

APPENDIX: STRENGTHEN OUT-OF-DISTRIBUTION DETECTION WITH UNCERTAINTY-DRIVEN ADAPTIVELY RECTIFIED BACKPROPAGATION

Anonymous authors

Paper under double-blind review

A APPENDIX

A.1 RELATED WORKS

A.1.1 OUT-OF-DISTRIBUTION DETECTION

Test-time OOD detection methods. Test-time OOD detection methods have the advantage of being easy to use without modifying the training procedure and objective (Yang et al., 2021). The test-time approaches do not require retraining the model, performs well, and is easy to implement in the real world. Test-time OOD detection methods can be categorized into confidence-based, feature-based, distance-based, gradient-based, pruning-based, and activation-based methods. Confidence-based methods use the confidence score of a pre-trained classifier to detect OOD data. The underlying assumption is that the ID data should receive a high confidence score, while the OOD data should receive a low confidence score. MSP (Hendrycks & Gimpel, 2016) directly uses the maximum SoftMax score to determine whether the test sample is an ID or OOD. ODIN (Liang et al., 2017) improves the SoftMax score by perturbing the input and applying temperature scaling to the logits. Energy (Liu et al., 2020) demonstrates that the Energy score (i.e., logsumexp of logits) outperforms the SoftMax score in distinguishing between ID and OOD data. Feature-based methods include GRAM (Sastry & Oore, 2020) and SHE (Zhang et al., 2023). GRAM computes the gram matrix within the hidden layers. SHE uses the energy function defined in modern Hopfield networks. Distance-based methods consider OOD data to be farther away from the training set than ID data. Mahalanobis (Lee et al., 2018) calculates the minimum Mahalanobis distance between the test data and the class centroids of the training set as an OOD score. Gradient-based approaches (Huang et al., 2021) uses gradient statistics to calculate OOD score. Pruning-based methods prunes the weights of model to address overconfident prediction of OOD data. DICE (Sun & Li, 2022) prunes the weights of the classification layer to address overconfidence in the model’s prediction of OOD data. Activation-based methods attempt to maximize the gap between ID and OOD data by truncating abnormally low or high activations. ReAct (Sun et al., 2021) observes that OOD inputs trigger abnormally high activations. BATS (Zhu et al., 2022; He et al., 2024) exhibits efficacy by truncating both abnormally low and abnormally high activations of each channel.

Training-time OOD detection methods. Unlike testing-time methods, training-time methods aim to mitigate overconfident predictions for OOD data during the training period. According to whether the OOD-supervised signals are used in the training process, training-time methods can be categorized into OOD-free and OOD-needed methods. The representatives of OOD-free methods are (Wei et al., 2022; Lin et al., 2021). Wei et al. (2022) decoupled the influence of logits’ norm from the training procedure by incorporating LogitNorm into the cross-entropy loss. Lin et al. (2021) exploited intermediate classifier outputs for dynamic and efficient OOD inference. The OOD-needed methods aim to calibrate the model by OOD-supervised signals (Ming et al., 2022; Katz-Samuels et al., 2022; Du et al., 2022).

A.1.2 SAMPLE SELECTION

Sample selection has emerged as a powerful technique for enhancing the efficiency and robustness of deep learning model training. Current approaches typically prioritize instances based on their informativeness (Alain et al., 2015), uniqueness (Shi et al., 2021), or confidence levels (Khim et al., 2020), though these methods often incur significant computational overhead. Existing selec-

Table 1: Comparison (FPR95 \downarrow) on CIFAR-10 benchmark. All values are percentages.

Method	CIFAR-100	TIN	Near-Avg	MNIST	SVHM	Texture	Places365	Far-Avg
MSP	52.70 \pm 1.47	42.13 \pm 2.26	47.41 \pm 0.99	25.33 \pm 1.50	23.77 \pm 4.86	28.23 \pm 0.23	40.83 \pm 1.74	29.54 \pm 1.91
MSP+UM	43.63 \pm 1.57	34.49 \pm 2.22	39.06 \pm 1.76	31.55 \pm 10.36	13.43 \pm 1.72	25.85 \pm 1.77	33.50 \pm 3.75	26.08 \pm 1.20
MSP+ours	34.99 \pm 0.97	29.96 \pm 1.03	32.47 \pm 1.00	18.56 \pm 2.17	13.67 \pm 0.86	23.42 \pm 0.92	29.48 \pm 1.05	21.28 \pm 0.75
Energy	67.57 \pm 0.98	56.21 \pm 2.24	61.89 \pm 1.48	28.40 \pm 3.34	36.60 \pm 11.38	43.33 \pm 1.81	52.94 \pm 1.80	40.32 \pm 3.99
Energy+UM	53.51 \pm 3.84	38.17 \pm 3.21	45.84 \pm 3.17	24.29 \pm 7.85	17.50 \pm 8.53	30.47 \pm 6.06	29.67 \pm 5.18	25.48 \pm 2.77
Energy+ours	38.53 \pm 0.69	29.61 \pm 0.82	34.07 \pm 0.61	11.54 \pm 1.38	11.47 \pm 2.53	23.12 \pm 0.58	26.98 \pm 1.47	18.28 \pm 0.54
KNN	37.44 \pm 0.75	29.68 \pm 0.80	33.56 \pm 0.73	20.18 \pm 0.98	20.19 \pm 2.72	21.34 \pm 0.56	28.73 \pm 1.00	22.61 \pm 1.21
KNN+UM	45.11 \pm 2.37	38.65 \pm 3.23	41.88 \pm 2.79	20.00 \pm 0.68	24.09 \pm 4.42	30.44 \pm 2.52	34.39 \pm 3.45	27.23 \pm 0.48
KNN+ours	37.66 \pm 0.58	30.61 \pm 0.64	34.14 \pm 0.59	15.89 \pm 2.45	18.56 \pm 2.97	21.82 \pm 0.98	29.09 \pm 2.12	21.34 \pm 0.82
ReAct	64.95 \pm 3.63	54.60 \pm 2.56	59.78 \pm 3.10	36.84 \pm 7.29	36.47 \pm 9.09	37.99 \pm 5.40	45.70 \pm 2.64	39.25 \pm 4.36
ReAct+UM	54.29 \pm 6.29	40.74 \pm 7.38	47.52 \pm 6.73	29.30 \pm 6.62	20.88 \pm 12.03	29.66 \pm 9.02	28.12 \pm 4.10	26.99 \pm 4.61
ReAct+ours	38.75 \pm 1.37	29.88 \pm 0.72	34.31 \pm 0.86	13.15 \pm 2.46	12.67 \pm 3.22	23.17 \pm 0.46	26.24 \pm 2.47	18.81 \pm 1.13
Relation	40.33 \pm 0.77	32.92 \pm 1.34	36.62 \pm 0.92	23.03 \pm 1.11	20.99 \pm 3.11	23.60 \pm 0.82	34.46 \pm 1.30	25.52 \pm 1.36
Relation+UM	47.41 \pm 4.96	39.67 \pm 3.33	43.54 \pm 5.38	26.97 \pm 1.47	17.98 \pm 1.23	24.23 \pm 1.11	38.10 \pm 1.11	26.82 \pm 1.24
Relation+ours	41.67 \pm 1.02	34.88 \pm 1.02	38.28 \pm 0.99	19.73 \pm 2.91	15.29 \pm 2.01	23.63 \pm 1.30	36.96 \pm 2.73	23.90 \pm 0.97
Fdbd	38.90 \pm 0.79	29.54 \pm 1.28	34.22 \pm 0.79	19.42 \pm 0.43	20.95 \pm 3.51	21.28 \pm 0.93	27.94 \pm 0.73	22.40 \pm 1.30
Fdbd+UM	40.39 \pm 0.43	35.01 \pm 1.13	37.70 \pm 2.06	21.43 \pm 4.46	17.81 \pm 2.95	25.12 \pm 2.84	32.84 \pm 1.89	24.30 \pm 0.93
Fdbd+ours	36.64 \pm 0.36	28.62 \pm 1.16	32.63 \pm 0.60	17.82 \pm 2.25	15.76 \pm 2.82	20.21 \pm 1.19	26.59 \pm 0.74	20.09 \pm 0.84
Nci	52.94 \pm 1.89	41.21 \pm 0.55	47.07 \pm 0.99	33.39 \pm 2.18	23.97 \pm 3.63	23.78 \pm 0.32	36.28 \pm 1.05	29.35 \pm 1.56
Nci+UM	49.84 \pm 3.06	43.58 \pm 3.59	46.71 \pm 3.31	32.19 \pm 1.69	17.26 \pm 2.06	25.36 \pm 1.30	45.03 \pm 4.49	29.96 \pm 1.09
Nci+ours	42.32 \pm 1.08	36.20 \pm 1.76	39.26 \pm 1.40	22.50 \pm 3.99	16.02 \pm 1.32	22.41 \pm 1.23	35.12 \pm 1.28	24.01 \pm 0.92

Table 2: Comparison (AUROC \uparrow) on CIFAR-10 benchmark. All values are percentages.

Method	CIFAR-100	TIN	Near-Avg	MNIST	SVHM	Texture	Places365	Far-Avg
MSP	87.25 \pm 0.34	89.10 \pm 0.38	88.18 \pm 0.34	91.88 \pm 0.39	91.92 \pm 1.12	90.94 \pm 0.19	89.28 \pm 0.35	91.01 \pm 0.47
MSP+UM	86.97 \pm 0.30	89.00 \pm 0.60	87.98 \pm 0.45	89.90 \pm 3.59	95.89 \pm 1.23	91.76 \pm 0.62	89.62 \pm 1.05	91.79 \pm 0.48
MSP+ours	88.55 \pm 0.21	90.27 \pm 0.30	89.41 \pm 0.25	93.85 \pm 0.81	95.30 \pm 1.26	92.24 \pm 0.09	90.54 \pm 0.45	92.98 \pm 0.36
Energy	86.24 \pm 0.48	88.95 \pm 0.50	87.60 \pm 0.47	93.51 \pm 0.59	91.54 \pm 1.91	90.78 \pm 0.34	89.75 \pm 0.41	91.40 \pm 0.75
Energy+UM	86.89 \pm 0.36	90.28 \pm 0.56	88.59 \pm 0.33	93.83 \pm 2.11	95.70 \pm 1.90	91.91 \pm 1.08	92.94 \pm 1.09	93.60 \pm 0.36
Energy+ours	89.73 \pm 0.05	92.12 \pm 0.22	90.93 \pm 0.13	97.22 \pm 0.45	97.17 \pm 1.02	93.75 \pm 0.21	93.11 \pm 0.52	95.31 \pm 0.23
KNN	89.78 \pm 0.30	91.76 \pm 0.28	90.77 \pm 0.28	93.95 \pm 0.12	93.52 \pm 0.83	93.65 \pm 0.07	92.14 \pm 0.32	93.31 \pm 0.29
KNN+UM	86.61 \pm 0.54	88.34 \pm 1.09	87.48 \pm 0.82	92.80 \pm 1.76	90.76 \pm 1.52	89.83 \pm 1.05	90.79 \pm 0.78	91.50 \pm 0.30
KNN+ours	89.19 \pm 0.13	91.22 \pm 0.17	90.20 \pm 0.14	95.66 \pm 1.06	93.83 \pm 1.76	93.55 \pm 0.14	92.37 \pm 0.58	93.85 \pm 0.63
ReAct	86.63 \pm 0.54	89.18 \pm 0.30	87.90 \pm 0.42	92.30 \pm 1.02	91.42 \pm 1.45	91.42 \pm 0.64	90.45 \pm 0.45	91.40 \pm 0.68
ReAct+UM	86.72 \pm 0.89	90.02 \pm 0.57	88.37 \pm 0.69	92.80 \pm 1.76	95.14 \pm 2.39	92.38 \pm 1.35	93.11 \pm 1.01	93.36 \pm 0.70
ReAct+ours	89.66 \pm 0.01	92.02 \pm 0.19	90.84 \pm 0.09	96.92 \pm 0.66	96.90 \pm 1.25	93.86 \pm 0.07	93.22 \pm 0.84	95.22 \pm 0.34
Relation	89.01 \pm 0.34	90.94 \pm 0.31	89.97 \pm 0.31	93.26 \pm 0.25	93.31 \pm 1.06	92.90 \pm 0.09	90.78 \pm 0.38	92.56 \pm 0.40
Relation+UM	86.27 \pm 0.49	88.85 \pm 0.70	87.56 \pm 0.79	91.39 \pm 0.42	94.73 \pm 0.25	92.62 \pm 0.32	88.78 \pm 0.21	91.88 \pm 0.19
Relation+ours	88.05 \pm 0.54	89.91 \pm 0.64	88.98 \pm 0.59	94.01 \pm 0.61	95.18 \pm 1.30	92.72 \pm 0.64	90.04 \pm 1.06	92.99 \pm 0.08
Fdbd	89.59 \pm 0.25	91.77 \pm 0.29	90.68 \pm 0.26	94.46 \pm 0.12	93.32 \pm 1.13	93.73 \pm 0.13	92.21 \pm 0.23	93.43 \pm 0.38
Fdbd+UM	87.67 \pm 0.89	91.09 \pm 0.57	89.38 \pm 0.51	92.62 \pm 1.72	95.08 \pm 1.56	92.35 \pm 1.44	91.79 \pm 1.01	93.21 \pm 0.22
Fdbd+ours	89.52 \pm 0.14	91.96 \pm 0.24	90.74 \pm 0.19	95.16 \pm 0.71	95.09 \pm 1.90	94.20 \pm 0.32	92.84 \pm 0.31	94.32 \pm 0.52
Nci	87.81 \pm 0.37	89.69 \pm 0.22	88.75 \pm 0.26	90.96 \pm 0.26	92.27 \pm 1.05	92.68 \pm 0.16	90.30 \pm 0.31	91.55 \pm 0.40
Nci+UM	86.33 \pm 0.63	88.16 \pm 0.88	87.24 \pm 0.76	91.19 \pm 0.62	94.88 \pm 0.85	92.68 \pm 0.33	87.96 \pm 1.10	91.68 \pm 0.14
Nci+ours	88.09 \pm 0.29	89.88 \pm 0.38	88.98 \pm 0.33	93.58 \pm 1.23	95.52 \pm 1.10	93.52 \pm 0.34	90.28 \pm 0.32	93.22 \pm 0.42

tion strategies can be broadly categorized into two paradigms: (1) static selection methods like Data Pruning (Killamsetty et al., 2021b) and Core Set (Xia et al., 2023), which identify representative subsets before training, and (2) dynamic approaches such as Dynamic Data Pruning (Qin et al., 2023) and Importance Sampling (Jiang et al., 2019), which continuously adjust sample selection during training. Various metrics have been proposed to quantify sample informativeness, including gradient norms (Killamsetty et al., 2021a), loss values (Mindermann et al., 2022), and prediction uncertainty (Chang et al., 2017). While our method UARB shares the fundamental objective of optimizing training through selective instance participation, our method introduces a novel uncertainty-driven selection criterion based on whether the model has “mastered” a given instance. This key distinction enables UARB to automatically adapt the training subset composition without requiring predefined schedules or fixed removal rates, offering significant advantages over conventional approaches. Different from the existing sample selection techniques have primarily focused on classification tasks, UARB is dedicated to enhancing OOD detection. Additionally, we develop an adaptive strategy by incorporating class-informed statistics to determine when mastering an instance.

A.2 USE OF LLMs

We only sought assistance from LLMs for language polishing.

Table 3: Comparison (FPR95 \downarrow) on CIFAR-100 benchmark. All values are percentages.

Method	CIFAR-10	TIN	Near-Avg	MNIST	SVHN	Texture	Places365	Far-Avg
MSP	59.70 \pm 0.73	50.60 \pm 0.95	55.15 \pm 0.13	59.49 \pm 1.46	56.51 \pm 5.82	61.59 \pm 0.85	56.70 \pm 0.75	58.57 \pm 1.50
MSP+UM	59.35 \pm 0.58	51.19 \pm 0.44	55.27 \pm 0.24	56.21 \pm 2.12	55.02 \pm 2.49	60.87 \pm 2.67	57.78 \pm 0.21	57.47 \pm 1.14
MSP+ours	59.13 \pm 0.26	50.87 \pm 0.56	55.00 \pm 0.34	54.85 \pm 1.31	50.32 \pm 7.88	60.22 \pm 0.45	55.82 \pm 0.41	55.30 \pm 2.17
Energy	60.21 \pm 0.81	51.77 \pm 0.53	55.99 \pm 0.27	55.13 \pm 1.33	49.04 \pm 7.13	63.00 \pm 0.29	56.92 \pm 0.82	56.02 \pm 1.58
Energy+UM	60.65 \pm 1.21	54.07 \pm 0.44	57.36 \pm 0.48	51.96 \pm 2.10	47.26 \pm 1.17	63.49 \pm 3.47	59.85 \pm 0.19	55.64 \pm 1.09
Energy+ours	59.50 \pm 0.47	51.59 \pm 0.65	55.54 \pm 0.54	51.65 \pm 1.62	43.20 \pm 9.54	60.32 \pm 0.69	57.02 \pm 0.67	53.05 \pm 2.70
KNN	73.70 \pm 1.92	49.47 \pm 0.37	61.59 \pm 0.83	50.81 \pm 1.44	56.54 \pm 10.90	52.73 \pm 1.86	59.99 \pm 0.59	55.02 \pm 3.43
KNN+UM	77.03 \pm 0.78	51.29 \pm 0.55	64.16 \pm 0.46	43.34 \pm 3.76	47.75 \pm 4.16	51.06 \pm 0.39	64.00 \pm 0.35	51.54 \pm 2.08
KNN+ours	72.09 \pm 2.04	49.73 \pm 0.36	60.91 \pm 1.18	46.46 \pm 3.99	46.32 \pm 7.30	53.40 \pm 2.27	60.02 \pm 0.87	51.55 \pm 2.94
ReAct	61.48 \pm 1.16	51.70 \pm 0.57	56.59 \pm 0.31	57.66 \pm 1.46	45.78 \pm 5.58	56.34 \pm 0.83	55.34 \pm 0.88	53.78 \pm 1.18
ReAct+UM	62.80 \pm 0.91	53.84 \pm 0.50	58.32 \pm 0.21	55.66 \pm 1.36	46.37 \pm 1.79	55.33 \pm 2.08	59.04 \pm 0.55	54.10 \pm 0.57
ReAct+ours	61.17 \pm 0.57	51.61 \pm 0.76	56.39 \pm 0.54	53.34 \pm 1.78	39.90 \pm 8.02	52.87 \pm 0.58	55.26 \pm 0.93	50.34 \pm 2.38
Relation	72.03 \pm 2.55	49.56 \pm 0.59	60.80 \pm 0.99	52.46 \pm 1.97	57.81 \pm 9.79	54.34 \pm 2.29	61.57 \pm 0.75	56.55 \pm 3.25
Relation+UM	72.16 \pm 0.75	50.13 \pm 0.16	61.14 \pm 0.30	46.34 \pm 2.01	49.38 \pm 3.92	52.30 \pm 0.81	62.60 \pm 0.42	52.66 \pm 1.63
Relation+ours	71.62 \pm 1.48	50.17 \pm 0.51	60.89 \pm 0.94	48.41 \pm 3.71	47.37 \pm 8.66	54.55 \pm 1.69	60.77 \pm 1.19	52.77 \pm 3.20
Fdbd	65.41 \pm 1.42	47.81 \pm 0.66	56.61 \pm 0.73	55.02 \pm 1.35	54.85 \pm 7.13	53.64 \pm 1.47	57.19 \pm 0.68	55.18 \pm 2.12
Fdbd+UM	66.76 \pm 0.86	48.67 \pm 0.24	57.72 \pm 0.55	47.22 \pm 1.00	48.51 \pm 3.38	52.05 \pm 1.47	58.62 \pm 0.24	51.60 \pm 1.04
Fdbd+ours	64.46 \pm 1.20	48.33 \pm 0.66	56.39 \pm 0.81	49.37 \pm 3.07	46.78 \pm 7.62	52.84 \pm 1.77	56.26 \pm 0.17	51.31 \pm 2.83
Nci	63.47 \pm 1.47	48.06 \pm 0.79	55.76 \pm 0.60	54.74 \pm 2.24	49.47 \pm 5.89	47.71 \pm 0.91	53.83 \pm 0.30	51.44 \pm 1.52
Nci+UM	63.84 \pm 0.40	49.34 \pm 0.34	56.59 \pm 0.18	47.64 \pm 3.48	45.04 \pm 0.97	46.00 \pm 1.06	55.09 \pm 0.41	48.44 \pm 0.97
Nci+ours	63.96 \pm 0.88	49.51 \pm 0.75	56.74 \pm 0.28	49.71 \pm 2.39	43.53 \pm 5.67	47.29 \pm 1.03	53.17 \pm 0.21	48.43 \pm 1.96

Table 4: Comparison (AUROC \uparrow) on CIFAR-100 benchmark. All values are percentages.

Method	CIFAR-10	TIN	Near-Avg	MNIST	SVHN	Texture	Places365	Far-Avg
MSP	78.37 \pm 0.20	82.16 \pm 0.21	80.27 \pm 0.11	75.55 \pm 0.50	79.97 \pm 2.60	77.34 \pm 0.78	79.37 \pm 0.35	78.06 \pm 0.88
MSP+UM	78.29 \pm 0.11	82.07 \pm 0.26	80.18 \pm 0.16	77.53 \pm 1.54	80.46 \pm 0.84	78.05 \pm 0.90	78.99 \pm 0.18	78.76 \pm 0.50
MSP+ours	78.64 \pm 0.06	82.25 \pm 0.13	80.45 \pm 0.09	76.93 \pm 0.22	82.42 \pm 2.58	78.04 \pm 0.40	79.62 \pm 0.26	79.25 \pm 0.71
Energy	78.95 \pm 0.23	82.80 \pm 0.12	80.88 \pm 0.08	78.55 \pm 0.32	84.11 \pm 2.92	77.92 \pm 0.89	79.73 \pm 0.56	80.08 \pm 0.99
Energy+UM	78.48 \pm 0.36	82.22 \pm 0.17	80.35 \pm 0.21	79.90 \pm 1.75	83.81 \pm 0.12	78.56 \pm 1.25	78.84 \pm 0.15	80.28 \pm 0.76
Energy+ours	79.25 \pm 0.15	82.87 \pm 0.20	81.06 \pm 0.17	79.35 \pm 1.15	85.93 \pm 3.07	79.08 \pm 0.76	79.90 \pm 0.26	81.06 \pm 1.08
KNN	76.97 \pm 0.33	83.56 \pm 0.13	80.26 \pm 0.10	82.26 \pm 1.26	83.60 \pm 2.95	84.04 \pm 0.24	79.60 \pm 0.12	82.38 \pm 1.10
KNN+UM	76.46 \pm 0.35	83.16 \pm 0.12	79.81 \pm 0.23	83.95 \pm 2.38	86.22 \pm 0.73	84.72 \pm 0.24	78.54 \pm 0.18	83.36 \pm 0.84
KNN+ours	77.26 \pm 0.26	83.56 \pm 0.05	80.41 \pm 0.13	82.29 \pm 1.66	85.45 \pm 1.88	84.07 \pm 0.60	79.84 \pm 0.20	82.91 \pm 1.01
ReAct	78.71 \pm 0.27	82.88 \pm 0.15	80.80 \pm 0.11	77.84 \pm 0.27	84.92 \pm 1.95	79.66 \pm 0.84	80.18 \pm 0.61	80.65 \pm 0.72
ReAct+UM	78.00 \pm 0.30	82.31 \pm 0.08	80.15 \pm 0.17	79.07 \pm 1.62	84.19 \pm 0.33	80.67 \pm 0.78	79.14 \pm 0.21	80.77 \pm 0.55
ReAct+ours	78.94 \pm 0.15	82.93 \pm 0.21	80.93 \pm 0.17	78.88 \pm 1.19	86.55 \pm 2.79	80.89 \pm 0.69	80.30 \pm 0.28	81.65 \pm 1.04
Relation	77.75 \pm 0.35	83.66 \pm 0.11	80.71 \pm 0.13	79.94 \pm 0.67	82.81 \pm 2.48	81.29 \pm 0.66	79.73 \pm 0.36	80.94 \pm 0.96
Relation+UM	77.67 \pm 0.04	83.41 \pm 0.12	80.54 \pm 0.05	81.54 \pm 1.29	84.33 \pm 0.61	82.10 \pm 0.41	79.18 \pm 0.05	81.79 \pm 0.50
Relation+ours	78.02 \pm 0.14	83.65 \pm 0.05	80.83 \pm 0.06	80.39 \pm 0.98	85.10 \pm 2.57	81.69 \pm 0.37	79.97 \pm 0.27	81.79 \pm 0.98
Fdbd	78.10 \pm 0.23	84.05 \pm 0.11	81.07 \pm 0.09	78.59 \pm 0.85	81.03 \pm 2.36	81.35 \pm 0.65	79.94 \pm 0.37	80.23 \pm 0.91
Fdbd+UM	77.74 \pm 0.07	83.87 \pm 0.17	80.80 \pm 0.11	81.16 \pm 0.84	82.64 \pm 0.99	81.70 \pm 0.54	79.49 \pm 0.09	81.25 \pm 0.40
Fdbd+ours	78.40 \pm 0.15	84.11 \pm 0.08	81.25 \pm 0.10	79.67 \pm 1.57	83.43 \pm 2.13	81.60 \pm 0.46	80.32 \pm 0.14	81.26 \pm 1.06
Nci	78.34 \pm 0.24	83.65 \pm 0.15	81.00 \pm 0.05	79.81 \pm 0.24	83.21 \pm 1.91	83.87 \pm 0.41	80.95 \pm 0.22	81.96 \pm 0.63
Nci+UM	78.19 \pm 0.19	83.54 \pm 0.18	80.87 \pm 0.18	81.18 \pm 2.16	84.73 \pm 0.09	84.53 \pm 0.32	80.46 \pm 0.11	82.73 \pm 0.55
Nci+ours	78.47 \pm 0.01	83.73 \pm 0.04	81.10 \pm 0.02	79.88 \pm 0.49	84.73 \pm 1.68	84.10 \pm 0.31	81.30 \pm 0.20	82.50 \pm 0.63

A.3 DETAILED RESULTS

A.3.1 DETAILED RESULTS ON CIFAR

Tables 1 and Table 2 present the detailed FPR95 and AUROC results of our method and baseline methods on CIFAR-10. Although performance varies across different OOD datasets, our method demonstrates nearly comprehensive superiority over baselines in terms of average performance. Furthermore, the proposed approach exhibits consistent effectiveness in both Near-OOD and Far-OOD scenarios. When combined with various baselines, our method also shows orthogonality, indicating its compatibility and complementary advantages.

Tables 3 and Table 4 present the detailed FPR95 and AUROC results of our method and baseline methods on CIFAR-100. The proposed approach exhibits consistent effectiveness in both Near-OOD and Far-OOD scenarios. When combined with various baselines, our method also shows orthogonality.

A.3.2 DETAILED RESULTS ON IMAGENET

Table 5 reports the detailed AUROC results of our method and the baseline methods on ImageNet. The results demonstrate that our method achieves near-comprehensive superiority over the baselines.

Furthermore, it exhibits consistent effectiveness in both Near-OOD and Far-OOD scenarios. When combined with different baselines, our method also demonstrates orthogonality, highlighting its compatibility and complementary advantages.

Table 5: Comparison (AUROC \uparrow) on ImageNet benchmark. All values are percentages.

Method	Ssb_hard	Ninco	Near-Avg	iNaturalist	Texture	Openimage-o	Far-Avg
MSP	89.52 \pm 0.08	90.77 \pm 0.04	90.15 \pm 0.04	94.11 \pm 0.13	93.64 \pm 0.12	92.74 \pm 0.06	93.50 \pm 0.07
MSP+UM	89.53 \pm 0.05	90.63 \pm 0.03	90.08 \pm 0.02	94.16 \pm 0.23	93.78 \pm 0.15	92.65 \pm 0.12	93.53 \pm 0.17
MSP+ours	89.85 \pm 0.10	91.02 \pm 0.16	90.43 \pm 0.12	94.44 \pm 0.15	93.91 \pm 0.20	93.00 \pm 0.18	93.78 \pm 0.17
Energy	90.12 \pm 0.08	91.48 \pm 0.12	90.80 \pm 0.06	95.39 \pm 0.25	94.91 \pm 0.12	94.29 \pm 0.05	94.87 \pm 0.07
Energy+UM	89.40 \pm 0.13	90.64 \pm 0.15	90.02 \pm 0.10	94.82 \pm 0.30	94.59 \pm 0.32	93.57 \pm 0.33	94.33 \pm 0.30
Energy+ours	89.40 \pm 0.13	90.64 \pm 0.15	90.02 \pm 0.10	94.82 \pm 0.30	94.59 \pm 0.32	93.57 \pm 0.33	94.33 \pm 0.30
KNN	90.88 \pm 0.09	93.12 \pm 0.13	92.00 \pm 0.10	95.97 \pm 0.11	97.97 \pm 0.08	95.51 \pm 0.12	96.48 \pm 0.09
KNN+UM	90.78 \pm 0.22	92.89 \pm 0.17	91.84 \pm 0.13	95.76 \pm 0.31	98.01 \pm 0.07	95.43 \pm 0.13	96.40 \pm 0.14
KNN+ours	91.28 \pm 0.15	93.22 \pm 0.14	92.25 \pm 0.10	96.21 \pm 0.06	98.03 \pm 0.02	95.71 \pm 0.01	96.65 \pm 0.01
ReAct	90.05 \pm 0.14	91.69 \pm 0.13	90.87 \pm 0.11	95.52 \pm 0.26	95.34 \pm 0.09	94.97 \pm 0.05	95.12 \pm 0.12
ReAct+UM	89.34 \pm 0.21	90.91 \pm 0.14	90.12 \pm 0.10	95.10 \pm 0.30	95.14 \pm 0.23	93.92 \pm 0.33	94.72 \pm 0.27
ReAct+ours	90.19 \pm 0.12	91.72 \pm 0.17	90.95 \pm 0.03	95.55 \pm 0.21	95.48 \pm 0.12	94.40 \pm 0.16	95.14 \pm 0.13
Relation	91.32 \pm 0.06	93.17 \pm 0.05	92.25 \pm 0.05	96.20 \pm 0.09	97.00 \pm 0.08	95.29 \pm 0.06	96.16 \pm 0.08
Relation+UM	91.24 \pm 0.10	92.99 \pm 0.18	92.11 \pm 0.13	96.14 \pm 0.11	96.99 \pm 0.03	95.17 \pm 0.05	96.10 \pm 0.05
Relation+ours	91.50 \pm 0.07	93.21 \pm 0.16	92.36 \pm 0.10	96.35 \pm 0.05	97.02 \pm 0.02	95.37 \pm 0.06	96.25 \pm 0.04
Fdbd	90.90 \pm 0.10	93.00 \pm 0.07	91.95 \pm 0.06	96.16 \pm 0.04	96.89 \pm 0.11	95.20 \pm 0.05	96.08 \pm 0.06
Fdbd+UM	91.04 \pm 0.07	93.11 \pm 0.09	92.07 \pm 0.02	96.38 \pm 0.11	97.00 \pm 0.06	95.36 \pm 0.04	96.25 \pm 0.05
Fdbd+ours	91.23 \pm 0.14	93.31 \pm 0.10	92.27 \pm 0.08	96.51 \pm 0.02	97.08 \pm 0.04	95.49 \pm 0.05	96.36 \pm 0.04
Nci	90.81 \pm 0.03	92.49 \pm 0.08	91.65 \pm 0.03	95.82 \pm 0.03	96.94 \pm 0.09	94.97 \pm 0.05	95.91 \pm 0.05
Nci+UM	90.90 \pm 0.09	92.53 \pm 0.06	91.71 \pm 0.02	96.04 \pm 0.09	97.03 \pm 0.06	95.10 \pm 0.05	96.06 \pm 0.05
Nci+ours	91.14 \pm 0.06	92.72 \pm 0.10	91.93 \pm 0.08	96.16 \pm 0.05	97.06 \pm 0.08	95.21 \pm 0.07	96.14 \pm 0.07

A.3.3 MORE RESULTS ABOUT SENSITIVITY ANALYSIS OF γ .

To further analyze the sensitivity of γ , we report the results for ID (CIFAR-100) vs OOD (SVHN), ID (CIFAR-100) vs OOD (MNIST), ID (CIFAR-100) vs OOD (Texture), and ID (CIFAR-100) vs OOD (Places365) under different γ values (0, 0.02, 0.05, 0.1). The experimental results are shown in Table 6 and Table 7. The experimental results consistently demonstrate a trend where FPR95 initially decreases and then increases as γ rises.

Table 6: Sensitivity analysis (FPR95 \downarrow) of γ on SVHN and MNIST.

Method	CIFAR100 vs SVHN				CIFAR100 vs MNIST			
	$\gamma = 0$	$\gamma = 0.02$	$\gamma = 0.05$	$\gamma = 0.1$	$\gamma = 0$	$\gamma = 0.02$	$\gamma = 0.05$	$\gamma = 0.1$
MSP+UARB	52.46 \pm 7.44	50.32 \pm 7.88	50.11 \pm 3.44	56.93 \pm 1.38	53.70 \pm 3.89	54.85 \pm 1.31	55.85 \pm 1.06	54.14 \pm 1.46
Energy+UARB	45.14 \pm 8.52	43.20 \pm 9.54	44.07 \pm 3.30	52.72 \pm 3.23	52.37 \pm 2.79	51.65 \pm 1.62	53.27 \pm 0.91	50.20 \pm 3.68
KNN+UARB	47.18 \pm 5.76	46.32 \pm 7.30	43.73 \pm 4.77	53.66 \pm 7.49	42.36 \pm 3.11	46.46 \pm 3.99	46.50 \pm 2.47	48.21 \pm 7.26
ReAct+UARB	42.03 \pm 6.60	39.90 \pm 8.02	41.47 \pm 3.33	50.85 \pm 5.38	56.68 \pm 2.88	53.34 \pm 1.78	55.98 \pm 0.43	51.48 \pm 1.48
Relation+UARB	49.48 \pm 7.64	47.37 \pm 8.66	46.07 \pm 5.32	55.24 \pm 5.58	46.30 \pm 3.53	48.41 \pm 3.71	48.94 \pm 1.67	50.33 \pm 6.70
Fdbd+UARB	48.83 \pm 7.57	46.78 \pm 7.62	44.72 \pm 4.66	53.87 \pm 4.67	47.40 \pm 3.41	49.37 \pm 3.07	49.70 \pm 2.26	50.31 \pm 4.46
Nci+UARB	45.55 \pm 5.40	43.53 \pm 5.67	41.90 \pm 4.61	48.91 \pm 2.88	46.79 \pm 2.69	49.71 \pm 2.39	49.55 \pm 2.70	49.39 \pm 5.70

Table 7: Sensitivity analysis (FPR95 \downarrow) of γ on Texture and Places365.

Method	CIFAR100 vs Texture				CIFAR100 vs Places365			
	$\gamma = 0$	$\gamma = 0.02$	$\gamma = 0.05$	$\gamma = 0.1$	$\gamma = 0$	$\gamma = 0.02$	$\gamma = 0.05$	$\gamma = 0.1$
MSP+UARB	62.21 \pm 1.15	60.22 \pm 0.45	59.97 \pm 1.50	58.88 \pm 0.29	56.06 \pm 0.21	55.82 \pm 0.41	55.99 \pm 0.19	56.87 \pm 0.88
Energy+UARB	61.31 \pm 1.62	60.32 \pm 0.69	61.05 \pm 0.59	59.53 \pm 0.64	56.87 \pm 0.65	57.02 \pm 0.67	57.11 \pm 0.57	57.18 \pm 0.95
KNN+UARB	57.27 \pm 1.37	53.40 \pm 2.27	53.17 \pm 3.07	51.97 \pm 1.67	60.49 \pm 1.04	60.02 \pm 0.87	61.14 \pm 1.33	61.57 \pm 1.26
ReAct+UARB	54.44 \pm 4.29	52.87 \pm 0.58	54.89 \pm 0.70	53.85 \pm 0.61	55.83 \pm 0.51	55.26 \pm 0.93	55.54 \pm 0.86	56.25 \pm 0.80
Relation+UARB	58.69 \pm 2.46	54.55 \pm 1.69	53.74 \pm 1.85	53.24 \pm 1.73	61.60 \pm 1.29	60.77 \pm 1.19	61.95 \pm 1.45	62.97 \pm 1.48
Fdbd+UARB	57.14 \pm 2.90	52.84 \pm 1.77	52.39 \pm 0.28	53.17 \pm 1.33	56.46 \pm 0.07	56.26 \pm 0.17	57.13 \pm 0.63	57.76 \pm 1.15
Nci+UARB	49.70 \pm 1.50	47.29 \pm 1.03	46.57 \pm 0.37	46.08 \pm 1.07	53.96 \pm 0.34	53.17 \pm 0.21	54.10 \pm 0.45	54.39 \pm 0.64

A.3.4 FURTHER ANALYSIS ON UARB WITH EMA

To further validate the performance of our method combined with Exponential Moving Average (EMA), we report the results of UARB+EMA integrated with different baselines at various epochs

(100 and 150). The experimental results are presented in Table 8. The experimental results demonstrate that UARB+EMA achieves performance improvements with increasing training epochs, consistently across both Far-OOD and Near-OOD scenarios, and regardless of the baseline method it is combined with.

Table 8: Comparison of OOD detection methods (with our proposed UARB) on large-scale ImageNet benchmark. All values are percentages. EMA denotes Exponential Moving Average.

Method	Epoch	Near-OOD		Far-OOD	
		FPR95↓	AUROC↑	FPR95↓	AUROC↑
Energy+UARB+EMA	100	36.90 ± 1.38	90.68 ± 0.24	19.47 ± 0.84	95.30 ± 0.21
Energy+UARB+EMA	150	32.08 ± 0.59	91.64 ± 0.28	16.57 ± 0.89	96.04 ± 0.10
ReAct+UARB+EMA	100	36.78 ± 0.66	90.68 ± 0.15	18.69 ± 1.05	95.44 ± 0.23
ReAct+UARB+EMA	150	33.92 ± 3.63	91.28 ± 0.81	15.95 ± 0.76	96.25 ± 0.10

REFERENCES

- Guillaume Alain, Alex Lamb, Chinnadhurai Sankar, Aaron Courville, and Yoshua Bengio. Variance reduction in sgd by distributed importance sampling. *arXiv preprint arXiv:1511.06481*, 2015.
- Haw-Shiuan Chang, Erik Learned-Miller, and Andrew McCallum. Active bias: Training more accurate neural networks by emphasizing high variance samples. *Advances in Neural Information Processing Systems*, 30, 2017.
- Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don’t know by virtual outlier synthesis. *arXiv preprint arXiv:2202.01197*, 2022.
- Rundong He, Yue Yuan, Zhongyi Han, Fan Wang, Wan Su, Yilong Yin, Tongliang Liu, and Yongshun Gong. Exploring channel-aware typical features for out-of-distribution detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 12402–12410, 2024.
- Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016.
- Rui Huang, Andrew Geng, and Yixuan Li. On the importance of gradients for detecting distributional shifts in the wild. *Advances in Neural Information Processing Systems*, 34:677–689, 2021.
- Angela H Jiang, Daniel L-K Wong, Giulio Zhou, David G Andersen, Jeffrey Dean, Gregory R Ganger, Gauri Joshi, Michael Kaminsky, Michael Kozuch, Zachary C Lipton, et al. Accelerating deep learning by focusing on the biggest losers. *arXiv preprint arXiv:1910.00762*, 2019.
- Julian Katz-Samuels, Julia B Nakhleh, Robert Nowak, and Yixuan Li. Training ood detectors in their natural habitats. In *ICML*, pp. 10848–10865. PMLR, 2022.
- Justin Khim, Liu Leqi, Adarsh Prasad, and Pradeep Ravikumar. Uniform convergence of rank-weighted learning. In *International conference on machine learning*, pp. 5254–5263. PMLR, 2020.
- Krishnateja Killamsetty, Sivasubramanian Durga, Ganesh Ramakrishnan, Abir De, and Rishabh Iyer. Grad-match: Gradient matching based data subset selection for efficient deep model training. In *International Conference on Machine Learning*, pp. 5464–5474. PMLR, 2021a.
- Krishnateja Killamsetty, Durga Sivasubramanian, Ganesh Ramakrishnan, and Rishabh Iyer. Glisten: Generalization based data subset selection for efficient and robust learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 8110–8118, 2021b.
- Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31, 2018.
- Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.
- Ziqian Lin, Sreya Dutta Roy, and Yixuan Li. Mood: Multi-level out-of-distribution detection. In *CVPR*, pp. 15313–15323, 2021.
- Weitang Liu, Xiaoyun Wang, John D Owens, and Yixuan Li. Energy-based out-of-distribution detection. *arXiv preprint arXiv:2010.03759*, 2020.
- Sören Mindermann, Jan M Brauner, Muhammed T Razzak, Mrinank Sharma, Andreas Kirsch, Winnie Xu, Benedikt Hölten, Aidan N Gomez, Adrien Morisot, Sebastian Farquhar, et al. Prioritized training on points that are learnable, worth learning, and not yet learnt. In *International Conference on Machine Learning*, pp. 15630–15649. PMLR, 2022.
- Yifei Ming, Ying Fan, and Yixuan Li. Poem: Out-of-distribution detection with posterior sampling. In *ICML*, pp. 15650–15665. PMLR, 2022.
- Ziheng Qin, Kai Wang, Zangwei Zheng, Jianyang Gu, Xiangyu Peng, Zhaopan Xu, Daquan Zhou, Lei Shang, Baigui Sun, Xuansong Xie, et al. Infobatch: Lossless training speed up by unbiased dynamic data pruning. *arXiv preprint arXiv:2303.04947*, 2023.

- Chandramouli Shama Sastry and Sageev Oore. Detecting out-of-distribution examples with gram matrices. In *International Conference on Machine Learning*, pp. 8491–8501. PMLR, 2020.
- Tianze Shi, Adrian Benton, Igor Malioutov, and Ozan Irsoy. Diversity-aware batch active learning for dependency parsing. *arXiv preprint arXiv:2104.13936*, 2021.
- Yiyu Sun and Yixuan Li. Dice: Leveraging sparsification for out-of-distribution detection. In *European Conference on Computer Vision*, pp. 691–708. Springer, 2022.
- Yiyu Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. *Advances in Neural Information Processing Systems*, 34, 2021.
- Hongxin Wei, Renchunzi Xie, Hao Cheng, Lei Feng, Bo An, and Yixuan Li. Mitigating neural network overconfidence with logit normalization. *arXiv preprint arXiv:2205.09310*, 2022.
- Xiaobo Xia, Jiale Liu, Shaokun Zhang, Qingyun Wu, Hongxin Wei, and Tongliang Liu. Refined coreset selection: Towards minimal coreset size under model performance constraints. *arXiv preprint arXiv:2311.08675*, 2023.
- Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*, 2021.
- Jinsong Zhang, Qiang Fu, Xu Chen, Lun Du, Zelin Li, Gang Wang, Xiaoguang Liu, Shi Han, and Dongmei Zhang. Out-of-distribution detection based on in-distribution data patterns memorization with modern hopfield energy. In *International Conference on Learning Representations (ICLR’23)*, February 2023.
- Yao Zhu, YueFeng Chen, Chuanlong Xie, Xiaodan Li, Rong Zhang, Hui Xue, Xiang Tian, bolun zheng, and Yaowu Chen. Boosting out-of-distribution detection with typical features, 2022.