x}' = \mathbf{x} + \alpha \cdot \mathbf{G}_{\text{inv}}$	**47.94**$_{\pm 5.38}$	**89.75**$_{\pm 1.01}$	**0.55**$_{\pm 0.07}$	**99.84**$_{\pm 0.02}$	**89.61**$_{\pm 0.19}$%	**40.76**$_{\pm 1.13}$	**91.86**$_{\pm 0.20}$	4.28$_{\pm 0.53}$	98.79$_{\pm 0.12}$	**94.20**$_{\pm 0.11}$%

Table 13. Ablation of gradient-based outlier synthesis methods (addition vs. subtraction) on Aircraft and Car datasets.

D.6. Hyperparameter studies in Waterbirds datasets



(a) $\alpha \sim 300$ (b) $\alpha \sim 30$ (c) $\alpha \sim 1$

Figure 11. UMAP visualization of feature space of Waterbirds datasets with varying values of α (used for outlier synthesis).

The hyperparameter α determines the nature of synthesized outliers. To illustrate its effect in feature space, Figure 11a, Figure 11b, and Figure 11c present UMAP visualizations (with cross-entropy baseline on Waterbirds) of virtual outliers alongside ID samples for different values of α. Figure 11a shows a clear separation of virtual outliers from the ID cluster at high α, while Figure 11b demonstrates partial overlap with some outliers remaining distant, reflecting a moderate α. In Figure 11c, virtual outliers and ID samples significantly overlap due to a very low α. Ideally, virtual outliers should be sufficiently challenging but lack invariant features of ID data for optimal spurious OOD detection. The results presented in Table 14b demonstrate that optimal spurious OOD detection is achieved by linearly decreasing α from 300 to 30 during training. Dynamically adjusting α during training – starting high to compensate for $\mathbf{G}$'s initial unreliability and then decreasing to generate more challenging outliers as $\mathbf{G}$ improves – facilitates enhanced spurious OOD detection. This dynamic adjustment, however, did not show gains in fine-grained and conventional OOD detection. The sensitivity of p_{inv} with both dynamic and static values of α is presented in Table 14a, which reinforces the same prior conclusion of dynamic α being superior. It is also evident that α is a more critical hyperparameter than p_{inv}. Regardless of p_{inv} not being highly critical, destroying invariant features while preserving environmental ones still occurs as gradients are higher in invariant regions. Moreover, we also present a sensitivity analysis of σ in Waterbirds datasets in Table 14c. It suggests the optimal value of σ is around 0.5.

α	p_{inv}	Spurious OOD		Conventional OOD	
		FPR@95↓	AUROC↑	FPR@95↓	AUROC↑
30	1%	23.41$_{\pm 2.52}$	94.71$_{\pm 0.41}$	13.98$_{\pm 1.58}$	97.19$_{\pm 0.22}$
	10%	17.26$_{\pm 1.19}$	96.27$_{\pm 0.28}$	10.18$_{\pm 0.54}$	98.05$_{\pm 0.12}$
	30%	18.18$_{\pm 1.13}$	96.20$_{\pm 0.27}$	9.52$_{\pm 0.79}$	98.24$_{\pm 0.15}$
	100%	17.29$_{\pm 1.65}$	96.39$_{\pm 0.27}$	**9.23$_{\pm 0.93}$**	**98.27$_{\pm 0.17}$**
300 → 30	1%	14.50$_{\pm 0.37}$	96.97$_{\pm 0.06}$	12.98$_{\pm 0.79}$	97.49$_{\pm 0.18}$
	10%	**11.37$_{\pm 0.42}$**	**97.48$_{\pm 0.10}$**	10.02$_{\pm 0.38}$	98.04$_{\pm 0.02}$
	30%	11.74$_{\pm 1.19}$	97.42$_{\pm 0.11}$	10.61$_{\pm 1.33}$	97.98$_{\pm 0.16}$
	100%	13.92$_{\pm 1.38}$	97.06$_{\pm 0.36}$	10.89$_{\pm 0.75}$	97.91$_{\pm 0.19}$

(a) Sensitivity of p_{inv}

α	Spurious OOD		Conventional OOD	
	FPR@95↓	AUROC↑	FPR@95↓	AUROC↑
300	17.92$_{\pm 2.20}$	96.37$_{\pm 0.46}$	15.47$_{\pm 1.28}$	97.04$_{\pm 0.19}$
100	15.93$_{\pm 1.81}$	96.86$_{\pm 0.40}$	12.34$_{\pm 0.77}$	97.72$_{\pm 0.18}$
30	17.26$_{\pm 1.19}$	96.27$_{\pm 0.28}$	10.18$_{\pm 0.54}$	98.04$_{\pm 0.12}$
10	35.73$_{\pm 6.31}$	90.61$_{\pm 1.53}$	18.41$_{\pm 1.49}$	96.34$_{\pm 0.23}$
300 → 30	**11.37$_{\pm 0.42}$**	**97.48$_{\pm 0.10}$**	**10.02$_{\pm 0.38}$**	**98.04$_{\pm 0.02}$**

(b) Sensitivity of α

σ	Spurious OOD	
	FPR@95↓	AUROC↑
0.1	34.01$_{\pm 0.70}$	90.21$_{\pm 1.16}$
0.5	**11.37$_{\pm 0.42}$**	**97.48$_{\pm 0.10}$**
2.5	51.75$_{\pm 31.35}$	80.33$_{\pm 18.22}$

(c) Sensitivity of α

Table 14. Sensitivity analysis in Waterbirds datasets

D.7. Hyperparameter studies in CelebA datasets

Table 15 presents the sensitivity analysis of the α hyperparameter in CelebA datasets. Due to the relative simplicity of the ID task, nearly perfect performance in conventional OOD detection is achieved across all values of α. The sensitivity analysis of α (keeping p_{inv} = 5%) and p_{inv} (keeping α = 50 $\rightarrow$ 30) is presented in the Table 15 and Table Table 16 respectively. The results indicate that optimal performance is achieved when α is reduced from 50 to 30 and p_{inv} is set to 5%. Table 17 shows the sensitivity analysis of σ for spurious OOD detection in CelebA ($p_{\text{inv}} = 5\%$) datasets. The results demonstrate that setting σ to 0.5 yields the best OOD detection performance in CelebA datasets. In each study, outliers are synthesized as $\mathbf{x}' = \mathbf{x} + \alpha \cdot \mathbf{G}_{\text{inv}}$.

α	**Spurious OOD**		**Conventional OOD**	
	FPR@95↓	AUROC↑	FPR@95↓	AUROC↑
40	51.90$_{\pm 8.35}$	83.14$_{\pm 1.81}$	**0.01**$_{\pm 0.01}$	**99.99**$_{\pm 0.00}$
30	47.99$_{\pm 7.47}$	83.98$_{\pm 2.01}$	**0.01**$_{\pm 0.02}$	**99.98**$_{\pm 0.00}$
10	50.49$_{\pm 3.86}$	83.09$_{\pm 0.99}$	**0.00**$_{\pm 0.00}$	**99.99**$_{\pm 0.00}$
5.0	55.03$_{\pm 8.16}$	82.16$_{\pm 2.04}$	**0.02**$_{\pm 0.03}$	**99.98**$_{\pm 0.01}$
50 $\rightarrow$ 30	**46.61**$_{\pm 0.63}$	**84.12**$_{\pm 0.55}$	**0.01**$_{\pm 0.01}$	**99.99**$_{\pm 0.00}$

Table 15. Sensitivity of α hyperparameter in CelebA datasets.

Metrics	p_{inv}		
	1%	5%	10%
FPR@95 ↓ / AUROC ↑	49.42$_{\pm 4.25}$ / 83.62$_{\pm 0.55}$	**46.61**$_{\pm 0.63}$ / **84.12**$_{\pm 0.55}$	52.32$_{\pm 5.67}$ / 82.19$_{\pm 2.15}$

Table 16. Sensitivity of p_{inv} in terms of spurious OOD detection in CelebA datasets.

σ	**Spurious OOD**	
	FPR@95↓	AUROC↑
0.1	66.16$_{\pm 0.59}$	66.08$_{\pm 3.24}$
0.5	**46.61**$_{\pm 0.63}$	**84.12**$_{\pm 0.55}$
2.5	61.08$_{\pm 12.02}$	80.89$_{\pm 3.43}$

Table 17. Sensitivity of σ hyperparameter in terms of spurious OOD detection in CelebA datasets.

D.8. Ablation study of outlier synthesis in CIFAR-10/100 datasets

The ablation study of outlier synthesis in CIFAR-10/100 datasets is presented in Table 18. Our study examines various mechanisms for synthesizing outliers and reveals that shuffling invariant pixels yields optimal results for the CIFAR-100 benchmark. Furthermore, we find that even utilizing mere random shuffling of pixels can achieve good OOD detection performance in the context of the relatively easier CIFAR-10 benchmark. This highlights the effectiveness of pixel shuffle perturbation in outlier synthesis. We observe that adding Gaussian noise and $\mathbf{G}_{\text{inv}}$ to ID for synthesizing virtual outliers performs comparably to other methods on the CIFAR-10 benchmark but significantly underperforms in average OOD detection on the CIFAR-100 benchmark. The datasets yielding higher accuracy are more likely to benefit from utilizing the gradient $\mathbf{G}$ addition for outlier synthesis, as increased accuracy suggests reliable outlier synthesis in later training stages. Hence, we hypothesize that using $\mathbf{G}_{\text{inv}}$ addition for outlier synthesis is less effective than mere random pixel shuffle perturbation in CIFAR-100 because of the modest accuracy ($\sim 76.63\%$).

Outlier synthesis	p_{inv}/α	CIFAR-10		CIFAR-100	
		FPR@95↓	AUROC↑	FPR@95↓	AUROC↑
Random pixel shuffle	5%	9.17$_{\pm 0.76}$	97.95$_{\pm 0.19}$	35.05$_{\pm 1.24}$	89.41$_{\pm 0.54}$
	10%	8.67$_{\pm 0.49}$	98.13$_{\pm 0.10}$	34.15$_{\pm 0.77}$	89.13$_{\pm 0.26}$
	15%	10.28$_{\pm 0.76}$	97.79$_{\pm 0.18}$	35.60$_{\pm 2.69}$	89.58$_{\pm 0.93}$
	20%	**7.69$_{\pm 0.29}$**	**98.32$_{\pm 0.07}$**	34.18$_{\pm 1.37}$	90.23$_{\pm 0.88}$
	30%	9.04$_{\pm 0.58}$	98.05$_{\pm 0.10}$	35.02$_{\pm 2.44}$	89.99$_{\pm 0.63}$
	100%	10.04$_{\pm 0.74}$	97.78$_{\pm 0.15}$	36.60$_{\pm 0.90}$	89.85$_{\pm 0.33}$
Invariant pixel shuffle	5%	8.88$_{\pm 0.29}$	98.07$_{\pm 0.07}$	32.47$_{\pm 3.19}$	89.91$_{\pm 1.31}$
	10%	8.09$_{\pm 0.42}$	98.29$_{\pm 0.06}$	**29.90$_{\pm 0.76}$**	**91.35$_{\pm 0.13}$**
	15%	7.84$_{\pm 0.27}$	98.31$_{\pm 0.07}$	32.15$_{\pm 1.30}$	90.42$_{\pm 0.31}$
	20%	8.66$_{\pm 0.55}$	98.14$_{\pm 0.15}$	35.43$_{\pm 1.67}$	89.63$_{\pm 0.87}$
$\mathbf{G}_{\text{inv}}$ addition	1.0	10.43$_{\pm 0.97}$	97.71$_{\pm 0.24}$	46.39$_{\pm 8.48}$	85.56$_{\pm 3.10}$
	10.0	9.28$_{\pm 0.95}$	98.00$_{\pm 0.19}$	39.23$_{\pm 1.63}$	88.31$_{\pm 0.95}$
	100.0	9.99$_{\pm 0.88}$	97.85$_{\pm 0.23}$	48.22$_{\pm 2.11}$	84.60$_{\pm 0.86}$
Gaussian noise addition	0.01	8.74$_{\pm 0.26}$	97.95$_{\pm 0.18}$	42.64$_{\pm 3.98}$	88.33$_{\pm 1.50}$
	0.1	8.61$_{\pm 0.25}$	98.10$_{\pm 0.05}$	43.81$_{\pm 5.01}$	86.48$_{\pm 1.44}$
	1.0	12.58$_{\pm 1.75}$	97.12$_{\pm 0.43}$	47.34$_{\pm 3.92}$	84.86$_{\pm 1.20}$
	10.0	13.90$_{\pm 1.21}$	96.75$_{\pm 0.28}$	53.82$_{\pm 2.08}$	81.40$_{\pm 1.32}$

Table 18. Ablation study of outlier synthesis in CIFAR-10/100 datasets.

D.9. Ablation study of outlier synthesis in Waterbirds datasets

Table 19 presents an ablation study of outlier synthesis approaches in the Waterbirds datasets. From the results, we can observe a trend in both random pixel shuffle and invariant pixel shuffle: as the level of pixel shuffling increases, the presence of ID environmental features in the outliers diminishes, making the outliers increasingly trivial. For a given level of pixel shuffling p_{inv}, shuffling invariant pixels derived from $\mathbf{G}_{\text{inv}}$ results in superior performance compared to random pixel shuffling. This indicates that $\mathbf{G}_{\text{inv}}$, as the model continues learning ID discrimination, indicates pixels most crucial for class recognition. Shuffling these pixels effectively disrupts the semantic content synthesizing reliable outliers. Moreover, synthesizing outliers by adding Gaussian noise to ID data leads to a relative decline in OOD detection performance. Gaussian noise introduces extraneous information and results in relatively more distortion of environmental/background features within the synthesized outliers. However, preserving environmental/background features in these outliers is critical in mitigating the effect of high spurious correlations in the training set. Hence, the outlier synthesis approach resulting in minimal alteration in environmental features is favorable in Waterbirds datasets.

OOD type	Random pixel shuffle			Invariant pixel shuffle			Gaussian noise		
	p_{inv}	FPR@95↓	AUROC↑	p_{inv}	FPR@95↓	AUROC↑	α	FPR@95↓	AUROC↑
Spurious	5%	**19.48**$_{\pm 1.18}$	**96.49**$_{\pm 0.10}$	5%	**13.21**$_{\pm 1.24}$	**97.43**$_{\pm 0.16}$	0.01	77.37$_{\pm 3.56}$	74.52$_{\pm 3.23}$
	10%	24.32$_{\pm 2.09}$	95.33$_{\pm 0.39}$	10%	14.80$_{\pm 0.19}$	97.26$_{\pm 0.05}$	0.05	**36.48**$_{\pm 6.03}$	89.89$_{\pm 2.37}$
	15%	28.71$_{\pm 2.45}$	94.32$_{\pm 0.35}$	15%	17.63$_{\pm 3.78}$	96.73$_{\pm 0.60}$	0.1	40.44$_{\pm 6.72}$	90.75$_{\pm 1.81}$
	20%	30.23$_{\pm 3.79}$	93.87$_{\pm 0.82}$	20%	20.91$_{\pm 1.20}$	96.17$_{\pm 0.20}$	0.5	37.86$_{\pm 4.93}$	**91.71**$_{\pm 1.38}$
	30%	32.43$_{\pm 3.45}$	92.81$_{\pm 1.02}$	30%	28.02$_{\pm 1.76}$	94.50$_{\pm 0.61}$	1.0	40.47$_{\pm 2.21}$	90.48$_{\pm 0.56}$
	100%	45.16$_{\pm 2.75}$	89.33$_{\pm 1.22}$	100%	45.16$_{\pm 2.75}$	89.33$_{\pm 1.22}$	10.0	49.93$_{\pm 4.98}$	86.40$_{\pm 1.27}$
Conventional	5%	**22.53**$_{\pm 1.44}$	**95.81**$_{\pm 0.35}$	5%	**15.36**$_{\pm 1.79}$	**97.12**$_{\pm 0.24}$	0.01	39.77$_{\pm 6.38}$	86.20$_{\pm 5.50}$
	10%	25.95$_{\pm 2.26}$	94.82$_{\pm 0.65}$	10%	16.58$_{\pm 0.74}$	96.99$_{\pm 0.07}$	0.05	29.12$_{\pm 2.01}$	94.01$_{\pm 0.57}$
	15%	27.76$_{\pm 1.83}$	94.14$_{\pm 0.33}$	15%	20.35$_{\pm 2.04}$	96.27$_{\pm 0.28}$	0.1	**23.49**$_{\pm 1.18}$	**95.88**$_{\pm 0.28}$
	20%	28.52$_{\pm 3.30}$	94.10$_{\pm 0.78}$	20%	21.51$_{\pm 1.48}$	96.05$_{\pm 0.32}$	0.5	33.57$_{\pm 2.17}$	92.43$_{\pm 0.65}$
	30%	31.27$_{\pm 2.66}$	92.94$_{\pm 0.92}$	30%	27.43$_{\pm 1.44}$	94.37$_{\pm 0.14}$	1.0	37.66$_{\pm 1.91}$	90.99$_{\pm 0.93}$
	100%	36.05$_{\pm 2.38}$	91.46$_{\pm 1.01}$	100%	36.05$_{\pm 2.38}$	91.46$_{\pm 1.01}$	10.0	45.55$_{\pm 2.06}$	87.72$_{\pm 0.65}$

Table 19. Ablation study of outlier synthesis in Waterbirds datasets.

D.10. Ablation study of outlier synthesis in CelebA datasets

Table 20 presents the ablation study of outlier synthesis approaches in the CelebA datasets. The table compares three methods of outlier synthesis. In any setting, it can be observed that ASCOOD achieves near-perfect conventional OOD Detection performance. It is due to the trivial nature of ID classification which is the task of color identification. Similar to the case in Waterbirds datasets, a notable trend observed in random pixel shuffling is that increasing the percentage of shuffled pixels degrades OOD detection performance. A similar effect is evident when using the invariant pixel shuffle approach. We attribute this to the disruption of environmental features as more pixels are shuffled, making the outliers less challenging and thereby diminishing detection efficacy. Additionally, we observe that Gaussian noise addition for outlier synthesis yields the best performance on the spurious OOD Detection in CelebA benchmark among these compared approaches. This approach also forms challenging outliers because of the use of ID data in outlier synthesis.

OOD type	Random pixel shuffle			Invariant pixel shuffle			Gaussian noise addition		
	p_{inv}	FPR@95↓	AUROC↑	p_{inv}	FPR@95↓	AUROC↑	α	FPR@95↓	AUROC↑
Spurious	5%	**52.79**$_{\pm 1.29}$	**84.77**$_{\pm 0.63}$	5%	**47.27**$_{\pm 1.39}$	**85.37**$_{\pm 0.38}$	0.01	76.83$_{\pm 6.17}$	61.75$_{\pm 5.48}$
	10%	55.05$_{\pm 2.80}$	83.59$_{\pm 0.89}$	10%	50.34$_{\pm 5.00}$	84.70$_{\pm 0.70}$	0.05	49.30$_{\pm 3.07}$	83.38$_{\pm 0.87}$
	15%	55.41$_{\pm 3.38}$	83.29$_{\pm 1.02}$	15%	51.39$_{\pm 4.82}$	84.50$_{\pm 1.86}$	0.1	**42.62**$_{\pm 0.76}$	**86.50**$_{\pm 0.89}$
	20%	59.05$_{\pm 2.85}$	82.19$_{\pm 0.78}$	20%	59.05$_{\pm 2.85}$	82.19$_{\pm 0.78}$	0.5	52.90$_{\pm 1.54}$	83.05$_{\pm 2.00}$
	30%	58.60$_{\pm 3.99}$	82.41$_{\pm 1.80}$	30%	58.60$_{\pm 3.99}$	82.41$_{\pm 1.80}$	1.0	60.30$_{\pm 4.89}$	82.58$_{\pm 1.63}$
	100%	62.05$_{\pm 1.81}$	81.74$_{\pm 1.07}$	100%	62.05$_{\pm 1.81}$	81.74$_{\pm 1.07}$	10.0	60.40$_{\pm 8.12}$	80.85$_{\pm 2.61}$
Conventional	5%	**0.02**$_{\pm 0.02}$	**99.97**$_{\pm 0.01}$	5%	**0.02**$_{\pm 0.03}$	**99.98**$_{\pm 0.00}$	0.01	74.26$_{\pm 6.23}$	37.09$_{\pm 7.04}$
	10%	0.03$_{\pm 0.04}$	99.97$_{\pm 0.01}$	10%	0.03$_{\pm 0.04}$	99.97$_{\pm 0.00}$	0.05	0.02$_{\pm 0.02}$	99.98$_{\pm 0.00}$
	15%	0.04$_{\pm 0.02}$	99.97$_{\pm 0.01}$	15%	0.03$_{\pm 0.02}$	99.97$_{\pm 0.00}$	0.1	**0.02**$_{\pm 0.02}$	**99.98**$_{\pm 0.00}$
	20%	0.05$_{\pm 0.03}$	99.97$_{\pm 0.01}$	20%	0.05$_{\pm 0.03}$	99.97$_{\pm 0.01}$	0.5	0.32$_{\pm 0.05}$	99.90$_{\pm 0.01}$
	30%	0.06$_{\pm 0.01}$	99.97$_{\pm 0.00}$	30%	0.06$_{\pm 0.01}$	99.97$_{\pm 0.00}$	1.0	0.08$_{\pm 0.05}$	99.95$_{\pm 0.02}$
	100%	0.13$_{\pm 0.09}$	99.95$_{\pm 0.01}$	100%	0.13$_{\pm 0.09}$	99.95$_{\pm 0.01}$	10.0	0.32$_{\pm 0.05}$	99.90$_{\pm 0.01}$

Table 20. Ablation study of outlier synthesis in CelebA datasets.

E. Confidence score plots with conventional OOD datasets:

E.1. Waterbirds datasets

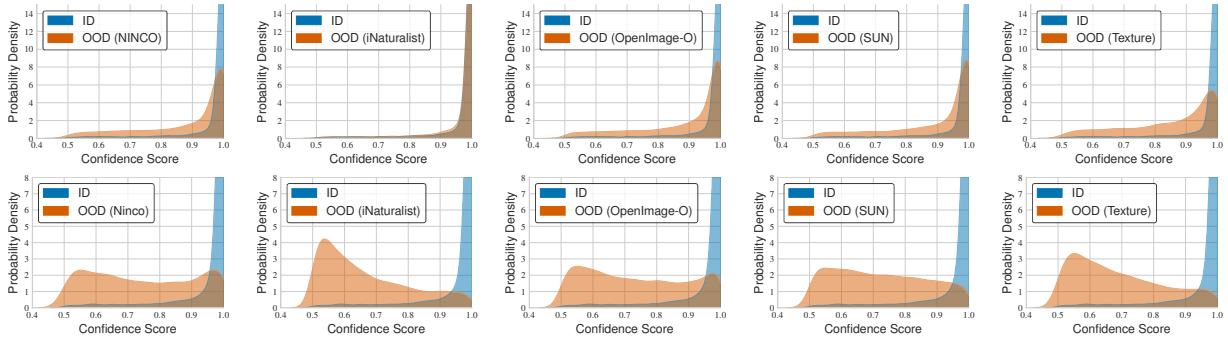


Figure 12. **Top row:** Visualization of confidence scores of cross-entropy baseline for Waterbirds ID and various conventional OOD datasets. **Bottom row:** Visualization of confidence scores of ASCOOD for Waterbirds ID and various conventional OOD datasets.

E.2. CelebA datasets

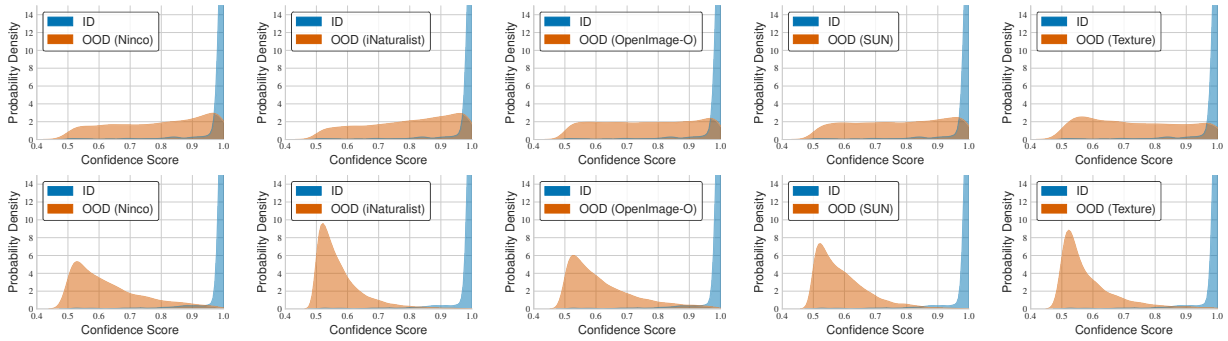


Figure 13. **Top row:** Visualization of confidence scores of cross-entropy baseline for CelebA ID and various conventional OOD datasets. **Bottom row:** Visualization of confidence scores of ASCOOD for CelebA ID and various conventional OOD datasets.

E.3. CIFAR-10 datasets

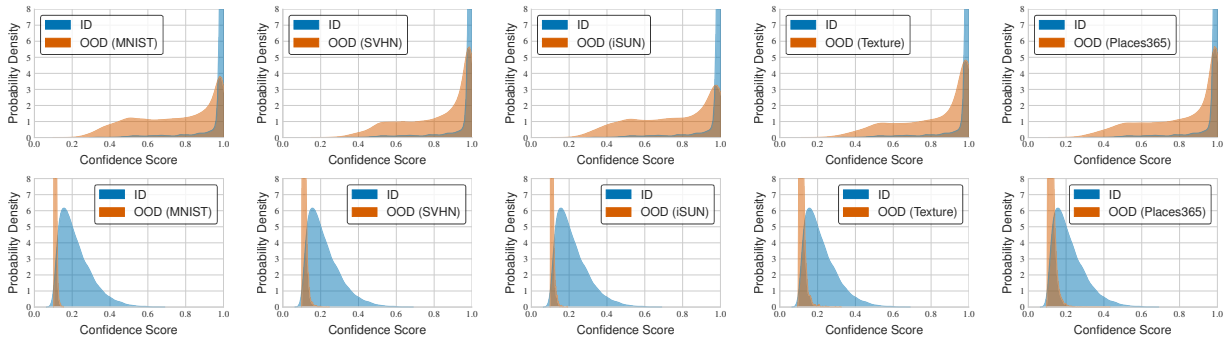


Figure 14. **Top row:** Visualization of confidence scores of cross-entropy baseline for CIFAR-10 ID and various conventional OOD datasets. **Bottom row:** Visualization of confidence scores of ASCOOD for CIFAR-10 ID and various conventional OOD datasets.

E.4. CIFAR-100 datasets

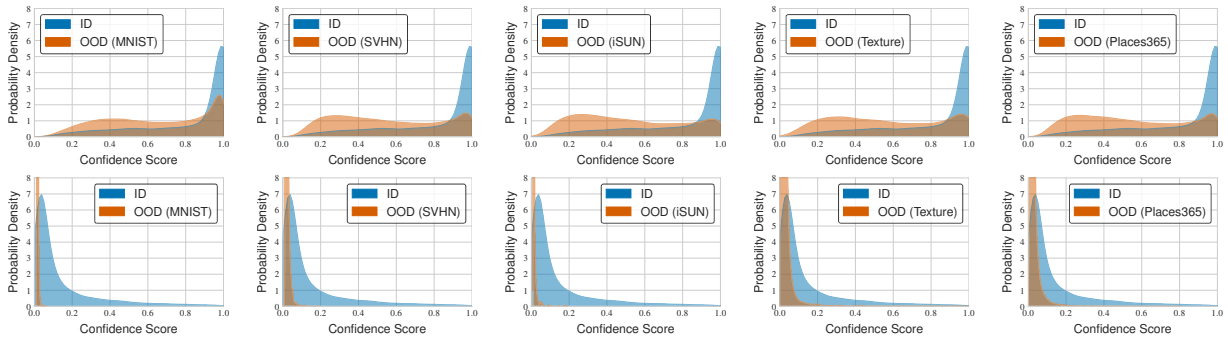


Figure 15. **Top row:** Visualization of confidence scores of cross-entropy baseline for CIFAR-100 ID and various conventional OOD datasets. **Bottom row:** Visualization of confidence scores of ASCOOD for CIFAR-100 ID and various conventional OOD datasets.

E.5. Aircraft datasets

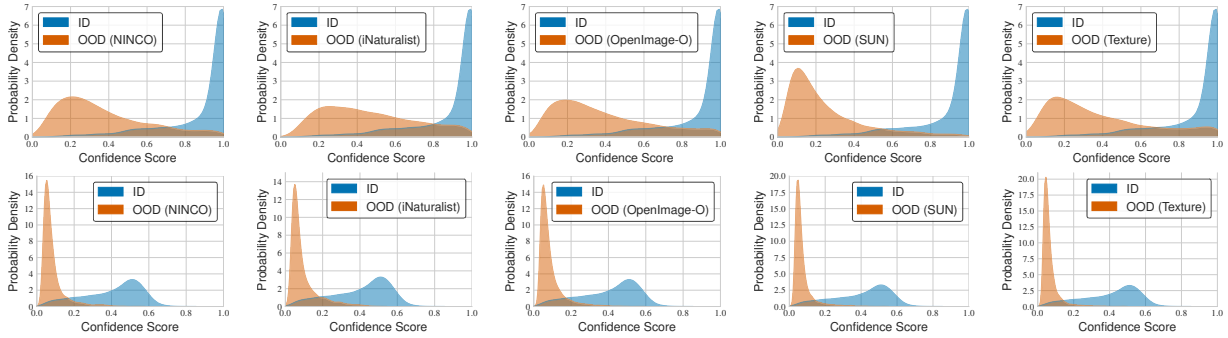


Figure 16. **Top row:** Visualization of confidence scores of cross-entropy baseline for Aircraft ID and various conventional OOD datasets. **Bottom row:** Visualization of confidence scores of ASCOOD for Aircraft ID and various conventional OOD datasets.

E.6. Car datasets

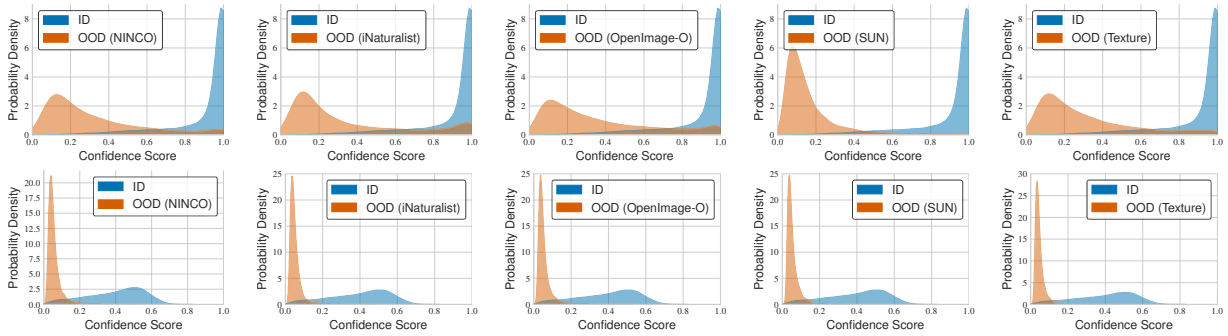
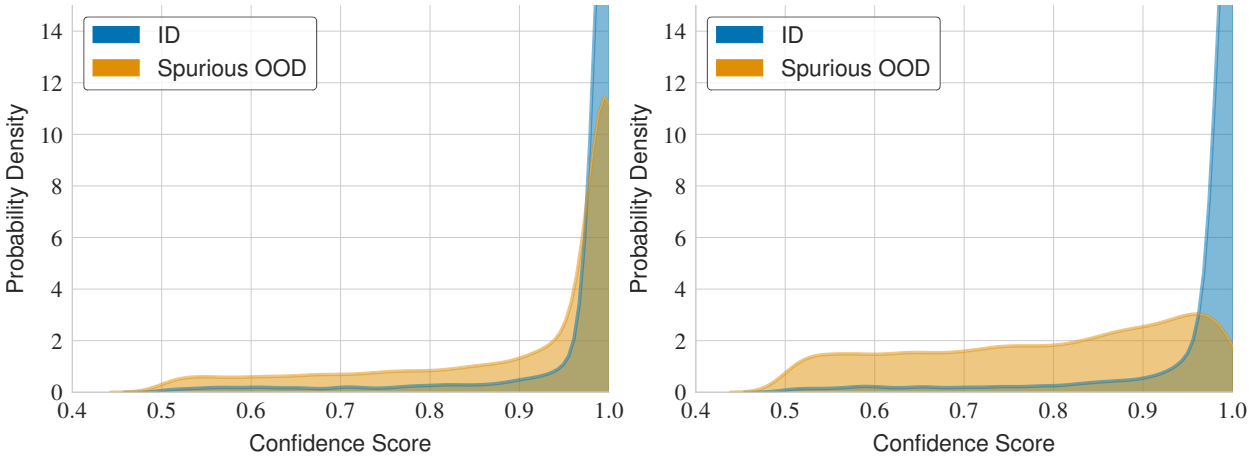


Figure 17. **Top row:** Visualization of confidence scores of cross-entropy baseline for Car ID and various conventional OOD datasets. **Bottom row:** Visualization of confidence scores of ASCOOD for Car ID and various conventional OOD datasets.

F. Confidence score plots with spurious and fine-grained OOD datasets:

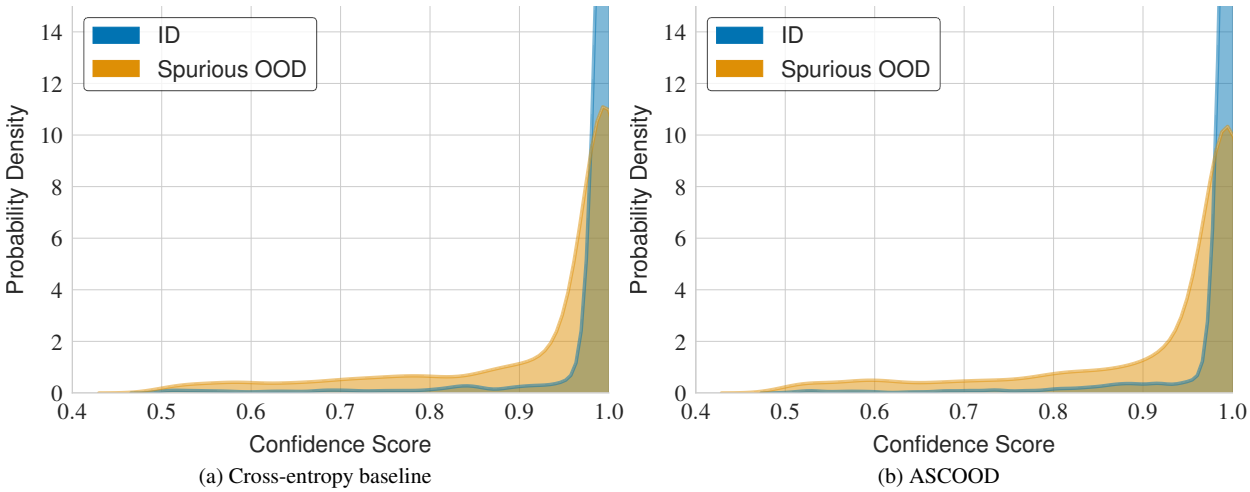
F.1. Waterbirds datasets



(a) Cross-entropy baseline (b) ASCOOD

Figure 18. Visualization of confidence scores for Waterbirds ID datasets and corresponding spurious OOD datasets.

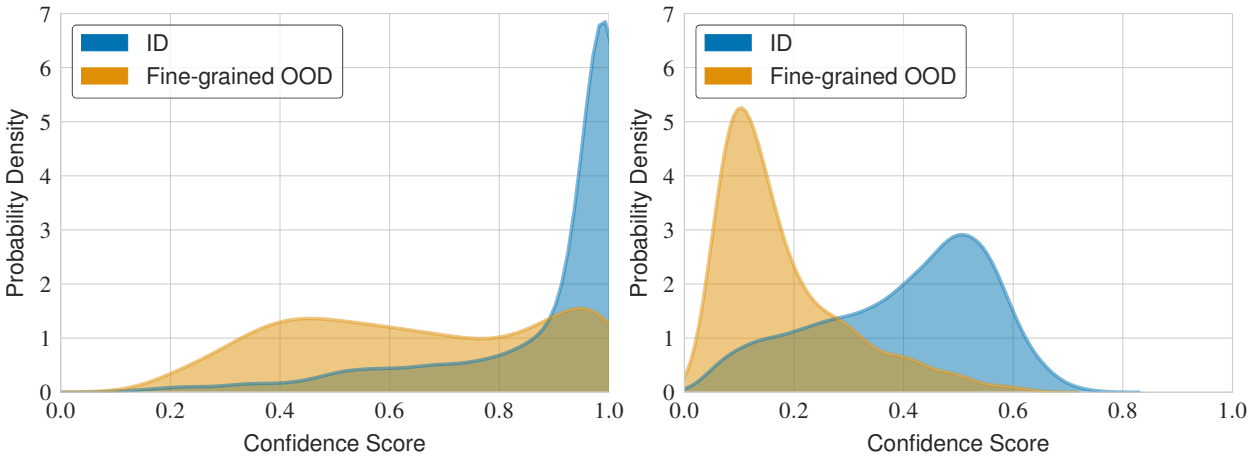
F.2. CelebA datasets



(a) Cross-entropy baseline (b) ASCOOD

Figure 19. Visualization of confidence scores for CelebA ID datasets and corresponding spurious OOD datasets.

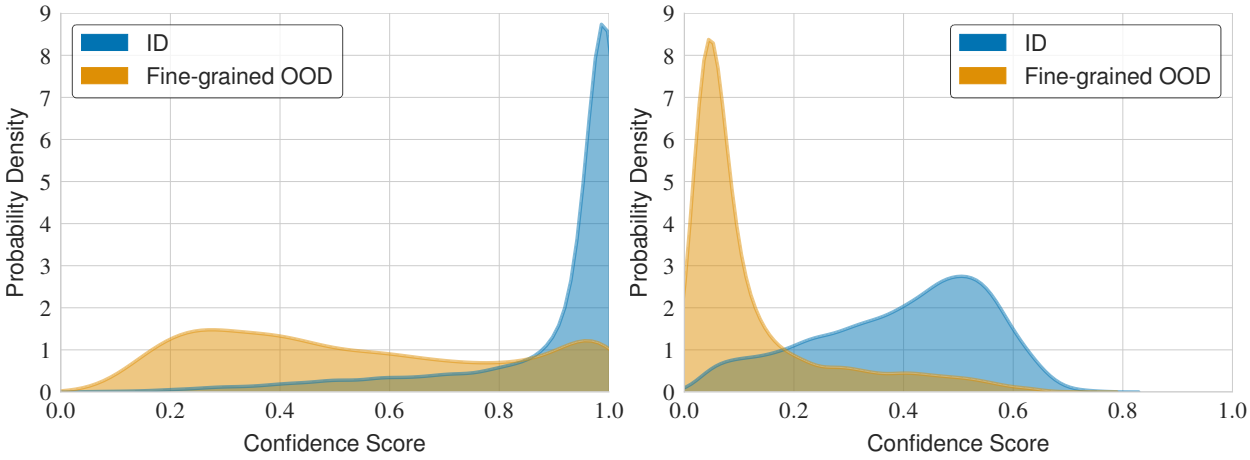
F.3. Aircraft datasets



(a) Cross-entropy baseline (b) ASCOOD

Figure 20. Visualization of confidence scores for Aircraft ID datasets and corresponding fine-grained OOD datasets.

F.4. Car datasets



(a) Cross-entropy baseline (b) ASCOOD

Figure 21. Visualization of confidence scores for Car ID datasets and corresponding fine-grained OOD datasets.

G. Complete results

G.1. ODIN vs. I-ODIN

We empirically find that perturbing only the most significant (one) color channel value, rather than all channel values (entire pixels of image), yields the best results across all benchmarks.

Method	**CIFAR-10 Benchmark**						**CIFAR-100 Benchmark**					
	CIFAR-100		TIN		Average		CIFAR-10		TIN		Average	
	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑
ODIN	77.00±5.74	82.18±1.87	75.38±6.42	83.55±1.84	76.19±6.08	82.87±1.85	60.64±0.56	78.18±0.14	55.19±0.57	81.63±0.08	57.91±0.51	79.90±0.11
I-ODIN	**61.33**±1.18	**86.87**±0.14	**50.20**±2.39	**88.96**±0.25	**55.76**±1.78	**87.92**±0.17	**59.09**±0.62	**79.24**±0.10	**51.57**±0.66	**82.96**±0.05	**55.33**±0.64	**81.10**±0.07

Table 21. ODIN vs. I-ODIN in near-OOD detection across CIFAR benchmarks of OpenOOD.

Method	**ImageNet-200 Benchmark**						**ImageNet-1k Benchmark**					
	SSB-Hard		NINCO		Average		SSB-Hard		NINCO		Average	
	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑
ODIN	73.51±0.38	77.19±0.06	60.00±0.80	83.34±0.12	66.76±0.26	80.27±0.08	76.83±0.00	71.74±0.00	68.16±0.00	77.77±0.00	72.50±0.00	74.75±0.11
I-ODIN	**69.47**±0.37	**80.22**±0.01	**49.37**±0.75	**85.76**±0.08	**59.42**±0.48	**82.99**±0.04	**76.15**±0.00	**72.49**±0.00	**58.86**±0.00	**80.57**±0.00	**67.51**±0.00	**76.93**±0.00

Table 22. ODIN vs. I-ODIN in near-OOD detection across ImageNet benchmarks of OpenOOD.

Method	**MINIST**		**SVHN**		**Textures**		**Places365**		**Average**	
	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑
CIFAR-10 Benchmark										
ODIN	**23.83**±12.34	**95.24**±1.96	68.61±0.52	84.58±0.77	67.70±11.06	86.94±2.26	70.36±6.96	85.07±1.24	57.62±4.24	87.96±0.61
I-ODIN	25.86±11.23	93.45±1.99	**31.80**±7.09	**91.55**±0.81	**43.36**±2.71	**89.78**±0.60	**49.20**±8.05	**89.16**±0.69	**37.55**±6.46	**90.99**±0.76
CIFAR-100 Benchmark										
ODIN	**45.94**±3.29	**83.79**±1.31	67.41±3.88	74.54±0.76	62.37±2.96	**79.33**±1.08	59.71±0.92	79.45±0.26	58.86±0.79	79.28±0.21
I-ODIN	52.92±3.90	78.89±1.50	**54.06**±2.95	**81.56**±1.44	**62.07**±1.97	78.48±0.83	**57.47**±0.93	**79.83**±0.24	**56.63**±1.30	**79.69**±0.56

Table 23. ODIN vs. I-ODIN in far-OOD detection across CIFAR benchmarks of OpenOOD.

Method	**iNaturalist**		**Textures**		**OpenImage-O**		**Average**	
	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑
ImageNet-200 Benchmark								
ODIN	**22.39**±1.87	**94.37**±0.41	42.99±1.56	**90.65**±0.20	37.30±0.59	**90.11**±0.15	34.23±1.05	**91.71**±0.19
I-ODIN	24.77±1.91	93.22±0.43	41.27±1.98	90.57±0.15	**35.24**±0.60	89.71±0.19	**33.76**±1.20	91.16±0.18
ImageNet-1k Benchmark								
ODIN	35.98±0.00	91.17±0.00	49.24±0.00	**89.00**±0.00	46.67±0.00	88.23±0.00	43.96±0.00	89.47±0.00
I-ODIN	**30.04**±0.00	**91.33**±0.00	**46.27**±0.00	88.37±0.00	**37.44**±0.00	**89.23**±0.00	**37.92**±0.00	**89.64**±0.00

Table 24. ODIN vs. I-ODIN in far-OOD detection across ImageNet benchmarks of OpenOOD.

G.2. CIFAR-10 results

Method	MNIST	SVHN	iSUN	Texture	Places365	Average
			Posthoc methods			
MSP	23.63$_{\pm 5.81}$ / 92.63$_{\pm 1.57}$	25.80$_{\pm 1.64}$ / 91.46$_{\pm 0.40}$	22.86$_{\pm 1.09}$ / 92.39$_{\pm 0.51}$	34.96$_{\pm 4.66}$ / 89.89$_{\pm 0.71}$	42.46$_{\pm 3.81}$ / 88.92$_{\pm 0.47}$	29.94$_{\pm 1.32}$ / 91.06$_{\pm 0.35}$
TempScale	23.53$_{\pm 7.05}$ / 93.11$_{\pm 1.77}$	26.97$_{\pm 2.65}$ / 91.66$_{\pm 0.52}$	22.95$_{\pm 1.07}$ / 92.75$_{\pm 0.50}$	38.16$_{\pm 5.87}$ / 90.01$_{\pm 0.74}$	45.28$_{\pm 4.52}$ / 89.11$_{\pm 0.52}$	31.38$_{\pm 1.78}$ / 91.33$_{\pm 0.42}$
ODIN	23.81$_{\pm 12.35}$ / 95.24$_{\pm 1.96}$	68.61$_{\pm 0.45}$ / 84.58$_{\pm 0.77}$	27.26$_{\pm 3.81}$ / 94.63$_{\pm 0.24}$	67.74$_{\pm 11.09}$ / 86.94$_{\pm 2.26}$	70.35$_{\pm 6.99}$ / 85.07$_{\pm 1.24}$	51.55$_{\pm 4.03}$ / 89.29$_{\pm 0.53}$
MDS	27.30$_{\pm 3.54}$ / 90.10$_{\pm 2.41}$	25.95$_{\pm 2.52}$ / 91.19$_{\pm 0.47}$	32.56$_{\pm 6.52}$ / 89.86$_{\pm 2.72}$	27.92$_{\pm 4.18}$ / 92.69$_{\pm 1.06}$	47.66$_{\pm 4.54}$ / 84.90$_{\pm 2.54}$	32.28$_{\pm 3.97}$ / 89.75$_{\pm 1.60}$
MDSEns	**1.30**$_{\pm \mathbf{0.51}}$ / 99.17$_{\pm 0.41}$	74.34$_{\pm 1.04}$ / 66.56$_{\pm 0.58}$	32.09$_{\pm 12.53}$ / 93.98$_{\pm 3.08}$	76.07$_{\pm 0.17}$ / 77.40$_{\pm 0.28}$	94.16$_{\pm 0.33}$ / 52.47$_{\pm 0.15}$	55.59$_{\pm 2.36}$ / 77.92$_{\pm 0.57}$
RMDS	21.48$_{\pm 2.32}$ / 93.22$_{\pm 0.80}$	23.45$_{\pm 1.48}$ / 91.84$_{\pm 0.26}$	20.30$_{\pm 1.17}$ / 93.95$_{\pm 0.51}$	25.26$_{\pm 0.54}$ / 92.23$_{\pm 0.23}$	31.21$_{\pm 0.28}$ / 91.51$_{\pm 0.11}$	24.34$_{\pm 0.74}$ / 92.55$_{\pm 0.22}$
Gram	70.29$_{\pm 8.95}$ / 72.64$_{\pm 2.34}$	33.92$_{\pm 17.36}$ / 91.51$_{\pm 4.45}$	98.90$_{\pm 0.75}$ / 34.46$_{\pm 2.42}$	94.64$_{\pm 2.71}$ / 62.35$_{\pm 8.27}$	90.49$_{\pm 1.93}$ / 60.44$_{\pm 3.41}$	77.65$_{\pm 5.39}$ / 64.28$_{\pm 2.69}$
EBO	24.99$_{\pm 12.94}$ / 94.32$_{\pm 2.53}$	35.15$_{\pm 6.17}$ / 91.78$_{\pm 0.98}$	27.26$_{\pm 0.23}$ / 93.32$_{\pm 0.44}$	51.84$_{\pm 6.09}$ / 89.47$_{\pm 0.70}$	54.86$_{\pm 6.51}$ / 89.25$_{\pm 0.78}$	38.82$_{\pm 4.25}$ / 91.63$_{\pm 0.76}$
GradNorm	85.40$_{\pm 4.85}$ / 63.72$_{\pm 7.37}$	91.66$_{\pm 2.42}$ / 53.91$_{\pm 6.36}$	87.74$_{\pm 0.27}$ / 66.84$_{\pm 1.08}$	98.08$_{\pm 0.49}$ / 52.07$_{\pm 4.09}$	92.45$_{\pm 2.28}$ / 60.50$_{\pm 5.33}$	91.07$_{\pm 1.76}$ / 59.41$_{\pm 2.67}$
ReAct	33.76$_{\pm 18.01}$ / 92.81$_{\pm 3.03}$	50.23$_{\pm 15.97}$ / 89.12$_{\pm 3.19}$	25.86$_{\pm 2.55}$ / 93.62$_{\pm 0.72}$	51.40$_{\pm 11.43}$ / 89.38$_{\pm 1.49}$	44.21$_{\pm 3.32}$ / 90.35$_{\pm 0.78}$	41.09$_{\pm 6.33}$ / 91.06$_{\pm 1.04}$
MLS	25.06$_{\pm 12.87}$ / 94.15$_{\pm 2.48}$	35.09$_{\pm 6.10}$ / 91.69$_{\pm 0.94}$	27.25$_{\pm 0.28}$ / 93.21$_{\pm 0.43}$	51.71$_{\pm 6.11}$ / 89.41$_{\pm 0.71}$	54.84$_{\pm 6.50}$ / 89.14$_{\pm 0.76}$	38.79$_{\pm 4.19}$ / 91.52$_{\pm 0.73}$
KLM	17.30$_{\pm 0.74}$ / 95.10$_{\pm 0.63}$	7.49$_{\pm 6.40}$ / 98.86$_{\pm 0.83}$	21.11$_{\pm 4.69}$ / 91.24$_{\pm 2.70}$	19.42$_{\pm 3.42}$ / 94.87$_{\pm 1.44}$	25.49$_{\pm 0.40}$ / 93.40$_{\pm 0.75}$	18.16$_{\pm 2.10}$ / 94.69$_{\pm 0.87}$
VIM	18.34$_{\pm 1.42}$ / 94.76$_{\pm 0.38}$	19.29$_{\pm 0.41}$ / 94.51$_{\pm 0.48}$	20.73$_{\pm 1.46}$ / 94.28$_{\pm 0.43}$	21.17$_{\pm 1.84}$ / 95.15$_{\pm 0.34}$	41.43$_{\pm 2.16}$ / 89.49$_{\pm 0.39}$	24.19$_{\pm 0.31}$ / 93.64$_{\pm 0.16}$
DICE	30.63$_{\pm 10.66}$ / 90.42$_{\pm 6.02}$	36.51$_{\pm 4.75}$ / 90.09$_{\pm 1.83}$	59.71$_{\pm 10.23}$ / 81.33$_{\pm 3.22}$	62.47$_{\pm 4.79}$ / 81.88$_{\pm 2.37}$	77.23$_{\pm 12.64}$ / 74.77$_{\pm 4.87}$	53.31$_{\pm 5.29}$ / 83.70$_{\pm 2.16}$
RankFeat	61.85$_{\pm 12.78}$ / 75.88$_{\pm 5.22}$	64.50$_{\pm 7.37}$ / 68.15$_{\pm 7.45}$	42.43$_{\pm 15.56}$ / 83.93$_{\pm 7.41}$	59.71$_{\pm 9.78}$ / 73.46$_{\pm 6.49}$	43.70$_{\pm 7.39}$ / 85.99$_{\pm 3.04}$	54.44$_{\pm 9.49}$ / 77.48$_{\pm 5.48}$
ASH	58.79$_{\pm 13.90}$ / 88.83$_{\pm 3.34}$	78.93$_{\pm 5.95}$ / 81.25$_{\pm 3.34}$	55.21$_{\pm 5.26}$ / 89.87$_{\pm 2.26}$	83.66$_{\pm 2.02}$ / 82.27$_{\pm 0.55}$	76.13$_{\pm 6.09}$ / 84.14$_{\pm 2.86}$	70.55$_{\pm 5.50}$ / 85.27$_{\pm 2.05}$
SHE	42.23$_{\pm 20.60}$ / 90.43$_{\pm 4.76}$	62.75$_{\pm 4.02}$ / 86.37$_{\pm 1.33}$	52.62$_{\pm 8.31}$ / 89.65$_{\pm 0.71}$	84.59$_{\pm 5.30}$ / 81.57$_{\pm 1.21}$	76.35$_{\pm 5.31}$ / 82.89$_{\pm 1.22}$	63.71$_{\pm 5.22}$ / 86.18$_{\pm 1.20}$
GEN	23.00$_{\pm 7.75}$ / 93.83$_{\pm 2.14}$	28.14$_{\pm 2.59}$ / 91.97$_{\pm 0.66}$	23.55$_{\pm 1.70}$ / 93.22$_{\pm 0.34}$	40.72$_{\pm 6.60}$ / 90.14$_{\pm 0.76}$	47.04$_{\pm 3.23}$ / 89.46$_{\pm 0.65}$	32.49$_{\pm 1.17}$ / 91.72$_{\pm 0.57}$
NNGuide	29.03$_{\pm 13.64}$ / 93.39$_{\pm 2.38}$	42.07$_{\pm 4.57}$ / 90.55$_{\pm 0.71}$	25.87$_{\pm 1.13}$ / 93.42$_{\pm 0.45}$	50.57$_{\pm 10.86}$ / 89.70$_{\pm 1.12}$	50.65$_{\pm 6.79}$ / 89.88$_{\pm 0.76}$	39.64$_{\pm 3.91}$ / 91.39$_{\pm 0.57}$
Relation	22.41$_{\pm 1.48}$ / 93.79$_{\pm 0.59}$	24.68$_{\pm 0.78}$ / 92.43$_{\pm 0.30}$	22.35$_{\pm 2.82}$ / 93.70$_{\pm 0.91}$	26.83$_{\pm 0.28}$ / 92.26$_{\pm 0.18}$	36.67$_{\pm 1.16}$ / 90.29$_{\pm 0.34}$	26.59$_{\pm 0.64}$ / 92.49$_{\pm 0.22}$
SCALE	48.69$_{\pm 15.40}$ / 90.58$_{\pm 2.96}$	70.53$_{\pm 8.72}$ / 84.63$_{\pm 3.05}$	45.64$_{\pm 4.20}$ / 91.34$_{\pm 1.50}$	80.40$_{\pm 3.92}$ / 83.94$_{\pm 0.35}$	70.51$_{\pm 5.98}$ / 86.41$_{\pm 1.92}$	63.15$_{\pm 5.95}$ / 87.38$_{\pm 1.51}$
FDBD	19.33$_{\pm 2.27}$ / 94.71$_{\pm 0.90}$	22.89$_{\pm 1.60}$ / 92.80$_{\pm 0.82}$	19.37$_{\pm 2.82}$ / 94.54$_{\pm 1.06}$	24.27$_{\pm 0.75}$ / 93.13$_{\pm 0.27}$	29.12$_{\pm 1.64}$ / 92.01$_{\pm 0.47}$	23.00$_{\pm 0.80}$ / 93.44$_{\pm 0.26}$
			Training methods			
ConfBranch	14.10$_{\pm 1.87}$ / 96.02$_{\pm 0.69}$	14.31$_{\pm 0.98}$ / 95.70$_{\pm 0.63}$	21.39$_{\pm 2.68}$ / 93.36$_{\pm 0.91}$	26.59$_{\pm 1.83}$ / 91.51$_{\pm 0.58}$	30.61$_{\pm 0.84}$ / 89.89$_{\pm 0.45}$	21.40$_{\pm 1.07}$ / 93.30$_{\pm 0.50}$
RotPred	7.26$_{\pm 1.28}$ / 97.89$_{\pm 0.18}$	**4.03**$_{\pm \mathbf{0.66}}$ / **98.68**$_{\pm \mathbf{0.13}}$	13.40$_{\pm 1.72}$ / 95.90$_{\pm 0.33}$	**10.49**$_{\pm \mathbf{0.51}}$ / 97.08$_{\pm 0.12}$	28.84$_{\pm 2.25}$ / 92.24$_{\pm 0.35}$	12.80$_{\pm 0.69}$ / 96.36$_{\pm 0.12}$
G-ODIN	5.45$_{\pm 0.93}$ / 98.77$_{\pm 0.26}$	11.17$_{\pm 1.23}$ / 97.58$_{\pm 0.25}$	12.02$_{\pm 3.51}$ / 97.51$_{\pm 0.85}$	27.10$_{\pm 2.75}$ / 94.97$_{\pm 0.49}$	44.37$_{\pm 5.25}$ / 90.13$_{\pm 1.11}$	20.02$_{\pm 0.91}$ / 95.79$_{\pm 0.19}$
MOS	71.20$_{\pm 14.94}$ / 70.86$_{\pm 9.56}$	59.55$_{\pm 4.48}$ / 73.85$_{\pm 5.59}$	55.50$_{\pm 6.34}$ / 80.61$_{\pm 3.42}$	68.14$_{\pm 1.96}$ / 72.33$_{\pm 1.53}$	50.20$_{\pm 1.07}$ / 87.11$_{\pm 1.24}$	60.32$_{\pm 2.90}$ / 77.39$_{\pm 2.75}$
VOS	24.23$_{\pm 5.16}$ / 94.05$_{\pm 1.19}$	24.94$_{\pm 6.89}$ / 94.04$_{\pm 1.37}$	30.26$_{\pm 8.44}$ / 93.27$_{\pm 1.36}$	55.67$_{\pm 7.29}$ / 89.37$_{\pm 0.82}$	55.32$_{\pm 5.44}$ / 89.49$_{\pm 0.63}$	38.08$_{\pm 3.29}$ / 92.04$_{\pm 0.52}$
LogitNorm	4.73$_{\pm 0.80}$ / 99.00$_{\pm 0.13}$	10.53$_{\pm 1.83}$ / 97.41$_{\pm 0.78}$	10.74$_{\pm 3.90}$ / 97.21$_{\pm 1.28}$	20.01$_{\pm 0.84}$ / 95.17$_{\pm 0.09}$	23.87$_{\pm 1.09}$ / 93.90$_{\pm 0.27}$	13.98$_{\pm 1.33}$ / 96.54$_{\pm 0.45}$
CIDER	11.86$_{\pm 4.03}$ / 97.11$_{\pm 1.31}$	18.35$_{\pm 4.46}$ / 94.62$_{\pm 1.95}$	17.82$_{\pm 3.74}$ / 95.09$_{\pm 1.13}$	25.64$_{\pm 1.24}$ / 93.10$_{\pm 0.94}$	27.48$_{\pm 1.75}$ / 92.56$_{\pm 0.72}$	20.23$_{\pm 2.90}$ / 94.49$_{\pm 1.18}$
NPOS	13.72$_{\pm 1.42}$ / 96.12$_{\pm 0.25}$	20.79$_{\pm 3.86}$ / 94.57$_{\pm 1.26}$	18.59$_{\pm 1.23}$ / 94.45$_{\pm 0.64}$	27.91$_{\pm 1.25}$ / 92.24$_{\pm 0.26}$	32.26$_{\pm 1.97}$ / 90.19$_{\pm 0.67}$	22.65$_{\pm 1.10}$ / 93.51$_{\pm 0.18}$
ASCOOD (MSP [65])	3.34$_{\pm 1.36}$ / 99.04$_{\pm 0.56}$	8.31$_{\pm 1.42}$ / 97.91$_{\pm 0.69}$	1.96$_{\pm 0.25}$ / 99.59$_{\pm 0.07}$	13.44$_{\pm 1.38}$ / 97.16$_{\pm 0.43}$	19.13$_{\pm 0.76}$ / 95.30$_{\pm 0.14}$	9.24$_{\pm 0.71}$ / 97.80$_{\pm 0.31}$
ASCOOD (SCALE)	1.37$_{\pm 0.35}$ / **99.68**$_{\pm \mathbf{0.11}}$	6.73$_{\pm 1.37}$ / 98.61$_{\pm 0.34}$	**0.94**$_{\pm \mathbf{0.19}}$ / **99.79**$_{\pm \mathbf{0.05}}$	10.93$_{\pm 0.52}$ / **97.95**$_{\pm \mathbf{0.06}}$	**18.47**$_{\pm \mathbf{1.12}}$ / **95.56**$_{\pm \mathbf{0.34}}$	**7.69**$_{\pm \mathbf{0.29}}$ / **98.32**$_{\pm \mathbf{0.07}}$

Table 25. OOD detection (FPR@95↓/ AUROC↑) results in CIFAR-10 benchmark. Refer to Table 8 for OE and MixOE results.

G.3. Near-OOD results

Method	CIFAR-100		Tiny ImageNet (TIN)		Average	
	FPR@95↓	AUROC↑	FPR@95↓	AUROC↑	FPR@95↓	AUROC↑
Posthoc methods						
MSP	53.08±4.86	87.19±0.33	43.27±3.00	88.87±0.19	48.17±3.92	88.03±0.25
TempScale	55.81±5.07	87.17±0.40	46.11±3.63	89.00±0.23	50.96±4.32	88.09±0.31
ODIN	77.00±5.74	82.18±1.87	75.38±6.42	83.55±1.84	76.19±6.08	82.87±1.85
MDS	52.81±3.62	83.59±2.27	46.99±4.36	84.81±2.53	49.90±3.98	84.20±2.40
MDSEns	91.87±0.10	61.29±0.23	92.66±0.42	59.57±0.53	92.26±0.20	60.43±0.26
RMDS	43.86±3.49	88.83±0.35	33.91±1.39	90.76±0.27	38.89±2.39	89.80±0.28
Gram	91.68±2.24	58.33±4.49	90.06±1.59	58.98±5.19	90.87±1.91	58.66±4.83
EBO	66.60±4.46	86.36±0.58	56.08±4.83	88.80±0.36	61.34±4.63	87.58±0.46
GradNorm	94.54±1.11	54.43±1.59	94.89±0.60	55.37±0.41	94.72±0.82	54.90±0.98
ReAct	67.40±7.34	85.93±0.83	59.71±7.31	88.29±0.44	63.56±7.33	87.11±0.61
MLS	66.59±4.44	86.31±0.59	56.06±4.82	88.72±0.36	61.32±4.62	87.52±0.47
KLM	90.55±5.83	77.89±0.75	85.18±7.60	80.49±0.85	87.86±6.37	79.19±0.80
VIM	49.19±3.15	87.75±0.28	40.49±1.55	89.62±0.33	44.84±2.31	88.68±0.28
DICE	73.71±7.67	77.01±0.88	66.37±7.68	79.67±0.87	70.04±7.64	78.34±0.79
RankFeat	65.32±3.48	77.98±2.24	56.44±5.76	80.94±2.80	60.88±4.60	79.46±2.52
ASH	87.31±2.06	74.11±1.55	86.25±1.58	76.44±0.61	86.78±1.82	75.27±1.04
SHE	81.00±3.42	80.31±0.69	78.30±3.52	82.76±0.43	79.65±3.47	81.54±0.51
GEN	58.75±3.97	87.21±0.36	48.59±2.34	89.20±0.25	53.67±3.14	88.20±0.30
NNGuide	67.89±6.67	86.33±0.74	57.26±7.15	88.80±0.52	62.57±6.90	87.57±0.62
Relation	41.31±0.58	88.84±0.13	34.26±0.50	90.69±0.20	37.78±0.54	89.77±0.16
SCALE	81.78±3.98	81.27±0.74	79.14±4.04	83.84±0.02	80.46±4.01	82.55±0.36
FDBD	39.58±1.79	89.56±0.15	31.04±1.21	91.60±0.18	37.12±0.99	90.42±0.36
Train-time regularization methods						
ConfBranch	34.44±0.81	88.91±0.25	28.11±0.61	90.77±0.25	31.28±0.66	89.84±0.24
RotPred	34.24±2.64	91.19±0.32	22.04±0.73	94.17±0.24	28.14±1.68	92.68±0.27
G-ODIN	48.86±2.91	88.14±0.60	42.21±2.18	90.09±0.54	45.54±2.52	89.12±0.57
MOS	79.38±5.06	70.57±3.04	78.05±6.69	72.34±3.16	78.72±5.86	71.45±3.09
VOS	61.57±3.24	86.57±0.57	52.49±0.73	88.84±0.48	57.03±1.92	87.70±0.48
LogitNorm	34.37±1.30	90.95±0.22	24.30±0.54	93.70±0.06	29.34±0.81	92.33±0.08
CIDER	35.60±0.78	89.47±0.19	28.61±1.10	91.94±0.19	32.11±0.94	90.71±0.16
NPOS	39.10±0.68	88.21±0.21	33.73±0.34	89.89±0.24	36.14±0.51	89.05±0.22
ASCOOD (MSP [65])	**31.50**±1.55	**91.48**±0.43	**20.82**±1.18	**94.58**±0.34	**26.16**±1.28	**93.03**±0.38

Table 26. Near-OOD detection results on CIFAR-100 and Tiny ImageNet datasets in CIFAR-10 benchmark. OOD detection results for OE and MixOE are omitted to avoid unfair comparison, as Tiny ImageNet is used to incentivize predictive uncertainty during training.

G.4. CIFAR-100 results

Method	MNIST	SVHN	iSUN	Texture	Places365	Average
			Posthoc methods			
MSP	57.23$_{\pm 4.67}$ / 76.08$_{\pm 1.86}$	59.07$_{\pm 2.53}$ / 78.42$_{\pm 0.89}$	52.64$_{\pm 0.50}$ / 81.37$_{\pm 0.37}$	61.89$_{\pm 1.28}$ / 77.32$_{\pm 0.70}$	56.63$_{\pm 0.86}$ / 79.23$_{\pm 0.29}$	57.49$_{\pm 0.85}$ / 78.48$_{\pm 0.42}$
TempScale	56.04$_{\pm 4.62}$ / 77.27$_{\pm 1.85}$	57.69$_{\pm 2.67}$ / 79.79$_{\pm 1.05}$	51.19$_{\pm 0.73}$ / 82.76$_{\pm 0.50}$	61.55$_{\pm 1.43}$ / 78.11$_{\pm 0.72}$	56.44$_{\pm 0.93}$ / 79.80$_{\pm 0.25}$	56.58$_{\pm 0.99}$ / 79.55$_{\pm 0.51}$
ODIN	45.94$_{\pm 3.23}$ / 83.79$_{\pm 1.30}$	67.43$_{\pm 3.86}$ / 74.54$_{\pm 0.76}$	36.97$_{\pm 1.24}$ / 90.10$_{\pm 0.41}$	62.37$_{\pm 2.95}$ / 79.34$_{\pm 1.08}$	59.73$_{\pm 0.93}$ / 79.45$_{\pm 0.26}$	54.49$_{\pm 0.72}$ / 81.44$_{\pm 0.20}$
MDS	71.71$_{\pm 2.94}$ / 67.47$_{\pm 0.81}$	67.22$_{\pm 6.07}$ / 70.67$_{\pm 6.40}$	77.91$_{\pm 4.91}$ / 64.07$_{\pm 6.27}$	70.49$_{\pm 2.48}$ / 76.26$_{\pm 0.69}$	79.60$_{\pm 0.35}$ / 63.15$_{\pm 0.49}$	73.39$_{\pm 1.69}$ / 68.32$_{\pm 1.24}$
MDSEns	2.83$_{\pm 0.86}$ / 98.21$_{\pm 0.78}$	82.57$_{\pm 2.58}$ / 53.76$_{\pm 1.63}$	48.26$_{\pm 3.74}$ / 91.25$_{\pm 1.35}$	84.94$_{\pm 0.83}$ / 69.75$_{\pm 1.14}$	96.61$_{\pm 0.17}$ / 42.27$_{\pm 0.73}$	63.04$_{\pm 0.16}$ / 71.05$_{\pm 0.33}$
RMDS	52.04$_{\pm 6.28}$ / 79.74$_{\pm 2.49}$	51.65$_{\pm 3.68}$ / 84.89$_{\pm 1.10}$	53.49$_{\pm 2.41}$ / 80.80$_{\pm 1.81}$	53.99$_{\pm 1.05}$ / 83.65$_{\pm 0.51}$	53.57$_{\pm 0.42}$ / 83.40$_{\pm 0.46}$	52.95$_{\pm 0.13}$ / 82.50$_{\pm 0.10}$
Gram	53.53$_{\pm 7.45}$ / 80.70$_{\pm 4.15}$	20.07$_{\pm 1.96}$ / 95.55$_{\pm 0.60}$	99.96$_{\pm 0.02}$ / 13.98$_{\pm 1.77}$	89.51$_{\pm 2.53}$ / 70.79$_{\pm 1.32}$	94.67$_{\pm 0.61}$ / 46.38$_{\pm 1.21}$	71.55$_{\pm 1.90}$ / 61.48$_{\pm 1.03}$
EBO	52.62$_{\pm 3.84}$ / 79.18$_{\pm 1.36}$	53.62$_{\pm 3.14}$ / 82.03$_{\pm 1.74}$	47.47$_{\pm 1.57}$ / 85.10$_{\pm 0.74}$	62.34$_{\pm 2.06}$ / 78.35$_{\pm 0.83}$	57.75$_{\pm 0.87}$ / 79.52$_{\pm 0.23}$	54.76$_{\pm 1.42}$ / 80.84$_{\pm 0.60}$
GradNorm	86.98$_{\pm 1.45}$ / 65.35$_{\pm 1.12}$	69.91$_{\pm 7.94}$ / 76.95$_{\pm 4.73}$	86.33$_{\pm 4.31}$ / 71.43$_{\pm 2.15}$	92.52$_{\pm 0.60}$ / 64.58$_{\pm 0.13}$	85.33$_{\pm 0.45}$ / 69.69$_{\pm 0.17}$	84.21$_{\pm 2.34}$ / 69.60$_{\pm 1.23}$
ReAct	56.03$_{\pm 5.66}$ / 78.37$_{\pm 1.59}$	50.43$_{\pm 2.06}$ / 83.01$_{\pm 0.97}$	44.43$_{\pm 1.32}$ / 85.76$_{\pm 0.79}$	55.03$_{\pm 0.81}$ / 80.15$_{\pm 0.46}$	55.33$_{\pm 0.40}$ / 80.03$_{\pm 0.11}$	52.25$_{\pm 1.48}$ / 81.46$_{\pm 0.55}$
MLS	52.94$_{\pm 3.83}$ / 78.91$_{\pm 1.47}$	53.90$_{\pm 3.05}$ / 81.65$_{\pm 1.49}$	47.72$_{\pm 1.44}$ / 84.54$_{\pm 0.67}$	62.39$_{\pm 2.14}$ / 78.39$_{\pm 0.84}$	57.67$_{\pm 0.91}$ / 79.75$_{\pm 0.24}$	54.92$_{\pm 1.35}$ / 80.65$_{\pm 0.57}$
KLM	72.88$_{\pm 6.55}$ / 74.15$_{\pm 2.60}$	50.31$_{\pm 7.02}$ / 79.32$_{\pm 0.44}$	84.16$_{\pm 2.35}$ / 77.41$_{\pm 0.96}$	81.80$_{\pm 5.80}$ / 75.76$_{\pm 0.47}$	81.62$_{\pm 1.36}$ / 75.68$_{\pm 0.26}$	74.15$_{\pm 1.15}$ / 76.46$_{\pm 0.50}$
VIM	48.30$_{\pm 1.07}$ / 81.89$_{\pm 1.02}$	46.19$_{\pm 5.45}$ / 83.14$_{\pm 3.72}$	46.70$_{\pm 3.56}$ / 82.28$_{\pm 3.75}$	46.85$_{\pm 2.28}$ / 85.91$_{\pm 0.78}$	61.57$_{\pm 0.76}$ / 75.85$_{\pm 0.37}$	49.92$_{\pm 1.06}$ / 81.81$_{\pm 0.76}$
DICE	51.80$_{\pm 3.67}$ / 79.86$_{\pm 1.89}$	49.58$_{\pm 3.32}$ / 84.22$_{\pm 2.00}$	47.74$_{\pm 1.35}$ / 84.62$_{\pm 0.93}$	64.21$_{\pm 1.63}$ / 77.63$_{\pm 0.34}$	59.39$_{\pm 1.26}$ / 78.33$_{\pm 0.66}$	54.55$_{\pm 0.45}$ / 80.93$_{\pm 0.30}$
RankFeat	75.02$_{\pm 5.82}$ / 63.03$_{\pm 3.85}$	58.49$_{\pm 2.29}$ / 72.14$_{\pm 1.39}$	61.72$_{\pm 0.94}$ / 74.22$_{\pm 0.92}$	66.87$_{\pm 3.80}$ / 69.40$_{\pm 3.08}$	77.42$_{\pm 1.96}$ / 63.82$_{\pm 1.83}$	67.90$_{\pm 0.96}$ / 68.52$_{\pm 1.32}$
ASH	51.57$_{\pm 3.60}$ / 80.71$_{\pm 1.13}$	47.61$_{\pm 2.24}$ / 85.19$_{\pm 0.88}$	50.28$_{\pm 3.17}$ / 83.86$_{\pm 1.43}$	57.77$_{\pm 1.78}$ / 81.10$_{\pm 0.63}$	56.97$_{\pm 0.45}$ / 80.45$_{\pm 0.24}$	52.84$_{\pm 1.20}$ / 82.26$_{\pm 0.54}$
SHE	58.78$_{\pm 2.71}$ / 76.76$_{\pm 1.06}$	59.16$_{\pm 7.62}$ / 80.97$_{\pm 3.98}$	53.93$_{\pm 2.95}$ / 82.28$_{\pm 0.96}$	73.28$_{\pm 3.20}$ / 73.64$_{\pm 1.28}$	65.26$_{\pm 0.99}$ / 76.30$_{\pm 0.51}$	62.08$_{\pm 2.73}$ / 77.99$_{\pm 1.09}$
GEN	53.93$_{\pm 5.72}$ / 78.29$_{\pm 2.05}$	55.44$_{\pm 2.77}$ / 81.41$_{\pm 1.50}$	48.72$_{\pm 1.87}$ / 83.66$_{\pm 1.67}$	61.23$_{\pm 1.40}$ / 78.74$_{\pm 0.81}$	56.25$_{\pm 1.01}$ / 80.28$_{\pm 0.27}$	55.11$_{\pm 1.65}$ / 80.48$_{\pm 0.93}$
NNGuide	50.78$_{\pm 4.59}$ / 80.49$_{\pm 1.31}$	50.53$_{\pm 2.28}$ / 83.74$_{\pm 1.42}$	43.01$_{\pm 0.49}$ / 86.78$_{\pm 0.72}$	57.08$_{\pm 2.45}$ / 80.94$_{\pm 1.12}$	58.04$_{\pm 1.09}$ / 79.68$_{\pm 0.39}$	51.89$_{\pm 1.35}$ / 82.33$_{\pm 0.66}$
Relation	49.59$_{\pm 3.83}$ / 80.20$_{\pm 1.69}$	55.24$_{\pm 2.16}$ / 82.10$_{\pm 0.16}$	47.72$_{\pm 1.17}$ / 84.91$_{\pm 0.55}$	54.65$_{\pm 2.43}$ / 81.24$_{\pm 0.71}$	61.99$_{\pm 1.42}$ / 79.55$_{\pm 0.35}$	53.84$_{\pm 0.32}$ / 81.60$_{\pm 0.34}$
SCALE	51.64$_{\pm 3.57}$ / 80.27$_{\pm 1.16}$	49.27$_{\pm 2.60}$ / 84.45$_{\pm 1.09}$	49.65$_{\pm 2.51}$ / 84.14$_{\pm 1.07}$	58.44$_{\pm 1.97}$ / 80.50$_{\pm 0.70}$	56.98$_{\pm 0.68}$ / 80.47$_{\pm 0.23}$	53.20$_{\pm 1.27}$ / 81.97$_{\pm 0.54}$
FDBD	51.35$_{\pm 4.46}$ / 79.05$_{\pm 2.00}$	53.81$_{\pm 2.23}$ / 80.48$_{\pm 0.49}$	42.33$_{\pm 0.50}$ / 85.58$_{\pm 0.55}$	53.66$_{\pm 1.04}$ / 81.18$_{\pm 0.72}$	57.17$_{\pm 0.89}$ / 79.85$_{\pm 0.24}$	51.66$_{\pm 0.41}$ / 81.23$_{\pm 0.41}$
			Train-time regularization methods			
ConfBranch	78.40$_{\pm 6.12}$ / 58.59$_{\pm 6.46}$	71.76$_{\pm 4.71}$ / 67.48$_{\pm 5.35}$	84.12$_{\pm 4.15}$ / 57.82$_{\pm 2.48}$	86.69$_{\pm 1.95}$ / 64.54$_{\pm 0.49}$	68.86$_{\pm 1.10}$ / 70.65$_{\pm 1.40}$	77.97$_{\pm 1.16}$ / 63.82$_{\pm 2.02}$
RotPred	25.49$_{\pm 1.79}$ / 92.00$_{\pm 0.42}$	**16.50$_{\pm 2.81}$ / 94.89$_{\pm 0.62}$**	36.97$_{\pm 3.95}$ / 88.39$_{\pm 1.28}$	39.05$_{\pm 0.61}$ / 88.39$_{\pm 0.05}$	59.85$_{\pm 0.37}$ / 76.79$_{\pm 0.11}$	35.57$_{\pm 1.04}$ / 88.09$_{\pm 0.33}$
G-ODIN	25.61$_{\pm 0.45}$ / 92.52$_{\pm 0.13}$	36.58$_{\pm 0.46}$ / 87.42$_{\pm 0.83}$	22.40$_{\pm 1.29}$ / 94.38$_{\pm 0.74}$	**36.84$_{\pm 2.99}$ / 89.45$_{\pm 0.92}$**	63.70$_{\pm 1.89}$ / 78.68$_{\pm 0.71}$	37.03$_{\pm 0.43}$ / 88.49$_{\pm 0.36}$
MOS	51.66$_{\pm 5.50}$ / 81.26$_{\pm 2.44}$	59.31$_{\pm 7.89}$ / 80.68$_{\pm 3.38}$	44.20$_{\pm 0.59}$ / 86.44$_{\pm 1.13}$	59.67$_{\pm 0.72}$ / 80.49$_{\pm 0.24}$	58.92$_{\pm 0.22}$ / 78.55$_{\pm 0.33}$	55.36$_{\pm 2.64}$ / 81.18$_{\pm 0.91}$
VOS	43.08$_{\pm 1.90}$ / 84.47$_{\pm 0.73}$	51.04$_{\pm 13.34}$ / 81.63$_{\pm 6.55}$	47.74$_{\pm 4.09}$ / 84.32$_{\pm 1.93}$	66.16$_{\pm 2.18}$ / 77.42$_{\pm 1.01}$	58.38$_{\pm 1.32}$ / 79.27$_{\pm 0.66}$	53.28$_{\pm 4.01}$ / 81.42$_{\pm 1.95}$
LogitNorm	28.43$_{\pm 5.36}$ / 90.87$_{\pm 2.23}$	41.68$_{\pm 3.68}$ / 85.54$_{\pm 0.93}$	57.74$_{\pm 8.32}$ / 75.04$_{\pm 4.79}$	78.59$_{\pm 2.39}$ / 71.48$_{\pm 1.55}$	55.66$_{\pm 1.57}$ / 80.58$_{\pm 0.77}$	52.42$_{\pm 1.67}$ / 80.70$_{\pm 1.15}$
CIDER	61.34$_{\pm 1.31}$ / 74.89$_{\pm 4.19}$	55.28$_{\pm 10.41}$ / 81.71$_{\pm 4.16}$	68.74$_{\pm 1.43}$ / 74.36$_{\pm 0.90}$	60.19$_{\pm 1.05}$ / 79.10$_{\pm 0.54}$	62.80$_{\pm 1.43}$ / 76.06$_{\pm 0.68}$	61.67$_{\pm 1.69}$ / 77.22$_{\pm 0.89}$
NPOS	44.37$_{\pm 2.20}$ / 87.53$_{\pm 1.09}$	54.47$_{\pm 7.76}$ / 84.15$_{\pm 3.33}$	50.81$_{\pm 5.61}$ / 84.78$_{\pm 2.34}$	52.43$_{\pm 1.95}$ / 83.83$_{\pm 0.51}$	58.37$_{\pm 1.03}$ / 79.99$_{\pm 0.28}$	52.09$_{\pm 2.02}$ / 84.06$_{\pm 0.99}$
ASCOOD (MSP [65])	12.56$_{\pm 1.55}$ / 97.02$_{\pm 0.29}$	39.33$_{\pm 4.45}$ / 87.58$_{\pm 1.45}$	4.28$_{\pm 0.89}$ / 99.09$_{\pm 0.10}$	55.88$_{\pm 1.30}$ / 84.72$_{\pm 0.86}$	**53.55$_{\pm 0.82}$ / 82.29$_{\pm 0.44}$**	33.12$_{\pm 0.89}$ / 90.14$_{\pm 0.17}$
ASCOOD (SCALE)	**6.76$_{\pm 1.50}$ / 98.58$_{\pm 0.27}$**	33.50$_{\pm 3.76}$ / 89.93$_{\pm 1.24}$	**1.88$_{\pm 0.56}$ / 99.37$_{\pm 0.09}$**	53.06$_{\pm 0.81}$ / 86.98$_{\pm 0.81}$	54.30$_{\pm 1.00}$ / 81.91$_{\pm 0.77}$	**29.90$_{\pm 0.76}$ / 91.35$_{\pm 0.13}$**

Table 27. OOD detection results (FPR@95 ↓/ AUROC ↑) in CIFAR-100 benchmark. Refer to Table 9 for OE and MixOE results.

G.5. ImageNet-100 results

Method	**iNaturalist**	**Textures**	**OpenImage-O**	**Average**
		Posthoc methods		
MSP	21.09 / 94.87	62.82 / 83.45	39.82 / 88.08	41.24 / 88.80
TempScale	20.22 / 95.16	63.56 / 83.53	39.91 / 88.35	41.23 / 89.02
ODIN	21.02 / 95.43	79.42 / 80.92	50.42 / 85.60	50.29 / 87.32
MDS	88.60 / 52.93	60.09 / 85.94	80.18 / 66.01	76.29 / 68.29
MDSEns	88.07 / 57.78	73.93 / 75.87	86.33 / 64.50	82.78 / 66.05
RMDS	17.29 / 91.38	39.16 / 85.80	37.40 / 86.95	31.28 / 88.04
Gram	80.33 / 60.61	79.09 / 76.22	79.20 / 68.39	79.54 / 68.40
EBO	22.24 / 94.03	74.67 / 81.04	43.98 / 86.66	46.96 / 87.24
GradNorm	22.98 / 95.18	87.67 / 81.64	66.42 / 82.99	59.02 / 86.60
ReAct	20.89 / 94.68	65.98 / 82.42	41.27 / 87.41	42.71 / 88.17
MLS	21.22 / 94.63	74.64 / 81.29	43.87 / 87.16	46.58 / 87.69
KLM	19.84 / 93.60	70.47 / 82.93	59.51 / 86.78	49.94 / 87.77
VIM	31.49 / 86.05	16.18 / 96.39	30.58 / 90.02	26.08 / 90.82
DICE	27.87 / 93.30	76.60 / 82.79	52.13 / 84.13	52.20 / 86.74
ASH	13.64 / 97.10	50.56 / 90.17	36.76 / 90.20	33.65 / 92.49
SHE	29.51 / 93.28	82.47 / 82.13	61.76 / 81.77	57.91 / 85.72
GEN	18.64 / 94.62	65.60 / 82.48	39.51 / 88.45	41.25 / 88.52
NNGuide	19.73 / 95.59	71.31 / 84.52	43.00 / 87.52	44.68 / 89.21
Relation	16.36 / 95.90	33.27 / 90.64	34.29 / 89.84	27.97 / 92.13
SCALE	12.91 / 97.32	54.13 / 89.77	38.58 / 89.96	35.21 / 92.35
FDBD	11.96 / 97.51	34.51 / 92.49	29.78 / 91.86	25.41 / 93.96
		Train-time regularization methods		
ConfBranch	25.71 / 93.92	72.80 / 83.21	54.58 / 83.82	51.03 / 86.98
RotPred	19.64 / 93.57	50.51 / 88.69	37.44 / 88.64	35.87 / 90.30
G-ODIN	34.98 / 92.22	73.53 / 80.51	57.16 / 84.72	55.22 / 85.82
MOS	- / -	- / -	- / -	- / -
VOS	21.20 / 94.25	64.96 / 83.72	36.82 / 88.26	40.99 / 88.74
LogitNorm	18.38 / 95.75	49.96 / 87.23	34.69 / 89.51	34.34 / 90.83
CIDER	24.91 / 95.03	21.84 / 96.20	46.29 / 89.41	31.01 / 93.54
NPOS	**11.13 / 97.40**	**11.80 / 97.27**	27.60 / **93.12**	**16.84 / 95.93**
Dream-OOD (EBO)	14.47 / 96.09	60.73 / 84.79	32.67 / 90.16	35.96 / 90.35
ASCOOD with $\mathcal{L}_{\text{OOD}} = \mathcal{L}_{\text{energy}}$ (EBO)	17.87 / 95.62	30.69 / 92.92	26.42 / 92.09	24.99 / 93.54
ASCOOD with $\mathcal{L}_{\text{OOD}} = \mathcal{L}_{\text{KL}}$ (EBO)	18.11 / 95.73	25.20 / 94.40	**26.04** / 91.95	23.12 / 94.02

Table 28. Conventional OOD detection results (FPR@95 ↓/ AUROC ↑) in large-scale settings using ImageNet-100 datasets.

G.6. ImageNet-100 results (SSB-Hard)

Method	SSB-Hard	
	FPR@95 ↓	AUROC ↑
Posthoc methods		
MSP	58.33	80.94
TempScale	58.38	80.94
ODIN	69.11	76.11
MDS	81.87	60.21
MDSEns	92.27	51.98
RMDS	57.64	83.65
Gram	89.18	54.09
EBO	61.33	79.23
GradNorm	80.71	70.53
ReAct	60.64	79.64
MLS	61.31	79.62
KLM	75.84	78.90
VIM	59.09	79.62
DICE	66.67	72.59
ASH	65.16	78.17
SHE	72.49	71.42
GEN	58.64	80.91
NNGuide	62.22	79.15
Relation	72.31	77.25
SCALE	65.51	79.04
FDBD	58.87	81.95
Train-time regularization methods		
ConfBranch	67.67	75.23
RotPred	67.64	75.72
G-ODIN	66.16	77.39
MOS	-	-
VOS	62.47	79.62
LogitNorm	59.31	80.88
CIDER	73.07	72.89
NPOS	58.80	80.18
Dream-OOD (EBO)	**55.67**	<u>83.30</u>
ASCOOD with $\mathcal{L}_{\text{OOD}} = \mathcal{L}_{\text{energy}}$ (EBO)	<u>56.09</u>	**83.91**
ASCOOD with $\mathcal{L}_{\text{OOD}} = \mathcal{L}_{\text{KL}}$ (EBO)	57.93	81.66

Table 29. near-OOD detection [81] results (FPR@95 ↓/ AUROC ↑) in large-scale settings using ImageNet-100 datasets.

G.7. Conventional OOD detection on Waterbirds benchmark

Method	NINCO	OpenImage-O	SUN	Texture	iNaturalist	Average
			Posthoc methods			
MSP	57.09$_{\pm 1.77}$ / 80.25$_{\pm 0.75}$	47.13$_{\pm 2.64}$ / 82.43$_{\pm 1.12}$	48.33$_{\pm 1.06}$ / 82.51$_{\pm 0.62}$	34.90$_{\pm 2.11}$ / 87.70$_{\pm 0.45}$	66.92$_{\pm 1.46}$ / 66.33$_{\pm 1.11}$	50.87$_{\pm 1.24}$ / 79.84$_{\pm 0.58}$
TempScale	57.09$_{\pm 1.78}$ / 80.26$_{\pm 0.75}$	47.13$_{\pm 2.64}$ / 82.43$_{\pm 1.11}$	48.31$_{\pm 1.07}$ / 82.51$_{\pm 0.62}$	34.90$_{\pm 2.11}$ / 87.70$_{\pm 0.45}$	66.82$_{\pm 1.50}$ / 66.34$_{\pm 1.11}$	50.85$_{\pm 1.24}$ / 79.85$_{\pm 0.58}$
ODIN	100.00$_{\pm 0.00}$ / 83.15$_{\pm 0.54}$	60.32$_{\pm 28.15}$ / 89.32$_{\pm 0.81}$	61.18$_{\pm 27.47}$ / 86.97$_{\pm 1.47}$	23.26$_{\pm 0.92}$ / 95.18$_{\pm 0.28}$	100.00$_{\pm 0.00}$ / 71.14$_{\pm 0.89}$	68.95$_{\pm 4.82}$ / 85.15$_{\pm 0.11}$
MDS	48.86$_{\pm 3.09}$ / 92.57$_{\pm 0.56}$	46.44$_{\pm 3.35}$ / 92.81$_{\pm 0.56}$	91.93$_{\pm 1.06}$ / 76.82$_{\pm 0.35}$	36.38$_{\pm 6.17}$ / 94.53$_{\pm 1.11}$	18.24$_{\pm 2.48}$ / 96.66$_{\pm 0.37}$	48.37$_{\pm 2.23}$ / 90.68$_{\pm 0.37}$
MDSEns	56.79$_{\pm 0.57}$ / 86.10$_{\pm 0.20}$	60.51$_{\pm 0.05}$ / 85.62$_{\pm 0.10}$	59.42$_{\pm 0.48}$ / 84.09$_{\pm 0.23}$	33.29$_{\pm 0.52}$ / 93.76$_{\pm 0.05}$	38.64$_{\pm 1.28}$ / 91.09$_{\pm 0.41}$	49.73$_{\pm 0.44}$ / 88.13$_{\pm 0.19}$
RMDS	85.76$_{\pm 3.27}$ / 70.98$_{\pm 2.01}$	81.95$_{\pm 3.18}$ / 71.13$_{\pm 2.43}$	74.09$_{\pm 2.64}$ / 75.90$_{\pm 0.67}$	75.93$_{\pm 3.64}$ / 72.41$_{\pm 2.22}$	83.27$_{\pm 2.28}$ / 57.45$_{\pm 4.35}$	80.20$_{\pm 2.41}$ / 69.57$_{\pm 2.08}$
Gram	98.51$_{\pm 0.11}$ / 50.12$_{\pm 0.75}$	98.59$_{\pm 0.18}$ / 55.72$_{\pm 1.43}$	91.95$_{\pm 1.09}$ / 68.88$_{\pm 2.21}$	98.96$_{\pm 0.70}$ / 67.78$_{\pm 2.34}$	98.20$_{\pm 0.44}$ / 45.59$_{\pm 1.52}$	97.24$_{\pm 0.19}$ / 57.62$_{\pm 0.37}$
EBO	57.20$_{\pm 2.32}$ / 80.46$_{\pm 0.87}$	46.09$_{\pm 3.06}$ / 83.33$_{\pm 0.88}$	48.06$_{\pm 2.56}$ / 82.69$_{\pm 1.69}$	34.74$_{\pm 2.44}$ / 88.29$_{\pm 0.38}$	66.52$_{\pm 2.82}$ / 67.66$_{\pm 1.60}$	50.52$_{\pm 2.32}$ / 80.49$_{\pm 0.91}$
GradNorm	91.39$_{\pm 1.88}$ / 69.92$_{\pm 1.55}$	75.22$_{\pm 3.57}$ / 79.29$_{\pm 1.51}$	72.99$_{\pm 5.66}$ / 80.96$_{\pm 1.96}$	51.40$_{\pm 6.44}$ / 88.05$_{\pm 1.25}$	95.08$_{\pm 0.68}$ / 56.54$_{\pm 0.50}$	77.22$_{\pm 0.71}$ / 74.95$_{\pm 0.39}$
ReAct	53.57$_{\pm 2.77}$ / 81.60$_{\pm 0.88}$	43.04$_{\pm 3.02}$ / 84.44$_{\pm 0.82}$	45.80$_{\pm 2.94}$ / 83.50$_{\pm 1.69}$	31.20$_{\pm 1.95}$ / 89.29$_{\pm 0.41}$	61.56$_{\pm 2.24}$ / 69.48$_{\pm 1.35}$	47.03$_{\pm 2.53}$ / 81.66$_{\pm 0.87}$
MLS	57.20$_{\pm 2.32}$ / 80.46$_{\pm 0.89}$	46.09$_{\pm 3.06}$ / 83.26$_{\pm 0.92}$	48.06$_{\pm 2.56}$ / 82.69$_{\pm 1.55}$	34.74$_{\pm 2.44}$ / 88.19$_{\pm 0.40}$	66.52$_{\pm 2.82}$ / 67.57$_{\pm 1.61}$	50.52$_{\pm 2.32}$ / 80.44$_{\pm 0.91}$
KLM	97.94$_{\pm 0.08}$ / 51.10$_{\pm 1.47}$	98.26$_{\pm 0.15}$ / 48.92$_{\pm 1.49}$	98.16$_{\pm 0.03}$ / 49.59$_{\pm 1.09}$	98.29$_{\pm 0.11}$ / 58.45$_{\pm 1.03}$	97.19$_{\pm 0.25}$ / 30.37$_{\pm 0.39}$	97.97$_{\pm 0.11}$ / 47.69$_{\pm 0.83}$
VIM	29.15$_{\pm 1.34}$ / 93.85$_{\pm 0.33}$	22.18$_{\pm 0.78}$ / 95.65$_{\pm 0.13}$	32.57$_{\pm 1.11}$ / 89.65$_{\pm 0.69}$	9.12$_{\pm 1.72}$ / 98.49$_{\pm 0.32}$	33.52$_{\pm 3.00}$ / 91.27$_{\pm 1.08}$	25.31$_{\pm 0.84}$ / 93.78$_{\pm 0.17}$
DICE	61.14$_{\pm 2.50}$ / 81.89$_{\pm 1.73}$	44.17$_{\pm 3.97}$ / 88.07$_{\pm 1.59}$	46.90$_{\pm 1.07}$ / 87.96$_{\pm 0.73}$	28.75$_{\pm 7.28}$ / 94.33$_{\pm 1.76}$	67.40$_{\pm 3.43}$ / 72.85$_{\pm 2.98}$	49.67$_{\pm 2.78}$ / 85.02$_{\pm 1.27}$
RankFeat	72.19$_{\pm 3.99}$ / 68.45$_{\pm 3.14}$	70.99$_{\pm 3.73}$ / 67.24$_{\pm 3.05}$	86.61$_{\pm 8.20}$ / 62.49$_{\pm 7.15}$	52.61$_{\pm 5.02}$ / 75.24$_{\pm 3.86}$	69.26$_{\pm 3.26}$ / 63.94$_{\pm 4.53}$	70.33$_{\pm 4.80}$ / 67.47$_{\pm 4.00}$
ASH	31.80$_{\pm 1.84}$ / 89.22$_{\pm 1.44}$	23.93$_{\pm 2.22}$ / 91.64$_{\pm 1.21}$	30.92$_{\pm 1.17}$ / 89.14$_{\pm 0.33}$	10.48$_{\pm 1.86}$ / 95.53$_{\pm 0.96}$	30.32$_{\pm 4.09}$ / 88.21$_{\pm 2.32}$	25.49$_{\pm 1.88}$ / 90.75$_{\pm 1.24}$
SHE	85.43$_{\pm 2.57}$ / 67.99$_{\pm 1.22}$	73.86$_{\pm 3.47}$ / 78.54$_{\pm 1.34}$	67.23$_{\pm 4.93}$ / 82.52$_{\pm 2.38}$	54.35$_{\pm 6.32}$ / 89.13$_{\pm 1.68}$	89.49$_{\pm 0.77}$ / 61.41$_{\pm 0.75}$	74.07$_{\pm 0.92}$ / 75.92$_{\pm 0.23}$
GEN	57.09$_{\pm 1.78}$ / 80.26$_{\pm 0.75}$	47.13$_{\pm 2.64}$ / 82.43$_{\pm 1.11}$	48.31$_{\pm 1.08}$ / 82.51$_{\pm 0.62}$	34.90$_{\pm 2.11}$ / 87.70$_{\pm 0.45}$	66.82$_{\pm 1.50}$ / 66.34$_{\pm 1.11}$	50.85$_{\pm 1.24}$ / 79.85$_{\pm 0.58}$
NNGuide	51.95$_{\pm 1.94}$ / 83.05$_{\pm 0.61}$	39.95$_{\pm 2.97}$ / 86.04$_{\pm 0.69}$	43.37$_{\pm 1.97}$ / 85.08$_{\pm 1.37}$	27.64$_{\pm 1.72}$ / 91.59$_{\pm 0.09}$	58.30$_{\pm 1.93}$ / 71.55$_{\pm 1.04}$	44.24$_{\pm 1.72}$ / 83.46$_{\pm 0.55}$
Relation	28.97$_{\pm 1.34}$ / 89.89$_{\pm 0.34}$	20.48$_{\pm 1.30}$ / 92.01$_{\pm 0.41}$	27.21$_{\pm 0.67}$ / 90.10$_{\pm 0.39}$	10.27$_{\pm 0.31}$ / 95.89$_{\pm 0.08}$	26.87$_{\pm 1.33}$ / 86.53$_{\pm 0.51}$	22.76$_{\pm 0.59}$ / 90.89$_{\pm 0.10}$
SCALE	33.66$_{\pm 0.96}$ / <u>92.32$_{\pm 0.28}$</u>	<u>21.99$_{\pm 0.58}$</u> / <u>95.27$_{\pm 0.17}$</u>	32.10$_{\pm 2.35}$ / 92.92$_{\pm 0.56}$	4.15$_{\pm 0.32}$ / 98.79$_{\pm 0.40}$	27.11$_{\pm 1.86}$ / 92.86$_{\pm 0.95}$	23.80$_{\pm 0.83}$ / 94.43$_{\pm 0.43}$
FDBD	41.43$_{\pm 0.74}$ / 85.03$_{\pm 0.36}$	35.88$_{\pm 1.64}$ / 86.15$_{\pm 0.70}$	39.62$_{\pm 0.65}$ / 84.99$_{\pm 0.55}$	24.73$_{\pm 0.51}$ / 90.97$_{\pm 0.17}$	47.85$_{\pm 0.88}$ / 75.30$_{\pm 0.62}$	37.90$_{\pm 0.43}$ / 84.49$_{\pm 0.27}$
			Train-time regularization methods			
ConfBranch	64.48$_{\pm 1.90}$ / 78.51$_{\pm 0.97}$	52.45$_{\pm 2.78}$ / 81.56$_{\pm 2.40}$	41.34$_{\pm 2.09}$ / 89.40$_{\pm 0.87}$	39.91$_{\pm 2.17}$ / 87.49$_{\pm 1.73}$	71.17$_{\pm 5.34}$ / 63.50$_{\pm 5.05}$	53.87$_{\pm 2.65}$ / 80.09$_{\pm 2.09}$
RotPred	36.42$_{\pm 1.52}$ / 92.08$_{\pm 0.26}$	30.42$_{\pm 1.56}$ / 93.03$_{\pm 0.53}$	35.28$_{\pm 2.18}$ / 88.30$_{\pm 1.12}$	<u>11.79$_{\pm 0.81}$</u> / <u>98.10$_{\pm 0.13}$</u>	47.36$_{\pm 3.19}$ / 84.00$_{\pm 1.91}$	32.25$_{\pm 1.14}$ / 91.10$_{\pm 0.62}$
G-ODIN	53.66$_{\pm 14.51}$ / 86.20$_{\pm 2.23}$	44.06$_{\pm 15.83}$ / 89.47$_{\pm 1.90}$	53.12$_{\pm 22.37}$ / 86.27$_{\pm 3.97}$	24.89$_{\pm 3.75}$ / 94.23$_{\pm 1.24}$	44.48$_{\pm 4.74}$ / 81.38$_{\pm 4.55}$	44.04$_{\pm 10.87}$ / 87.51$_{\pm 2.03}$
MOS	- / -	- / -	- / -	- / -	- / -	- / -
VOS	59.45$_{\pm 6.24}$ / 80.58$_{\pm 2.20}$	48.07$_{\pm 3.43}$ / 83.32$_{\pm 1.41}$	45.86$_{\pm 4.10}$ / 83.59$_{\pm 1.03}$	34.85$_{\pm 3.95}$ / 88.76$_{\pm 1.11}$	72.62$_{\pm 7.04}$ / 64.42$_{\pm 5.06}$	52.17$_{\pm 4.86}$ / 80.13$_{\pm 2.11}$
LogitNorm	62.43$_{\pm 3.52}$ / 78.29$_{\pm 1.80}$	53.08$_{\pm 2.99}$ / 79.40$_{\pm 1.82}$	51.19$_{\pm 5.01}$ / 80.92$_{\pm 0.92}$	40.10$_{\pm 3.19}$ / 84.98$_{\pm 1.32}$	79.42$_{\pm 16.03}$ / 61.83$_{\pm 5.47}$	57.24$_{\pm 5.36}$ / 77.08$_{\pm 2.00}$
CIDER	40.63$_{\pm 21.77}$ / 91.17$_{\pm 4.98}$	33.62$_{\pm 24.74}$ / 93.93$_{\pm 4.11}$	44.45$_{\pm 22.80}$ / 88.14$_{\pm 7.34}$	9.53$_{\pm 8.44}$ / 98.17$_{\pm 1.49}$	24.44$_{\pm 13.93}$ / 92.85$_{\pm 4.63}$	30.53$_{\pm 18.29}$ / 92.85$_{\pm 4.50}$
NPOS	23.69$_{\pm 2.52}$ / 95.30$_{\pm 0.56}$	15.17$_{\pm 2.87}$ / 97.08$_{\pm 0.44}$	32.29$_{\pm 6.42}$ / 92.92$_{\pm 1.30}$	1.04$_{\pm 0.22}$ / 99.67$_{\pm 0.13}$	18.21$_{\pm 1.71}$ / 95.07$_{\pm 0.62}$	18.08$_{\pm 2.42}$ / 96.01$_{\pm 0.44}$
OE (ODIN)	79.86$_{\pm 28.28}$ / 89.86$_{\pm 1.69}$	23.06$_{\pm 5.11}$ / 95.11$_{\pm 1.08}$	17.20$_{\pm 1.52}$ / 96.28$_{\pm 0.32}$	10.52$_{\pm 0.73}$ / 98.02$_{\pm 0.07}$	79.86$_{\pm 28.48}$ / 82.48$_{\pm 3.94}$	42.10$_{\pm 12.35}$ / 92.35$_{\pm 1.38}$
MixOE (ODIN)	54.56$_{\pm 6.05}$ / 86.20$_{\pm 1.96}$	34.04$_{\pm 5.88}$ / 91.60$_{\pm 1.98}$	32.71$_{\pm 3.41}$ / 92.06$_{\pm 1.13}$	22.01$_{\pm 4.99}$ / 94.80$_{\pm 1.52}$	42.33$_{\pm 5.73}$ / 83.77$_{\pm 4.54}$	37.13$_{\pm 4.17}$ / 89.69$_{\pm 2.15}$
ASCOOD (MSP)	<u>25.85$_{\pm 1.23}$</u> / 91.85$_{\pm 0.28}$	23.27$_{\pm 1.08}$ / 92.02$_{\pm 0.31}$	<u>18.58$_{\pm 1.52}$</u> / <u>93.14$_{\pm 0.11}$</u>	16.18$_{\pm 1.06}$ / 94.23$_{\pm 0.33}$	<u>17.93$_{\pm 0.61}$</u> / <u>93.77$_{\pm 0.61}$</u>	<u>20.36$_{\pm 0.84}$</u> / <u>93.00$_{\pm 0.11}$</u>
ASCOOD (ODIN)	**16.30$_{\pm 1.20}$ / 96.93$_{\pm 0.07}$**	**12.52$_{\pm 0.51}$ / 97.51$_{\pm 0.07}$**	**9.08$_{\pm 1.04}$ / 98.26$_{\pm 0.12}$**	**3.94$_{\pm 0.17}$ / 99.19$_{\pm 0.06}$**	**8.28$_{\pm 1.07}$ / 98.34$_{\pm 0.11}$**	**10.02$_{\pm 0.38}$ / 98.04$_{\pm 0.02}$**

Table 30. OOD detection results (FPR@95 ↓/ AUROC ↑) on various conventional OOD datasets in Waterbirds benchmark.

G.8. Spurious OOD detection on Waterbirds benchmark with MSP

Method	FPR@95↓	AUROC↑
Cross-entropy	60.41$_{\pm 1.52}$	77.18$_{\pm 0.66}$
ASCOOD	**24.11$_{\pm 0.57}$**	**90.75$_{\pm 0.08}$**

Table 31. Spurious OOD detection results in Waterbirds benchmark with MSP scoring.

G.9. Conventional OOD detection on CelebA benchmark

Method	NINCO	OpenImage-O	SUN	Texture	iNaturalist	Average
			Posthoc methods			
MSP	11.01$_{\pm 0.73}$ / 96.12$_{\pm 0.27}$	9.11$_{\pm 0.74}$ / 96.86$_{\pm 0.31}$	9.28$_{\pm 1.50}$ / 96.42$_{\pm 0.91}$	9.32$_{\pm 1.09}$ / 97.05$_{\pm 0.26}$	8.38$_{\pm 1.03}$ / 96.99$_{\pm 0.54}$	9.42$_{\pm 0.97}$ / 96.69$_{\pm 0.45}$
TempScale	11.01$_{\pm 0.73}$ / 96.12$_{\pm 0.27}$	9.11$_{\pm 0.74}$ / 96.86$_{\pm 0.31}$	9.28$_{\pm 1.50}$ / 96.42$_{\pm 0.91}$	9.32$_{\pm 1.09}$ / 97.05$_{\pm 0.26}$	8.38$_{\pm 1.03}$ / 96.99$_{\pm 0.54}$	9.42$_{\pm 0.97}$ / 96.69$_{\pm 0.45}$
ODIN	1.83$_{\pm 0.41}$ / 99.50$_{\pm 0.25}$	0.94$_{\pm 0.20}$ / 99.71$_{\pm 0.11}$	0.59$_{\pm 0.34}$ / 99.85$_{\pm 0.08}$	0.73$_{\pm 0.31}$ / 99.59$_{\pm 0.13}$	0.45$_{\pm 0.20}$ / 99.81$_{\pm 0.09}$	0.91$_{\pm 0.28}$ / 99.69$_{\pm 0.13}$
MDS	27.69$_{\pm 9.50}$ / 95.37$_{\pm 1.48}$	27.91$_{\pm 8.93}$ / 95.37$_{\pm 1.32}$	12.26$_{\pm 12.44}$ / 97.30$_{\pm 2.49}$	34.11$_{\pm 14.72}$ / 94.77$_{\pm 1.75}$	57.76$_{\pm 14.71}$ / 87.45$_{\pm 6.28}$	32.15$_{\pm 10.27}$ / 94.05$_{\pm 2.39}$
MDSEns	5.38$_{\pm 0.21}$ / 98.87$_{\pm 0.03}$	6.71$_{\pm 0.12}$ / 98.65$_{\pm 0.02}$	1.56$_{\pm 0.10}$ / 99.64$_{\pm 0.02}$	3.22$_{\pm 0.07}$ / 99.28$_{\pm 0.02}$	2.45$_{\pm 0.03}$ / 99.43$_{\pm 0.01}$	3.87$_{\pm 0.07}$ / 99.17$_{\pm 0.02}$
RMDS	30.52$_{\pm 9.83}$ / 91.57$_{\pm 1.90}$	27.59$_{\pm 7.92}$ / 92.00$_{\pm 1.67}$	43.20$_{\pm 24.73}$ / 87.92$_{\pm 6.43}$	28.15$_{\pm 5.65}$ / 92.28$_{\pm 0.77}$	23.92$_{\pm 12.72}$ / 92.78$_{\pm 3.20}$	30.68$_{\pm 11.44}$ / 91.31$_{\pm 2.68}$
Gram	80.49$_{\pm 7.02}$ / 76.83$_{\pm 3.52}$	77.24$_{\pm 4.80}$ / 78.47$_{\pm 2.16}$	67.09$_{\pm 10.60}$ / 81.19$_{\pm 3.92}$	87.39$_{\pm 2.52}$ / 82.98$_{\pm 1.05}$	41.30$_{\pm 9.28}$ / 92.66$_{\pm 1.81}$	70.70$_{\pm 6.70}$ / 82.43$_{\pm 2.46}$
EBO	12.09$_{\pm 2.27}$ / 95.66$_{\pm 1.08}$	9.81$_{\pm 1.54}$ / 96.44$_{\pm 0.88}$	9.68$_{\pm 3.02}$ / 96.24$_{\pm 1.87}$	10.15$_{\pm 1.91}$ / 96.64$_{\pm 0.60}$	8.83$_{\pm 1.86}$ / 96.89$_{\pm 0.95}$	10.11$_{\pm 2.07}$ / 96.37$_{\pm 1.05}$
GradNorm	24.21$_{\pm 9.83}$ / 94.79$_{\pm 1.60}$	13.81$_{\pm 5.63}$ / 96.64$_{\pm 0.96}$	13.34$_{\pm 6.65}$ / 96.55$_{\pm 1.46}$	7.52$_{\pm 2.49}$ / 97.87$_{\pm 0.54}$	7.57$_{\pm 3.02}$ / 97.77$_{\pm 0.82}$	13.29$_{\pm 5.48}$ / 96.72$_{\pm 1.07}$
ReAct	10.85$_{\pm 1.49}$ / 96.06$_{\pm 0.84}$	8.99$_{\pm 0.84}$ / 96.72$_{\pm 0.72}$	8.40$_{\pm 1.71}$ / 96.69$_{\pm 1.42}$	8.43$_{\pm 0.82}$ / 97.15$_{\pm 0.47}$	8.02$_{\pm 1.21}$ / 97.11$_{\pm 0.84}$	8.94$_{\pm 1.15}$ / 96.75$_{\pm 0.82}$
MLS	12.09$_{\pm 2.27}$ / 95.77$_{\pm 0.92}$	9.83$_{\pm 1.53}$ / 96.59$_{\pm 0.75}$	9.71$_{\pm 2.97}$ / 96.25$_{\pm 1.69}$	10.14$_{\pm 1.89}$ / 96.82$_{\pm 0.52}$	8.81$_{\pm 1.85}$ / 96.93$_{\pm 0.86}$	10.12$_{\pm 2.05}$ / 96.47$_{\pm 0.93}$
KLM	98.04$_{\pm 0.34}$ / 80.85$_{\pm 3.10}$	96.37$_{\pm 1.21}$ / 86.60$_{\pm 2.90}$	95.52$_{\pm 3.77}$ / 83.16$_{\pm 8.37}$	95.30$_{\pm 2.06}$ / 88.09$_{\pm 2.67}$	81.01$_{\pm 23.13}$ / 87.56$_{\pm 4.41}$	93.25$_{\pm 6.05}$ / 85.25$_{\pm 4.24}$
VIM	1.77$_{\pm 0.51}$ / 99.67$_{\pm 0.08}$	1.68$_{\pm 0.44}$ / 99.70$_{\pm 0.06}$	0.76$_{\pm 0.45}$ / 99.87$_{\pm 0.08}$	1.87$_{\pm 0.80}$ / 99.67$_{\pm 0.11}$	3.38$_{\pm 0.99}$ / 99.29$_{\pm 0.30}$	1.89$_{\pm 0.52}$ / 99.64$_{\pm 0.10}$
DICE	17.06$_{\pm 3.08}$ / 96.17$_{\pm 0.93}$	11.23$_{\pm 1.91}$ / 97.76$_{\pm 0.55}$	9.66$_{\pm 4.26}$ / 98.03$_{\pm 0.96}$	7.84$_{\pm 2.65}$ / 98.61$_{\pm 0.40}$	6.19$_{\pm 1.82}$ / 98.90$_{\pm 0.32}$	10.40$_{\pm 2.72}$ / 97.89$_{\pm 0.63}$
RankFeat	84.88$_{\pm 6.75}$ / 49.20$_{\pm 14.50}$	81.43$_{\pm 7.29}$ / 50.91$_{\pm 15.30}$	80.72$_{\pm 10.27}$ / 48.16$_{\pm 19.08}$	76.10$_{\pm 8.17}$ / 51.79$_{\pm 16.03}$	81.57$_{\pm 5.17}$ / 54.01$_{\pm 10.52}$	80.94$_{\pm 7.09}$ / 50.81$_{\pm 14.96}$
ASH	4.84$_{\pm 0.37}$ / 98.15$_{\pm 0.33}$	3.49$_{\pm 0.49}$ / 98.46$_{\pm 0.42}$	2.73$_{\pm 0.83}$ / 98.73$_{\pm 0.51}$	2.48$_{\pm 0.93}$ / 98.69$_{\pm 0.45}$	2.72$_{\pm 0.93}$ / 98.70$_{\pm 0.45}$	3.25$_{\pm 0.68}$ / 98.55$_{\pm 0.42}$
SHE	29.81$_{\pm 10.18}$ / 93.31$_{\pm 2.70}$	20.10$_{\pm 7.29}$ / 96.08$_{\pm 1.64}$	19.28$_{\pm 9.80}$ / 96.01$_{\pm 2.31}$	10.66$_{\pm 5.53}$ / 98.20$_{\pm 0.75}$	10.92$_{\pm 6.43}$ / 98.04$_{\pm 1.08}$	18.16$_{\pm 7.80}$ / 96.33$_{\pm 1.69}$
GEN	11.01$_{\pm 0.73}$ / 96.12$_{\pm 0.27}$	9.11$_{\pm 0.74}$ / 96.86$_{\pm 0.31}$	9.28$_{\pm 1.50}$ / 96.42$_{\pm 0.91}$	9.32$_{\pm 1.09}$ / 97.05$_{\pm 0.26}$	8.38$_{\pm 1.03}$ / 96.99$_{\pm 0.54}$	9.42$_{\pm 0.97}$ / 96.69$_{\pm 0.45}$
NNGuide	10.18$_{\pm 1.30}$ / 96.85$_{\pm 0.81}$	7.85$_{\pm 0.84}$ / 97.69$_{\pm 0.62}$	7.49$_{\pm 2.18}$ / 97.60$_{\pm 1.33}$	7.21$_{\pm 1.19}$ / 98.28$_{\pm 0.38}$	7.11$_{\pm 1.31}$ / 98.07$_{\pm 0.70}$	7.97$_{\pm 1.30}$ / 97.70$_{\pm 0.75}$
Relation	0.65$_{\pm 0.05}$ / 99.82$_{\pm 0.01}$	0.27$_{\pm 0.04}$ / 99.88$_{\pm 0.01}$	0.14$_{\pm 0.07}$ / 99.93$_{\pm 0.01}$	0.07$_{\pm 0.02}$ / 99.92$_{\pm 0.00}$	0.17$_{\pm 0.09}$ / 99.92$_{\pm 0.02}$	0.26$_{\pm 0.05}$ / 99.89$_{\pm 0.01}$
SCALE	1.82$_{\pm 0.67}$ / 99.42$_{\pm 0.18}$	0.97$_{\pm 0.30}$ / 99.62$_{\pm 0.20}$	0.44$_{\pm 0.28}$ / 99.75$_{\pm 0.19}$	0.27$_{\pm 0.23}$ / 99.73$_{\pm 0.20}$	0.44$_{\pm 0.36}$ / 99.74$_{\pm 0.22}$	0.79$_{\pm 0.31}$ / 99.65$_{\pm 0.19}$
FDBD	7.67$_{\pm 0.26}$ / 97.22$_{\pm 0.12}$	6.45$_{\pm 0.26}$ / 97.76$_{\pm 0.18}$	6.28$_{\pm 0.75}$ / 97.55$_{\pm 0.54}$	5.97$_{\pm 0.33}$ / 98.07$_{\pm 0.13}$	6.25$_{\pm 0.72}$ / 97.81$_{\pm 0.38}$	6.52$_{\pm 0.41}$ / 97.68$_{\pm 0.26}$
			Train-time regularization methods			
ConfBranch	10.77$_{\pm 1.23}$ / 97.31$_{\pm 0.39}$	10.11$_{\pm 0.91}$ / 97.62$_{\pm 0.19}$	7.29$_{\pm 1.04}$ / 98.47$_{\pm 0.20}$	9.97$_{\pm 0.99}$ / 97.92$_{\pm 0.11}$	3.34$_{\pm 0.29}$ / 99.40$_{\pm 0.06}$	8.29$_{\pm 0.59}$ / 98.14$_{\pm 0.13}$
RotPred	2.59$_{\pm 1.98}$ / 99.58$_{\pm 0.23}$	1.25$_{\pm 1.74}$ / 99.72$_{\pm 0.16}$	2.62$_{\pm 1.90}$ / 99.60$_{\pm 0.29}$	1.18$_{\pm 1.00}$ / 99.63$_{\pm 0.09}$	2.49$_{\pm 1.86}$ / 99.61$_{\pm 0.24}$	2.03$_{\pm 1.63}$ / 99.63$_{\pm 0.20}$
G-ODIN	17.84$_{\pm 22.70}$ / 97.99$_{\pm 1.48}$	17.47$_{\pm 22.86}$ / 98.18$_{\pm 1.46}$	1.52$_{\pm 1.85}$ / 99.28$_{\pm 0.71}$	8.19$_{\pm 9.51}$ / 98.49$_{\pm 0.87}$	0.56$_{\pm 0.30}$ / 99.59$_{\pm 0.24}$	9.12$_{\pm 11.43}$ / 98.71$_{\pm 0.95}$
MOS	- / -	- / -	- / -	- / -	- / -	- / -
VOS	12.18$_{\pm 1.86}$ / 95.71$_{\pm 1.04}$	10.05$_{\pm 1.76}$ / 96.56$_{\pm 0.91}$	11.08$_{\pm 3.47}$ / 95.55$_{\pm 2.05}$	10.18$_{\pm 2.02}$ / 96.89$_{\pm 0.86}$	8.81$_{\pm 1.57}$ / 97.00$_{\pm 0.67}$	10.46$_{\pm 2.08}$ / 96.34$_{\pm 1.07}$
LogitNorm	8.94$_{\pm 2.10}$ / 96.98$_{\pm 0.77}$	8.08$_{\pm 1.49}$ / 97.27$_{\pm 0.60}$	5.67$_{\pm 1.41}$ / 98.02$_{\pm 0.50}$	9.63$_{\pm 2.10}$ / 96.82$_{\pm 0.72}$	6.29$_{\pm 1.97}$ / 97.77$_{\pm 0.72}$	7.72$_{\pm 1.80}$ / 97.37$_{\pm 0.65}$
CIDER	2.17$_{\pm 2.89}$ / 99.54$_{\pm 0.55}$	1.06$_{\pm 1.34}$ / 99.73$_{\pm 0.30}$	1.14$_{\pm 1.55}$ / 99.72$_{\pm 0.37}$	1.04$_{\pm 1.41}$ / 99.76$_{\pm 0.26}$	1.06$_{\pm 1.40}$ / 99.77$_{\pm 0.31}$	1.29$_{\pm 1.72}$ / 99.70$_{\pm 0.36}$
NPOS	0.34$_{\pm 0.24}$ / 99.90$_{\pm 0.07}$	0.20$_{\pm 0.16}$ / 99.93$_{\pm 0.02}$	0.14$_{\pm 0.11}$ / 99.97$_{\pm 0.02}$	0.24$_{\pm 0.22}$ / 99.92$_{\pm 0.04}$	0.23$_{\pm 0.23}$ / 99.95$_{\pm 0.05}$	0.23$_{\pm 0.19}$ / 99.93$_{\pm 0.04}$
OE (ODIN)	0.07$_{\pm 0.02}$ / 99.97$_{\pm 0.01}$	0.04$_{\pm 0.03}$ / 99.95$_{\pm 0.00}$	**0.00**$_{\pm 1.55}$ / **100.00**$_{\pm 0.00}$	0.08$_{\pm 0.00}$ / 99.87$_{\pm 0.00}$	**0.00**$_{\pm 0.00}$ / 99.99$_{\pm 0.00}$	0.04$_{\pm 0.01}$ / 99.96$_{\pm 0.00}$
MixOE (ODIN)	8.08$_{\pm 1.97}$ / 98.38$_{\pm 0.38}$	4.19$_{\pm 0.80}$ / 99.07$_{\pm 0.19}$	2.25$_{\pm 0.67}$ / 99.41$_{\pm 0.23}$	9.84$_{\pm 1.84}$ / 97.81$_{\pm 0.56}$	0.83$_{\pm 0.26}$ / 99.71$_{\pm 0.06}$	5.04$_{\pm 0.98}$ / 98.88$_{\pm 0.25}$
ASCOOD (MSP)	3.27$_{\pm 0.79}$ / 99.05$_{\pm 0.24}$	3.04$_{\pm 0.67}$ / 99.06$_{\pm 0.24}$	1.62$_{\pm 0.42}$ / 99.50$_{\pm 0.21}$	2.18$_{\pm 0.46}$ / 99.27$_{\pm 0.17}$	1.87$_{\pm 0.24}$ / 99.37$_{\pm 0.14}$	2.40$_{\pm 0.48}$ / 99.25$_{\pm 0.20}$
ASCOOD (ODIN)	**0.03**$_{\pm 0.02}$ / **99.98**$_{\pm 0.01}$	**0.03**$_{\pm 0.02}$ / **99.95**$_{\pm 0.01}$	0.01$_{\pm 0.02}$ / 99.99$_{\pm 0.00}$	**0.03**$_{\pm 0.02}$ / **99.97**$_{\pm 0.01}$	0.01$_{\pm 0.02}$ / **100.00**$_{\pm 0.00}$	**0.02**$_{\pm 0.02}$ / **99.98**$_{\pm 0.00}$

Table 32. OOD detection results (FPR@95 ↓/ AUROC ↑) on various conventional OOD datasets in CelebA benchmark.

G.10. Spurious OOD detection on CelebA benchmark with MSP

Method	FPR@95↓	AUROC↑
Cross-entropy	56.00$_{\pm 2.73}$	82.53$_{\pm 1.19}$
ASCOOD	**49.56**$_{\pm 0.09}$	**83.30**$_{\pm 0.37}$

Table 33. Spurious OOD detection results in CelebA benchmark with MSP scoring.

G.11. Conventional OOD detection on Aircraft benchmark

Method	NINCO	OpenImage-O	SUN	Texture	iNaturalist	Average
			Posthoc methods			
MSP	27.21$_{\pm 1.68}$ / 93.92$_{\pm 0.45}$	29.67$_{\pm 1.49}$ / 93.31$_{\pm 0.35}$	13.86$_{\pm 0.98}$ / 97.53$_{\pm 0.17}$	34.94$_{\pm 2.37}$ / 92.50$_{\pm 0.52}$	29.98$_{\pm 1.25}$ / 92.48$_{\pm 0.69}$	27.13$_{\pm 0.56}$ / 93.95$_{\pm 0.04}$
TempScale	23.43$_{\pm 1.53}$ / 95.06$_{\pm 0.42}$	26.41$_{\pm 1.70}$ / 94.44$_{\pm 0.36}$	8.39$_{\pm 1.19}$ / 98.12$_{\pm 0.15}$	32.62$_{\pm 2.35}$ / 93.40$_{\pm 0.47}$	25.65$_{\pm 1.12}$ / 93.94$_{\pm 0.51}$	23.30$_{\pm 0.51}$ / 94.99$_{\pm 0.08}$
ODIN	18.23$_{\pm 5.06}$ / 96.07$_{\pm 0.84}$	25.63$_{\pm 4.61}$ / 94.90$_{\pm 0.71}$	4.58$_{\pm 0.83}$ / 98.84$_{\pm 0.16}$	46.55$_{\pm 6.48}$ / 93.23$_{\pm 0.48}$	11.85$_{\pm 3.12}$ / 97.15$_{\pm 0.67}$	21.37$_{\pm 3.66}$ / 96.04$_{\pm 0.49}$
MDS	3.79$_{\pm 0.50}$ / 98.94$_{\pm 0.11}$	2.61$_{\pm 0.02}$ / 99.21$_{\pm 0.08}$	20.96$_{\pm 2.13}$ / 95.16$_{\pm 0.43}$	0.79$_{\pm 0.13}$ / 99.82$_{\pm 0.03}$	0.92$_{\pm 0.25}$ / 99.70$_{\pm 0.11}$	5.82$_{\pm 0.42}$ / 98.57$_{\pm 0.13}$
MDSEns	25.16$_{\pm 1.46}$ / 95.09$_{\pm 0.22}$	25.12$_{\pm 2.04}$ / 95.51$_{\pm 0.27}$	54.06$_{\pm 2.01}$ / 90.02$_{\pm 0.23}$	10.16$_{\pm 0.32}$ / 97.95$_{\pm 0.06}$	5.39$_{\pm 0.30}$ / 98.66$_{\pm 0.05}$	23.98$_{\pm 1.17}$ / 95.45$_{\pm 0.17}$
RMDS	3.51$_{\pm 0.40}$ / 98.97$_{\pm 0.08}$	3.11$_{\pm 0.18}$ / 99.25$_{\pm 0.04}$	4.93$_{\pm 0.47}$ / 97.76$_{\pm 0.16}$	2.02$_{\pm 0.31}$ / 99.66$_{\pm 0.03}$	2.01$_{\pm 0.28}$ / 99.55$_{\pm 0.07}$	3.12$_{\pm 0.30}$ / 99.04$_{\pm 0.05}$
Gram	86.97$_{\pm 3.06}$ / 76.55$_{\pm 2.40}$	94.05$_{\pm 0.78}$ / 70.57$_{\pm 2.69}$	45.90$_{\pm 6.07}$ / 92.11$_{\pm 0.62}$	98.72$_{\pm 0.20}$ / 72.63$_{\pm 1.20}$	82.01$_{\pm 6.23}$ / 72.54$_{\pm 3.97}$	81.53$_{\pm 2.76}$ / 76.88$_{\pm 1.99}$
EBO	13.49$_{\pm 2.56}$ / 96.50$_{\pm 0.54}$	18.86$_{\pm 2.61}$ / 95.54$_{\pm 0.45}$	4.36$_{\pm 0.93}$ / 98.98$_{\pm 0.21}$	47.44$_{\pm 3.90}$ / 92.26$_{\pm 0.21}$	13.52$_{\pm 3.25}$ / 96.34$_{\pm 0.84}$	19.53$_{\pm 1.86}$ / 95.92$_{\pm 0.32}$
GradNorm	71.51$_{\pm 14.21}$ / 90.88$_{\pm 2.02}$	85.76$_{\pm 4.96}$ / 88.28$_{\pm 1.52}$	2.43$_{\pm 0.85}$ / 98.64$_{\pm 0.40}$	94.22$_{\pm 2.06}$ / 87.33$_{\pm 1.09}$	39.38$_{\pm 19.77}$ / 94.50$_{\pm 2.00}$	58.66$_{\pm 6.78}$ / 91.93$_{\pm 1.00}$
ReAct	5.56$_{\pm 0.48}$ / 98.48$_{\pm 0.15}$	6.19$_{\pm 0.34}$ / 98.28$_{\pm 0.12}$	2.57$_{\pm 0.57}$ / 99.48$_{\pm 0.09}$	9.28$_{\pm 0.27}$ / 97.80$_{\pm 0.02}$	4.68$_{\pm 1.08}$ / 98.61$_{\pm 0.30}$	5.66$_{\pm 0.45}$ / 98.53$_{\pm 0.11}$
MLS	13.27$_{\pm 2.39}$ / 96.52$_{\pm 0.51}$	18.05$_{\pm 2.56}$ / 95.62$_{\pm 0.44}$	4.30$_{\pm 0.67}$ / 98.95$_{\pm 0.18}$	45.72$_{\pm 3.50}$ / 92.49$_{\pm 0.24}$	13.22$_{\pm 2.96}$ / 96.30$_{\pm 0.75}$	18.91$_{\pm 1.53}$ / 95.98$_{\pm 0.27}$
KLM	15.93$_{\pm 2.39}$ / 95.80$_{\pm 0.40}$	17.68$_{\pm 3.02}$ / 95.65$_{\pm 0.41}$	7.60$_{\pm 0.93}$ / 97.11$_{\pm 0.08}$	26.09$_{\pm 4.48}$ / 94.61$_{\pm 0.44}$	18.13$_{\pm 1.64}$ / 95.32$_{\pm 0.17}$	17.09$_{\pm 2.09}$ / 95.70$_{\pm 0.26}$
VIM	1.20$_{\pm 0.11}$ / 99.59$_{\pm 0.03}$	0.94$_{\pm 0.04}$ / 99.71$_{\pm 0.03}$	4.17$_{\pm 0.40}$ / 98.82$_{\pm 0.09}$	0.44$_{\pm 0.11}$ / 99.93$_{\pm 0.01}$	0.53$_{\pm 0.09}$ / 99.86$_{\pm 0.05}$	1.46$_{\pm 0.13}$ / 99.58$_{\pm 0.04}$
DICE	34.96$_{\pm 9.02}$ / 94.31$_{\pm 1.28}$	47.42$_{\pm 6.34}$ / 92.53$_{\pm 0.94}$	3.56$_{\pm 1.67}$ / 99.02$_{\pm 0.31}$	77.20$_{\pm 4.63}$ / 90.09$_{\pm 0.58}$	24.45$_{\pm 10.23}$ / 95.71$_{\pm 1.58}$	37.52$_{\pm 4.87}$ / 94.33$_{\pm 0.74}$
RankFeat	69.62$_{\pm 15.84}$ / 65.61$_{\pm 13.49}$	73.76$_{\pm 14.55}$ / 61.67$_{\pm 13.68}$	51.12$_{\pm 18.15}$ / 79.98$_{\pm 12.03}$	82.63$_{\pm 13.63}$ / 63.10$_{\pm 16.77}$	63.84$_{\pm 21.22}$ / 67.44$_{\pm 16.00}$	68.19$_{\pm 16.47}$ / 67.56$_{\pm 14.16}$
ASH	3.80$_{\pm 1.16}$ / 98.76$_{\pm 0.35}$	3.36$_{\pm 0.45}$ / 98.75$_{\pm 0.25}$	0.86$_{\pm 0.10}$ / 99.78$_{\pm 0.05}$	1.88$_{\pm 0.08}$ / 99.27$_{\pm 0.03}$	1.49$_{\pm 0.23}$ / 99.59$_{\pm 0.11}$	2.28$_{\pm 0.28}$ / 99.23$_{\pm 0.09}$
SHE	66.00$_{\pm 10.21}$ / 90.02$_{\pm 1.85}$	77.10$_{\pm 5.39}$ / 87.86$_{\pm 1.22}$	7.39$_{\pm 3.71}$ / 98.21$_{\pm 0.51}$	87.35$_{\pm 3.68}$ / 88.35$_{\pm 0.90}$	50.28$_{\pm 14.43}$ / 91.97$_{\pm 2.45}$	57.63$_{\pm 5.79}$ / 91.28$_{\pm 1.07}$
GEN	11.49$_{\pm 1.61}$ / 96.79$_{\pm 0.43}$	14.78$_{\pm 1.91}$ / 96.14$_{\pm 0.37}$	4.49$_{\pm 0.74}$ / 98.93$_{\pm 0.17}$	31.48$_{\pm 2.78}$ / 93.71$_{\pm 0.22}$	11.93$_{\pm 1.73}$ / 96.45$_{\pm 0.48}$	14.83$_{\pm 1.18}$ / 96.40$_{\pm 0.20}$
NNGuide	9.25$_{\pm 1.92}$ / 97.69$_{\pm 0.45}$	11.59$_{\pm 2.22}$ / 97.17$_{\pm 0.42}$	2.55$_{\pm 0.53}$ / 99.45$_{\pm 0.12}$	20.18$_{\pm 3.20}$ / 96.02$_{\pm 0.29}$	7.71$_{\pm 2.06}$ / 98.13$_{\pm 0.48}$	10.26$_{\pm 1.64}$ / 97.69$_{\pm 0.28}$
Relation	4.19$_{\pm 0.39}$ / 98.72$_{\pm 0.12}$	4.43$_{\pm 0.25}$ / 98.66$_{\pm 0.06}$	2.36$_{\pm 0.25}$ / 99.48$_{\pm 0.05}$	3.47$_{\pm 0.26}$ / 99.09$_{\pm 0.03}$	3.87$_{\pm 0.25}$ / 98.79$_{\pm 0.06}$	3.66$_{\pm 0.20}$ / 98.95$_{\pm 0.04}$
SCALE	2.13$_{\pm 0.00}$ / 99.25$_{\pm 0.00}$	2.37$_{\pm 0.00}$ / 99.03$_{\pm 0.00}$	0.77$_{\pm 0.00}$ / 99.80$_{\pm 0.00}$	1.67$_{\pm 0.00}$ / 99.36$_{\pm 0.00}$	0.97$_{\pm 0.00}$ / 99.73$_{\pm 0.00}$	1.58$_{\pm 0.00}$ / 99.43$_{\pm 0.00}$
FDBD	3.57$_{\pm 0.19}$ / 99.00$_{\pm 0.07}$	3.75$_{\pm 0.11}$ / 98.99$_{\pm 0.04}$	1.90$_{\pm 0.05}$ / 99.61$_{\pm 0.03}$	2.91$_{\pm 0.20}$ / 99.33$_{\pm 0.02}$	2.93$_{\pm 0.17}$ / 99.15$_{\pm 0.03}$	3.01$_{\pm 0.11}$ / 99.22$_{\pm 0.02}$
			Train-time regularization methods			
ConfBranch	15.66$_{\pm 2.08}$ / 96.92$_{\pm 0.66}$	21.46$_{\pm 1.41}$ / 95.96$_{\pm 0.35}$	9.85$_{\pm 0.75}$ / 97.99$_{\pm 0.18}$	30.13$_{\pm 4.95}$ / 95.03$_{\pm 0.49}$	24.43$_{\pm 1.93}$ / 94.44$_{\pm 0.87}$	20.31$_{\pm 0.86}$ / 96.07$_{\pm 0.34}$
RotPred	3.29$_{\pm 1.35}$ / 99.30$_{\pm 0.35}$	3.17$_{\pm 1.07}$ / 99.29$_{\pm 0.22}$	2.02$_{\pm 0.39}$ / 99.56$_{\pm 0.12}$	6.11$_{\pm 1.70}$ / 98.53$_{\pm 0.36}$	1.43$_{\pm 0.18}$ / 99.70$_{\pm 0.05}$	3.21$_{\pm 0.92}$ / 99.28$_{\pm 0.22}$
G-ODIN	9.87$_{\pm 1.76}$ / 97.26$_{\pm 0.55}$	10.79$_{\pm 2.04}$ / 97.13$_{\pm 0.70}$	6.24$_{\pm 0.79}$ / 98.57$_{\pm 0.23}$	8.56$_{\pm 1.24}$ / 97.98$_{\pm 0.35}$	7.67$_{\pm 2.30}$ / 97.95$_{\pm 0.81}$	8.63$_{\pm 1.49}$ / 97.78$_{\pm 0.49}$
MOS	- / -	- / -	- / -	- / -	- / -	- / -
VOS	21.03$_{\pm 5.46}$ / 94.91$_{\pm 1.11}$	39.20$_{\pm 10.18}$ / 91.96$_{\pm 1.61}$	5.52$_{\pm 1.20}$ / 98.63$_{\pm 0.29}$	68.71$_{\pm 12.27}$ / 89.39$_{\pm 1.72}$	27.31$_{\pm 5.21}$ / 93.29$_{\pm 0.85}$	32.36$_{\pm 4.67}$ / 93.63$_{\pm 0.93}$
LogitNorm	10.38$_{\pm 2.54}$ / 97.29$_{\pm 0.60}$	17.05$_{\pm 4.02}$ / 95.91$_{\pm 0.93}$	2.52$_{\pm 0.14}$ / 99.41$_{\pm 0.06}$	26.15$_{\pm 5.33}$ / 94.82$_{\pm 0.81}$	10.26$_{\pm 3.26}$ / 97.08$_{\pm 0.89}$	13.27$_{\pm 3.00}$ / 96.90$_{\pm 0.64}$
CIDER	1.03$_{\pm 0.12}$ / 99.60$_{\pm 0.03}$	1.01$_{\pm 0.12}$ / 99.59$_{\pm 0.03}$	0.70$_{\pm 0.07}$ / 99.78$_{\pm 0.03}$	0.72$_{\pm 0.04}$ / 99.81$_{\pm 0.02}$	0.80$_{\pm 0.11}$ / 99.70$_{\pm 0.07}$	0.85$_{\pm 0.08}$ / 99.69$_{\pm 0.03}$
NPOS	1.07$_{\pm 0.09}$ / 99.63$_{\pm 0.04}$	0.96$_{\pm 0.10}$ / 99.63$_{\pm 0.03}$	0.57$_{\pm 0.03}$ / 99.80$_{\pm 0.04}$	0.40$_{\pm 0.05}$ / 99.90$_{\pm 0.02}$	0.64$_{\pm 0.04}$ / 99.75$_{\pm 0.05}$	0.73$_{\pm 0.05}$ / 99.74$_{\pm 0.04}$
OE	0.87$_{\pm 0.03}$ / 99.72$_{\pm 0.02}$	**0.72$_{\pm 0.02}$** / 99.71$_{\pm 0.01}$	**0.19$_{\pm 0.04}$** / 99.91$_{\pm 0.05}$	0.71$_{\pm 0.14}$ / 99.59$_{\pm 0.07}$	0.19$_{\pm 0.04}$ / 99.92$_{\pm 0.04}$	**0.54$_{\pm 0.02}$** / 99.77$_{\pm 0.04}$
MixOE	1.02$_{\pm 0.20}$ / 99.64$_{\pm 0.10}$	0.93$_{\pm 0.15}$ / 99.61$_{\pm 0.11}$	0.21$_{\pm 0.03}$ / **99.96$_{\pm 0.00}$**	1.60$_{\pm 1.33}$ / 99.13$_{\pm 0.49}$	**0.18$_{\pm 0.04}$ / 99.97$_{\pm 0.01}$**	0.79$_{\pm 0.32}$ / 99.66$_{\pm 0.13}$
ASCOOD (MSP)	10.73$_{\pm 1.31}$ / 97.03$_{\pm 0.13}$	11.80$_{\pm 1.80}$ / 96.88$_{\pm 0.25}$	7.05$_{\pm 0.34}$ / 98.13$_{\pm 0.14}$	7.25$_{\pm 0.47}$ / 98.02$_{\pm 0.05}$	13.67$_{\pm 2.19}$ / 96.60$_{\pm 0.27}$	10.10$_{\pm 1.06}$ / 97.33$_{\pm 0.10}$
ASCOOD (Relation)	5.61$_{\pm 0.18}$ / 98.46$_{\pm 0.03}$	5.84$_{\pm 0.35}$ / 98.41$_{\pm 0.13}$	3.55$_{\pm 0.24}$ / 99.08$_{\pm 0.08}$	3.37$_{\pm 0.05}$ / 99.19$_{\pm 0.07}$	5.46$_{\pm 0.55}$ / 98.59$_{\pm 0.09}$	4.76$_{\pm 0.19}$ / 98.75$_{\pm 0.04}$
ASCOOD (NNGuide)	**0.86$_{\pm 0.08}$ / 99.72$_{\pm 0.04}$**	0.76$_{\pm 0.15}$ / **99.76$_{\pm 0.05}$**	0.54$_{\pm 0.09}$ / 99.84$_{\pm 0.01}$	**0.18$_{\pm 0.04}$ / 99.97$_{\pm 0.00}$**	0.40$_{\pm 0.05}$ / 99.92$_{\pm 0.01}$	0.55$_{\pm 0.07}$ / **99.84$_{\pm 0.02}$**

Table 34. OOD detection results (FPR@95 ↓/ AUROC ↑) on various conventional OOD datasets in Aircraft benchmark.

G.12. Fine-grained OOD detection on Aircraft benchmark with MSP

Method	FPR@95↓	AUROC↑
Cross-entropy	63.79$_{\pm 5.71}$	80.53$_{\pm 1.61}$
ASCOOD	**52.97$_{\pm 0.98}$**	**86.22$_{\pm 0.09}$**

Table 35. Fine-grained OOD detection results in Aircraft benchmark with MSP scoring.

G.13. Conventional OOD detection on Car benchmark

Method	NINCO	OpenImage-O	SUN	Texture	iNaturalist	Average
			Posthoc methods			
MSP	18.04$_{\pm 5.67}$ / 96.23$_{\pm 1.12}$	23.95$_{\pm 8.61}$ / 95.29$_{\pm 1.62}$	4.48$_{\pm 1.73}$ / 99.15$_{\pm 0.25}$	19.88$_{\pm 0.82}$ / 95.99$_{\pm 0.10}$	24.39$_{\pm 18.59}$ / 95.88$_{\pm 3.10}$	18.14$_{\pm 6.55}$ / 96.51$_{\pm 1.17}$
TempScaling	15.31$_{\pm 5.33}$ / 96.77$_{\pm 1.10}$	21.68$_{\pm 8.65}$ / 95.85$_{\pm 1.55}$	2.74$_{\pm 0.96}$ / 99.41$_{\pm 0.16}$	19.10$_{\pm 1.23}$ / 96.31$_{\pm 0.14}$	21.27$_{\pm 18.31}$ / 96.51$_{\pm 2.80}$	16.02$_{\pm 6.60}$ / 96.97$_{\pm 1.10}$
ODIN	13.40$_{\pm 6.54}$ / 97.27$_{\pm 1.17}$	26.22$_{\pm 12.37}$ / 95.50$_{\pm 1.74}$	0.74$_{\pm 0.18}$ / 99.78$_{\pm 0.06}$	33.53$_{\pm 6.80}$ / 94.92$_{\pm 0.61}$	10.19$_{\pm 10.64}$ / 98.13$_{\pm 1.68}$	16.82$_{\pm 7.28}$ / 97.12$_{\pm 1.05}$
MDS	13.16$_{\pm 1.74}$ / 97.07$_{\pm 0.39}$	12.02$_{\pm 1.26}$ / 97.70$_{\pm 0.25}$	32.44$_{\pm 1.72}$ / 90.61$_{\pm 1.01}$	1.44$_{\pm 0.29}$ / 99.70$_{\pm 0.05}$	5.66$_{\pm 1.93}$ / 98.77$_{\pm 0.42}$	12.94$_{\pm 1.27}$ / 96.77$_{\pm 0.41}$
MDSEns	34.59$_{\pm 0.59}$ / 91.03$_{\pm 0.02}$	41.52$_{\pm 0.55}$ / 90.43$_{\pm 0.10}$	36.89$_{\pm 1.28}$ / 90.98$_{\pm 0.25}$	17.97$_{\pm 0.26}$ / 96.23$_{\pm 0.06}$	15.06$_{\pm 0.54}$ / 96.31$_{\pm 0.11}$	29.20$_{\pm 0.39}$ / 93.00$_{\pm 0.10}$
RMDS	5.15$_{\pm 0.17}$ / 98.27$_{\pm 0.04}$	4.78$_{\pm 0.05}$ / 98.67$_{\pm 0.07}$	6.37$_{\pm 0.23}$ / 96.81$_{\pm 0.10}$	2.37$_{\pm 0.21}$ / 99.62$_{\pm 0.03}$	4.60$_{\pm 0.20}$ / 98.46$_{\pm 0.14}$	4.65$_{\pm 0.13}$ / 98.37$_{\pm 0.05}$
Gram	35.64$_{\pm 8.58}$ / 93.61$_{\pm 1.73}$	49.55$_{\pm 7.35}$ / 91.25$_{\pm 1.82}$	4.85$_{\pm 1.23}$ / 99.04$_{\pm 0.20}$	89.12$_{\pm 1.87}$ / 85.93$_{\pm 1.35}$	22.86$_{\pm 6.48}$ / 94.78$_{\pm 2.00}$	40.40$_{\pm 5.02}$ / 92.92$_{\pm 1.41}$
EBO	9.18$_{\pm 3.84}$ / 97.86$_{\pm 0.90}$	18.48$_{\pm 8.67}$ / 96.52$_{\pm 1.35}$	0.54$_{\pm 0.15}$ / 99.84$_{\pm 0.03}$	31.20$_{\pm 4.48}$ / 94.91$_{\pm 0.44}$	7.12$_{\pm 7.49}$ / 98.54$_{\pm 1.35}$	13.30$_{\pm 4.88}$ / 97.53$_{\pm 0.81}$
GradNorm	7.58$_{\pm 3.06}$ / 98.17$_{\pm 0.50}$	18.01$_{\pm 5.92}$ / 96.69$_{\pm 0.72}$	0.08$_{\pm 0.02}$ / 99.96$_{\pm 0.01}$	46.33$_{\pm 6.12}$ / 94.50$_{\pm 0.46}$	0.53$_{\pm 0.53}$ / 99.77$_{\pm 0.15}$	14.51$_{\pm 2.95}$ / 97.82$_{\pm 0.36}$
ReAct	4.92$_{\pm 1.34}$ / 99.08$_{\pm 0.25}$	6.92$_{\pm 1.62}$ / 98.68$_{\pm 0.34}$	0.19$_{\pm 0.05}$ / 99.93$_{\pm 0.02}$	7.73$_{\pm 1.01}$ / 98.52$_{\pm 0.18}$	1.39$_{\pm 1.12}$ / 99.65$_{\pm 0.25}$	4.23$_{\pm 0.90}$ / 99.17$_{\pm 0.17}$
MLS	8.84$_{\pm 3.87}$ / 97.87$_{\pm 0.88}$	17.35$_{\pm 8.20}$ / 96.59$_{\pm 1.33}$	0.73$_{\pm 0.23}$ / 99.80$_{\pm 0.04}$	29.08$_{\pm 4.03}$ / 95.15$_{\pm 0.42}$	7.68$_{\pm 7.57}$ / 98.43$_{\pm 1.41}$	12.74$_{\pm 4.71}$ / 97.57$_{\pm 0.81}$
KLM	13.53$_{\pm 3.74}$ / 95.32$_{\pm 0.76}$	18.60$_{\pm 5.57}$ / 94.65$_{\pm 0.87}$	4.32$_{\pm 0.26}$ / 97.67$_{\pm 0.12}$	14.37$_{\pm 0.95}$ / 95.63$_{\pm 0.04}$	20.53$_{\pm 17.88}$ / 95.49$_{\pm 1.94}$	14.27$_{\pm 5.56}$ / 95.75$_{\pm 0.74}$
VIM	1.61$_{\pm 0.22}$ / 99.66$_{\pm 0.03}$	1.29$_{\pm 0.15}$ / 99.74$_{\pm 0.02}$	3.15$_{\pm 0.28}$ / 99.25$_{\pm 0.12}$	0.08$_{\pm 0.04}$ / 99.98$_{\pm 0.00}$	0.33$_{\pm 0.13}$ / 99.92$_{\pm 0.02}$	1.29$_{\pm 0.13}$ / 99.71$_{\pm 0.04}$
DICE	12.89$_{\pm 7.46}$ / 97.66$_{\pm 1.05}$	26.18$_{\pm 14.40}$ / 95.97$_{\pm 1.61}$	0.05$_{\pm 0.01}$ / 99.94$_{\pm 0.01}$	44.23$_{\pm 7.81}$ / 94.51$_{\pm 0.60}$	3.25$_{\pm 4.38}$ / 99.10$_{\pm 0.97}$	17.32$_{\pm 6.71}$ / 97.43$_{\pm 0.84}$
ASH	0.94$_{\pm 0.27}$ / 99.68$_{\pm 0.08}$	1.51$_{\pm 0.31}$ / 99.53$_{\pm 0.06}$	0.04$_{\pm 0.02}$ / 99.99$_{\pm 0.00}$	0.58$_{\pm 0.06}$ / 99.63$_{\pm 0.08}$	0.13$_{\pm 0.07}$ / 99.96$_{\pm 0.01}$	0.64$_{\pm 0.11}$ / 99.76$_{\pm 0.02}$
SHE	11.68$_{\pm 5.64}$ / 97.84$_{\pm 0.78}$	21.16$_{\pm 8.83}$ / 96.50$_{\pm 1.08}$	0.06$_{\pm 0.01}$ / 99.95$_{\pm 0.01}$	42.04$_{\pm 9.18}$ / 94.78$_{\pm 0.66}$	1.23$_{\pm 1.52}$ / 99.55$_{\pm 0.44}$	15.23$_{\pm 4.99}$ / 97.72$_{\pm 0.59}$
NNGuide	5.41$_{\pm 2.89}$ / 98.74$_{\pm 0.70}$	10.68$_{\pm 6.04}$ / 97.86$_{\pm 1.00}$	0.05$_{\pm 0.01}$ / 99.95$_{\pm 0.02}$	11.31$_{\pm 2.62}$ / 97.52$_{\pm 0.38}$	2.91$_{\pm 3.69}$ / 99.32$_{\pm 0.74}$	6.07$_{\pm 3.05}$ / 98.68$_{\pm 0.57}$
Relation	11.38$_{\pm 3.72}$ / 97.98$_{\pm 0.53}$	13.44$_{\pm 3.26}$ / 97.54$_{\pm 0.67}$	1.14$_{\pm 1.12}$ / 99.73$_{\pm 0.16}$	9.20$_{\pm 0.12}$ / 98.34$_{\pm 0.03}$	6.90$_{\pm 4.73}$ / 98.88$_{\pm 0.68}$	8.41$_{\pm 2.22}$ / 98.49$_{\pm 0.37}$
SCALE	1.06$_{\pm 0.26}$ / 99.68$_{\pm 0.06}$	1.44$_{\pm 0.24}$ / 99.53$_{\pm 0.06}$	0.04$_{\pm 0.01}$ / 99.99$_{\pm 0.00}$	0.66$_{\pm 0.04}$ / 99.62$_{\pm 0.05}$	0.12$_{\pm 0.07}$ / 99.97$_{\pm 0.01}$	0.66$_{\pm 0.12}$ / 99.76$_{\pm 0.03}$
FDBD	2.43$_{\pm 0.95}$ / 99.49$_{\pm 0.21}$	3.24$_{\pm 1.10}$ / 99.37$_{\pm 0.22}$	0.10$_{\pm 0.01}$ / 99.97$_{\pm 0.01}$	1.33$_{\pm 0.28}$ / 99.72$_{\pm 0.03}$	1.43$_{\pm 1.42}$ / 99.68$_{\pm 0.27}$	1.71$_{\pm 0.74}$ / 99.65$_{\pm 0.14}$
			Train-time regularization methods			
ConfBranch	3.35$_{\pm 0.43}$ / 99.23$_{\pm 0.05}$	5.44$_{\pm 1.70}$ / 98.68$_{\pm 0.32}$	0.40$_{\pm 0.06}$ / 99.85$_{\pm 0.01}$	7.53$_{\pm 1.68}$ / 97.88$_{\pm 0.44}$	1.57$_{\pm 0.38}$ / 99.57$_{\pm 0.07}$	3.66$_{\pm 0.68}$ / 99.04$_{\pm 0.15}$
RotPred	0.62$_{\pm 0.31}$ / 99.83$_{\pm 0.06}$	0.88$_{\pm 0.48}$ / 99.66$_{\pm 0.13}$	0.28$_{\pm 0.07}$ / 99.91$_{\pm 0.01}$	0.91$_{\pm 0.40}$ / 99.36$_{\pm 0.26}$	0.14$_{\pm 0.13}$ / 99.95$_{\pm 0.03}$	0.56$_{\pm 0.25}$ / 99.74$_{\pm 0.07}$
G-ODIN	14.14$_{\pm 4.64}$ / 97.20$_{\pm 0.82}$	18.54$_{\pm 6.88}$ / 96.57$_{\pm 1.10}$	1.26$_{\pm 0.07}$ / 99.62$_{\pm 0.03}$	6.59$_{\pm 2.47}$ / 98.66$_{\pm 0.34}$	11.89$_{\pm 8.62}$ / 97.90$_{\pm 1.29}$	10.48$_{\pm 3.85}$ / 97.99$_{\pm 0.62}$
MOS	- / -	- / -	- / -	- / -	- / -	- / -
VOS	6.85$_{\pm 0.55}$ / 98.43$_{\pm 0.13}$	11.64$_{\pm 0.46}$ / 97.56$_{\pm 0.16}$	0.54$_{\pm 0.18}$ / 99.83$_{\pm 0.05}$	35.19$_{\pm 5.19}$ / 94.50$_{\pm 0.54}$	2.56$_{\pm 1.17}$ / 99.32$_{\pm 0.25}$	11.36$_{\pm 1.42}$ / 97.93$_{\pm 0.16}$
LogitNorm	3.38$_{\pm 1.15}$ / 99.22$_{\pm 0.23}$	7.14$_{\pm 2.31}$ / 98.48$_{\pm 0.48}$	0.27$_{\pm 0.02}$ / 99.91$_{\pm 0.01}$	12.29$_{\pm 3.99}$ / 97.54$_{\pm 0.53}$	2.31$_{\pm 1.17}$ / 99.48$_{\pm 0.21}$	5.08$_{\pm 1.55}$ / 98.93$_{\pm 0.27}$
CIDER	1.00$_{\pm 0.06}$ / 99.78$_{\pm 0.01}$	0.90$_{\pm 0.07}$ / 99.81$_{\pm 0.01}$	0.08$_{\pm 0.07}$ / 99.98$_{\pm 0.02}$	0.11$_{\pm 0.01}$ / 99.94$_{\pm 0.01}$	0.04$_{\pm 0.01}$ / 99.98$_{\pm 0.00}$	0.43$_{\pm 0.03}$ / 99.90$_{\pm 0.00}$
NPOS	1.17$_{\pm 0.09}$ / 99.76$_{\pm 0.03}$	1.23$_{\pm 0.11}$ / 99.74$_{\pm 0.02}$	0.04$_{\pm 0.01}$ / 99.98$_{\pm 0.00}$	0.19$_{\pm 0.12}$ / 99.95$_{\pm 0.01}$	0.13$_{\pm 0.05}$ / 99.96$_{\pm 0.01}$	0.55$_{\pm 0.05}$ / 99.88$_{\pm 0.01}$
OE (NNGuide)	**0.04**$_{\pm 0.01}$ / 99.92$_{\pm 0.02}$	**0.05**$_{\pm 0.02}$ / 99.88$_{\pm 0.01}$	0.03$_{\pm 0.01}$ / 99.99$_{\pm 0.00}$	0.09$_{\pm 0.07}$ / 99.81$_{\pm 0.05}$	0.03$_{\pm 0.01}$ / 99.99$_{\pm 0.00}$	**0.05**$_{\pm 0.02}$ / 99.92$_{\pm 0.02}$
MixOE (NNGuide)	0.73$_{\pm 0.11}$ / 99.73$_{\pm 0.10}$	0.89$_{\pm 0.08}$ / 99.67$_{\pm 0.04}$	0.10$_{\pm 0.03}$ / 99.94$_{\pm 0.02}$	2.13$_{\pm 0.14}$ / 99.35$_{\pm 0.05}$	0.22$_{\pm 0.05}$ / 99.92$_{\pm 0.01}$	0.81$_{\pm 0.01}$ / 99.72$_{\pm 0.04}$
ASCOOD (MSP)	7.41$_{\pm 0.60}$ / 97.85$_{\pm 0.13}$	7.37$_{\pm 0.09}$ / 97.97$_{\pm 0.06}$	5.28$_{\pm 1.23}$ / 98.50$_{\pm 0.33}$	4.76$_{\pm 0.19}$ / 98.67$_{\pm 0.08}$	5.79$_{\pm 0.27}$ / 98.29$_{\pm 0.13}$	6.12$_{\pm 0.45}$ / 98.26$_{\pm 0.14}$
ASCOOD (Relation)	5.88$_{\pm 0.56}$ / 98.40$_{\pm 0.11}$	6.25$_{\pm 0.82}$ / 98.39$_{\pm 0.19}$	2.37$_{\pm 0.55}$ / 99.20$_{\pm 0.12}$	2.53$_{\pm 0.15}$ / 99.16$_{\pm 0.07}$	4.37$_{\pm 0.99}$ / 98.79$_{\pm 0.20}$	4.28$_{\pm 0.53}$ / 98.79$_{\pm 0.12}$
ASCOOD (NNGuide)	0.27$_{\pm 0.05}$ / **99.93**$_{\pm 0.01}$	0.12$_{\pm 0.03}$ / **99.96**$_{\pm 0.01}$	**0.03**$_{\pm 0.01}$ / **99.99**$_{\pm 0.00}$	**0.03**$_{\pm 0.00}$ / **99.99**$_{\pm 0.00}$	**0.03**$_{\pm 0.01}$ / **99.99**$_{\pm 0.00}$	0.09$_{\pm 0.02}$ / **99.97**$_{\pm 0.00}$

Table 36. OOD detection results (FPR@95 ↓/ AUROC ↑) on various conventional OOD datasets in Car benchmark.

G.14. Fine-grained OOD detection on Car benchmark with MSP

Method	FPR@95↓	AUROC↑
Cross-entropy	58.17$_{\pm 0.99}$	87.12$_{\pm 0.16}$
ASCOOD	**51.37**$_{\pm 2.82}$	**90.51**$_{\pm 0.36}$

Table 37. Fine-grained OOD detection results in Car benchmark with MSP scoring.