Long-Tailed Out-of-Distribution Detection via Normalized Outlier Distribution Adaptation

Wenjun Miao , Guansong Pang, Xiao Bai, Jin Zheng

Background

Out-of-Distribution (OOD) Detection:

OOD detection aims at accurately identifying OOD data from different distributions while also accurately classifying in-distribution (ID) data.

Long-tailed Recognition (LTR) :

Long-tailed Recognition means that training ID data is imbalance across different classes, but test ID data is balanced.

Test time adaptation (TTA) :

TTA focuses on utilizing test data during the inference stage to improve the model performance, without accessing the training data.

Background

OOD detection in LTR is to learn a classifier that for any test data:

- if it drawn from ID data (from either head or tail classes), then this classifier can classify it into the correct ID class;
- if it is drawn from OOD data, then this classifier can detect it as OOD data.

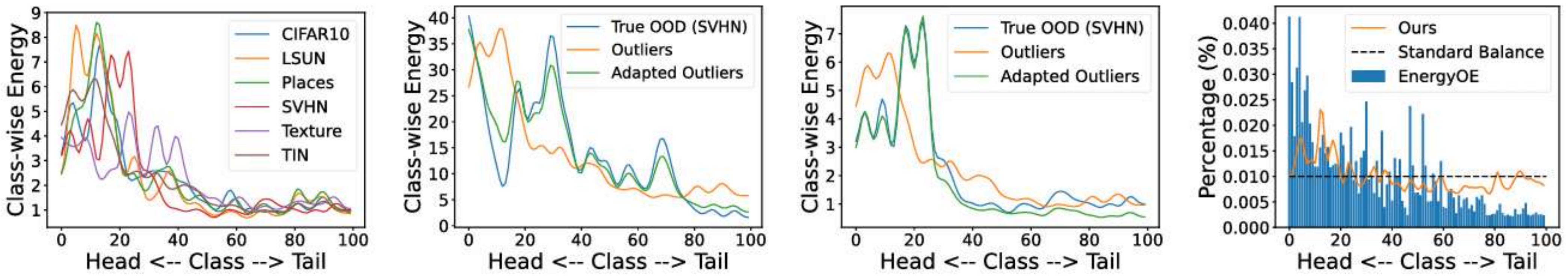
TTA for OOD detection in LTR is to online update the above pre-trained classifier with test data during the inference stage:

- For any unlabeled single test sample, utilizing pre-trained classifier to predict whether it belongs to ID or OOD data at the current iteration, then using the predicted label and the test sample to update the classifier.
- At the next iteration, the updated classifier is used to identify a new test sample and continuously update the classifier. Notably, each sample can only be seen by the classifier once during inference.

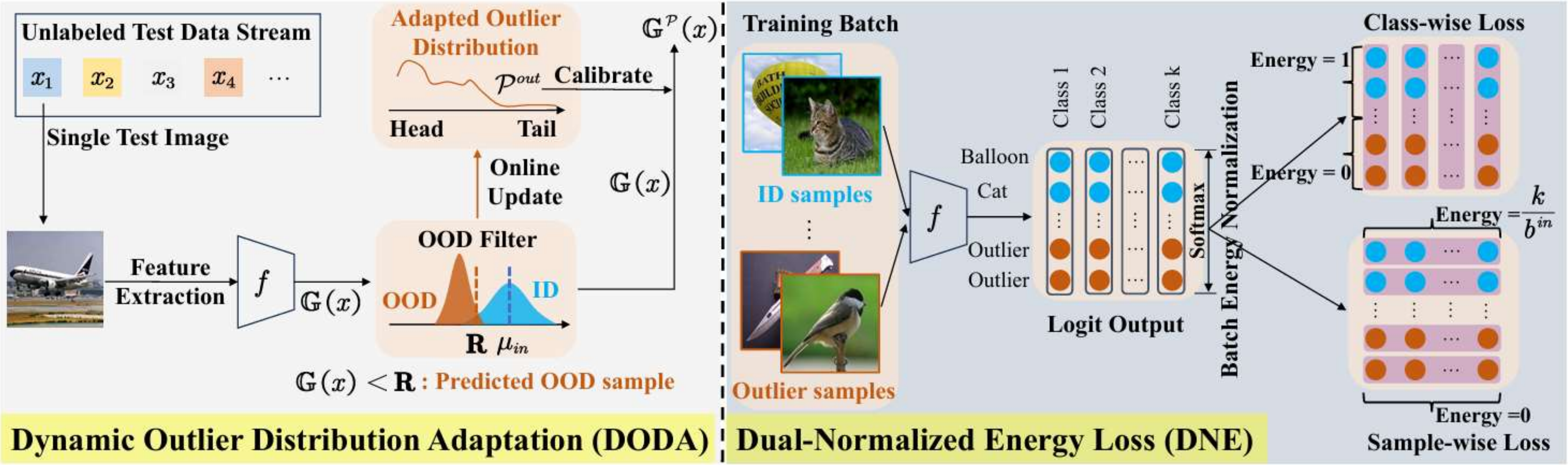
Challenges

One notorious challenge in OOD detection is the lack of ground-truth information on OOD samples, as they can be drawn from any unknown distribution. One popular solution to tackle this challenge is to use samples from external datasets as outliers to train OOD detectors, assuming that the distribution of the outliers is well aligned with that of the true OOD samples in the target data.

- However, the outliers often present a distribution shift compared to the true OOD samples, especially in LTR scenarios. The true OOD samples exhibit very different probability distribution to the head and tailed ID classes from the outliers.



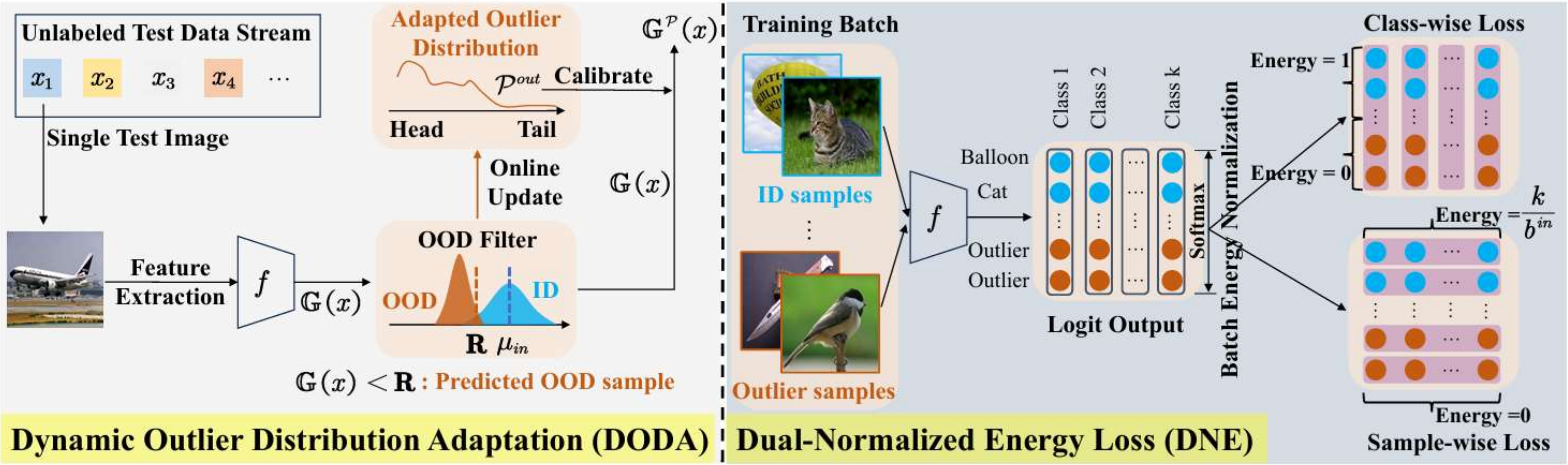
Our Approach: AdaptOD



We introduce a novel AdaptOD approach to tackle the distribution shift problem.

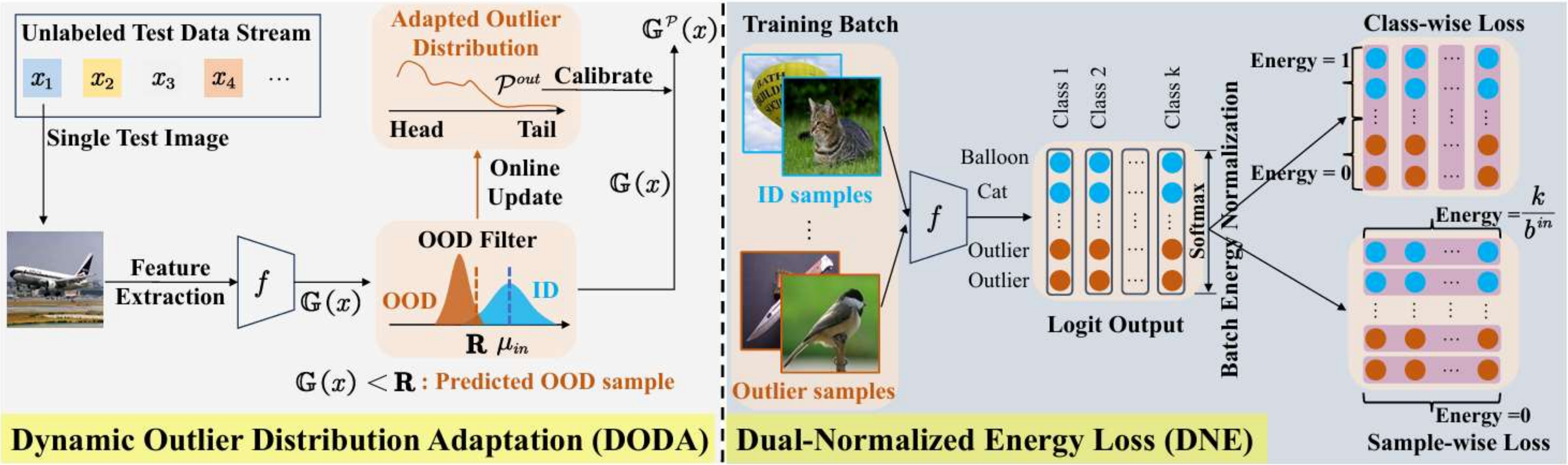
AdaptOD consists of two components, namely Dynamic Outlier Distribution Adaptation (DODA) and Dual-Normalized Energy Loss (DNE).

Our Approach: AdaptOD



DODA builds upon a vanilla outlier distribution and performs test time adaptation to dynamically adapt the this distribution to the true OOD distribution with the OOD knowledge obtained at testing time, and further use it to calibrate the prediction output of test samples at the inference stage. By contrast, DODA focuses on the calibration of the outlier distribution, effectively eliminating the retraining or memory overheads.

Our Approach: AdaptOD



DNE is designed to perform class- and sample-wise normalized energy training, which enforces more balanced prediction energy for imbalanced ID samples, enabling the learning of largely enhanced vanilla outlier distribution for more effective DODA. This guarantees a better starting point for the outlier distribution adaptation and the accuracy of the predicted OOD samples at testing time in DODA, and thus yielding substantially better aligned outlier distribution. It also provides stable energy margins, eliminating the need of manual tuning of these margins.

Experience

Table 1: Comparison of AdaptOD with EnergyOE and COCL on six OOD datasets.

OOD Dataset	Method	ID Dataset: CIFAR10-LT				ID Dataset: CIFAR100-LT			
		AUC↑	AP-in↑	AP-out↑	FPR↓	AUC↑	AP-in↑	AP-out↑	FPR↓
Texture [5]	EnergyOE [24]	95.53	97.42	92.93	18.44	79.56	86.03	70.88	79.45
	COCL [30]	96.81	98.21	93.86	14.65	81.99	88.05	74.38	59.79
	AdaptOD(Ours)	**98.22**	**98.81**	**94.91**	**11.60**	**83.88**	**89.43**	**76.47**	**58.47**
SVHN [31]	EnergyOE [24]	96.63	92.33	98.46	14.37	86.19	81.42	91.74	34.36
	COCL [30]	96.98	93.25	98.61	12.59	89.20	81.57	94.21	54.46
	AdaptOD(Ours)	**98.13**	**94.34**	**99.11**	**10.33**	**93.09**	**91.32**	**96.86**	**17.63**
CIFAR [19]	EnergyOE [24]	84.44	85.74	84.63	61.73	61.15	67.12	56.66	91.42
	COCL [30]	86.63	86.66	86.28	52.21	62.05	66.14	56.82	93.88
	AdaptOD(Ours)	**89.05**	**89.93**	**88.22**	**45.51**	**72.77**	**76.37**	**70.58**	**86.04**
TIN [20]	EnergyOE [24]	88.40	91.65	84.95	46.23	70.78	79.40	55.90	90.74
	COCL [30]	90.43	92.52	87.03	46.12	71.87	81.89	57.12	83.93
	AdaptOD(Ours)	**91.40**	**93.85**	**88.18**	**42.77**	**72.87**	**82.06**	**58.92**	**88.24**
LSUN [51]	EnergyOE [24]	94.00	94.78	93.70	28.42	81.61	86.57	69.16	80.57
	COCL [30]	94.85	95.43	93.98	27.48	84.10	89.89	69.80	74.67
	AdaptOD(Ours)	**96.16**	**96.84**	**95.86**	**24.12**	**85.70**	**90.55**	**72.70**	**70.20**
Place365 [58]	EnergyOE [24]	92.51	84.26	97.14	33.63	79.12	63.38	89.09	81.43
	COCL [30]	93.97	87.36	97.56	32.25	80.30	68.65	89.16	77.83
	AdaptOD(Ours)	**95.19**	**89.56**	**98.44**	**29.22**	**83.27**	**68.82**	**91.44**	**71.63**

Table 2: Comparison to different long-tailed OOD detection methods.

Method	ID Dataset: CIFAR10-LT					ID Dataset: CIFAR100-LT				
	AUC↑	AP-in↑	AP-out↑	FPR↓	ACC↑	AUC↑	AP-in↑	AP-out↑	FPR↓	ACC↑
OE [12]	89.76	89.45	87.22	53.19	73.59	73.52	75.06	67.27	86.30	39.42
EnergyOE [24]	91.92	91.03	91.97	33.80	74.57	76.40	77.32	72.24	76.33	41.32
PASCL [40]	90.99	90.56	89.24	42.90	77.08	73.32	74.84	67.18	79.38	43.10
EAT [45]	92.87	91.76	92.40	32.42	81.31	75.45	76.02	70.87	77.83	46.23
Class Prior [17]	92.08	91.17	90.86	34.42	74.33	76.03	77.31	72.26	76.43	40.77
BERL [4]	92.56	91.41	91.94	32.83	81.37	77.75	78.61	73.10	74.86	45.88
COCL [30]	93.28	92.24	92.89	30.88	81.56	78.25	79.37	73.58	74.09	46.41
OE [12]+**DODA(Ours)**	91.62	90.55	89.39	49.02	73.59	75.46	77.14	69.88	83.67	39.42
EnergyOE [24]+**DODA(Ours)**	93.36	92.17	92.97	30.82	74.57	79.40	80.89	76.54	72.63	41.32
BERL [4]+**DODA(Ours)**	93.77	92.62	93.15	29.41	81.37	79.45	81.15	75.52	70.51	45.88
COCL [30]+**DODA(Ours)**	93.89	93.06	93.39	29.25	81.56	79.81	81.26	75.93	70.14	46.41
AdaptOD(Ours)	**94.69**	**93.89**	**94.12**	**27.26**	**82.27**	**81.93**	**83.09**	**77.83**	**67.37**	**47.91**

Table 3: Comparison to different TTA-based OOD detection methods.

Training Method	TTA Method	ID Dataset: CIFAR10-LT				ID Dataset: CIFAR100-LT			
		AUC↑	AP-in↑	AP-out↑	FPR↓	AUC↑	AP-in↑	AP-out↑	FPR↓
OE [12]	w/o TTA	$89.76_{\pm0.27}$	$89.45_{\pm0.56}$	$87.22_{\pm0.61}$	$53.19_{\pm0.42}$	$73.52_{\pm0.68}$	$75.06_{\pm0.59}$	$67.27_{\pm0.57}$	$86.30_{\pm0.92}$
	AUTO [49]	$90.49_{\pm0.29}$	$89.83_{\pm0.52}$	$87.45_{\pm0.83}$	$52.63_{\pm0.47}$	$73.93_{\pm0.89}$	$75.98_{\pm0.81}$	$67.74_{\pm0.65}$	$85.71_{\pm1.00}$
	AdaODD [56]	$90.89_{\pm0.26}$	$90.17_{\pm0.51}$	$87.88_{\pm0.84}$	$51.44_{\pm0.56}$	$74.67_{\pm0.92}$	$76.53_{\pm0.64}$	$67.89_{\pm0.82}$	$85.34_{\pm0.94}$
	DODA(Ours)	$91.62_{\pm0.23}$	$90.55_{\pm0.45}$	$89.39_{\pm0.68}$	$49.02_{\pm0.41}$	$75.46_{\pm0.77}$	$77.14_{\pm0.59}$	$69.88_{\pm0.80}$	$83.67_{\pm0.88}$
EnergyOE [24]	w/o TTA	$91.92_{\pm0.30}$	$91.03_{\pm0.53}$	$91.97_{\pm0.62}$	$33.80_{\pm0.56}$	$76.40_{\pm0.86}$	$77.32_{\pm0.59}$	$72.24_{\pm0.62}$	$76.33_{\pm1.03}$
	AUTO [49]	$92.48_{\pm0.32}$	$91.43_{\pm0.55}$	$92.44_{\pm0.79}$	$31.99_{\pm0.36}$	$77.65_{\pm1.01}$	$78.11_{\pm0.62}$	$74.18_{\pm0.78}$	$74.66_{\pm0.99}$
	AdaODD [56]	$92.28_{\pm0.26}$	$91.63_{\pm0.56}$	$91.73_{\pm0.61}$	$32.83_{\pm0.59}$	$77.67_{\pm0.82}$	$78.47_{\pm0.81}$	$74.05_{\pm0.83}$	$74.86_{\pm0.98}$
	DODA(Ours)	$93.36_{\pm0.28}$	$92.17_{\pm0.53}$	$92.97_{\pm0.70}$	$30.82_{\pm0.51}$	$79.40_{\pm0.98}$	$80.89_{\pm0.84}$	$76.54_{\pm0.64}$	$72.63_{\pm0.94}$
BERL [4]	w/o TTA	$92.56_{\pm0.40}$	$91.41_{\pm0.83}$	$91.94_{\pm0.85}$	$32.83_{\pm0.38}$	$77.75_{\pm0.77}$	$78.61_{\pm0.56}$	$73.10_{\pm0.73}$	$74.86_{\pm1.07}$
	AUTO [49]	$92.41_{\pm0.49}$	$91.73_{\pm0.56}$	$92.42_{\pm0.90}$	$31.91_{\pm0.36}$	$77.99_{\pm0.75}$	$78.50_{\pm0.84}$	$73.50_{\pm0.87}$	$74.03_{\pm1.00}$
	AdaODD [56]	$92.68_{\pm0.26}$	$91.79_{\pm0.54}$	$92.20_{\pm0.67}$	$31.41_{\pm0.51}$	$78.26_{\pm0.97}$	$78.94_{\pm0.81}$	$73.61_{\pm0.75}$	$73.76_{\pm1.12}$
	DODA(Ours)	$93.77_{\pm0.30}$	$92.62_{\pm0.51}$	$93.15_{\pm0.73}$	$29.41_{\pm0.37}$	$79.45_{\pm0.83}$	$81.15_{\pm0.79}$	$75.52_{\pm0.69}$	$70.51_{\pm0.91}$
COCL [30]	w/o TTA	$93.28_{\pm0.30}$	$92.24_{\pm0.78}$	$92.89_{\pm0.72}$	$30.88_{\pm0.63}$	$78.25_{\pm0.99}$	$79.37_{\pm0.65}$	$73.58_{\pm0.76}$	$74.09_{\pm0.85}$
	AUTO [49]	$93.62_{\pm0.43}$	$92.74_{\pm0.83}$	$93.10_{\pm0.59}$	$30.41_{\pm0.40}$	$78.85_{\pm0.97}$	$79.99_{\pm0.72}$	$74.01_{\pm0.86}$	$72.75_{\pm0.95}$
	AdaODD [56]	$93.48_{\pm0.22}$	$92.60_{\pm0.66}$	$93.05_{\pm0.81}$	$30.79_{\pm0.39}$	$79.07_{\pm0.70}$	$80.00_{\pm0.60}$	$74.60_{\pm0.84}$	$73.09_{\pm0.91}$
	DODA(Ours)	$93.89_{\pm0.36}$	$93.06_{\pm0.56}$	$93.39_{\pm0.74}$	$29.25_{\pm0.40}$	$79.81_{\pm0.96}$	$81.26_{\pm0.59}$	$75.93_{\pm0.72}$	$70.14_{\pm0.98}$
DNE (Ours)	w/o TTA	$92.77_{\pm0.48}$	$92.18_{\pm0.71}$	$92.62_{\pm0.61}$	$31.48_{\pm0.36}$	$77.92_{\pm0.75}$	$78.97_{\pm0.61}$	$73.92_{\pm0.81}$	$74.44_{\pm0.99}$
	AUTO [49]	$92.89_{\pm0.44}$	$92.69_{\pm0.86}$	$92.25_{\pm0.60}$	$30.85_{\pm0.62}$	$79.36_{\pm0.91}$	$80.19_{\pm0.63}$	$74.81_{\pm0.80}$	$72.10_{\pm1.19}$
	AdaODD [56]	$93.39_{\pm0.46}$	$92.27_{\pm0.69}$	$92.92_{\pm0.59}$	$30.78_{\pm0.55}$	$80.26_{\pm0.81}$	$81.72_{\pm0.68}$	$75.62_{\pm0.88}$	$71.96_{\pm0.95}$
	DODA(Ours)	$\mathbf{94.69}_{\pm0.22}$	$\mathbf{93.89}_{\pm0.68}$	$\mathbf{94.12}_{\pm0.58}$	$\mathbf{27.26}_{\pm0.49}$	$\mathbf{81.93}_{\pm0.71}$	$\mathbf{83.09}_{\pm0.64}$	$\mathbf{77.83}_{\pm0.76}$	$\mathbf{67.37}_{\pm0.93}$

Ablation and Quantative Analysis

Table 5: Ablation study results on CIFAR10-LT, CIFAR100-LT and ImageNet-LT.

DODA	DNE-C	DNE-S	ID Dataset: CIFAR10-LT				ID Dataset: CIFAR100-LT				ID Dataset: ImageNet-LT			
			AUC↑	AP-in↑	AP-out↑	FPR↓	AUC↑	AP-in↑	AP-out↑	FPR↓	AUC↑	AP-in↑	AP-out↑	FPR↓
Baseline (EnergyOE [24])			91.92	91.03	91.97	33.80	76.40	77.32	72.24	76.33	69.43	45.12	84.75	76.89
✗	✗	✗	80.33	81.46	77.02	78.71	67.42	68.29	63.86	85.44	58.33	38.40	77.61	89.73
✓	✗	✗	92.63	92.05	92.46	30.17	78.10	80.22	74.17	71.65	71.71	45.99	86.37	74.31
✗	✓	✗	92.12	91.54	92.33	31.85	76.89	77.94	72.76	74.97	71.11	45.59	85.77	76.83
✗	✗	✓	91.98	91.36	91.92	32.44	76.53	77.46	72.55	74.62	70.55	45.36	84.95	77.02
✗	✓	✓	92.77	92.18	92.62	31.48	77.92	78.97	73.92	74.44	72.04	46.53	86.06	75.82
✓	✓	✗	93.81	93.32	93.53	28.69	80.07	82.13	75.73	68.64	73.14	47.61	87.19	73.67
✓	✗	✓	93.49	92.98	93.02	29.52	79.76	81.89	75.31	69.19	72.76	47.32	86.83	74.48
✓	✓	✓	**94.69**	**93.89**	**94.12**	**27.26**	**81.93**	**83.09**	**77.83**	**67.37**	**74.32**	**49.02**	**88.63**	**72.91**
Oracle Model			95.33	94.75	94.96	25.02	83.60	85.09	78.85	65.37	75.84	50.20	89.97	70.71

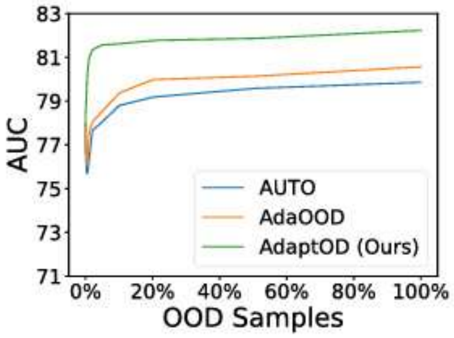


Figure 3: The average performance over six OOD datasets on CIFAR100-LT with an increasing percentage of true OOD samples fed to TTA methods.

Conclusion

- We propose the novel approach AdaptOD to address the distribution shift problem in long-tailed OOD detection, which is the first approach for adapting the outlier distribution to the true OOD distribution from both the training and inference stages.
- AdaptOD utilizes dual-normalized energy loss (DNE) to learn balanced prediction energy on imbalanced ID samples and enhanced vanilla outlier distribution, then uses adynamic outlier distribution adaptation (DODA) to adapt the outlier distribution to the true OOD distribution.
- Experiments on three popular benchmarks demonstrated that AdaptOD significantly enhances the performance of both OOD detection and long-tailed classification.

Thank you