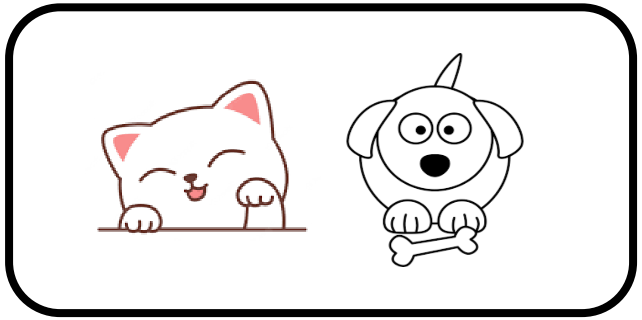
Partial and Asymmetric Contrastive Learning for Out-of-Distribution Detection in Long-Tailed Recognition

Haotao Wang[1] Aston Zhang[2] Yi Zhu[2] Shuai Zheng[2] Mu Li[2] Alex Smola[2] Zhangyang Wang[1]

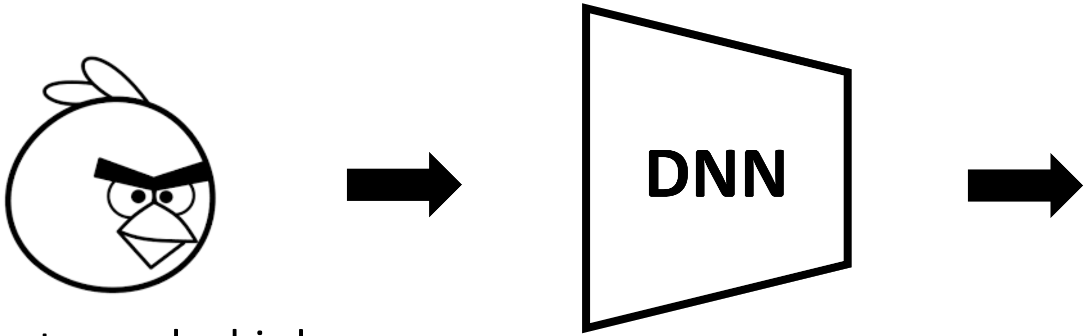
[1]University of Texas at Austin, [2]Amazon Web Services

Out-of-Distribution (OOD) Detection



In-distribution sample

Out-of-distribution sample



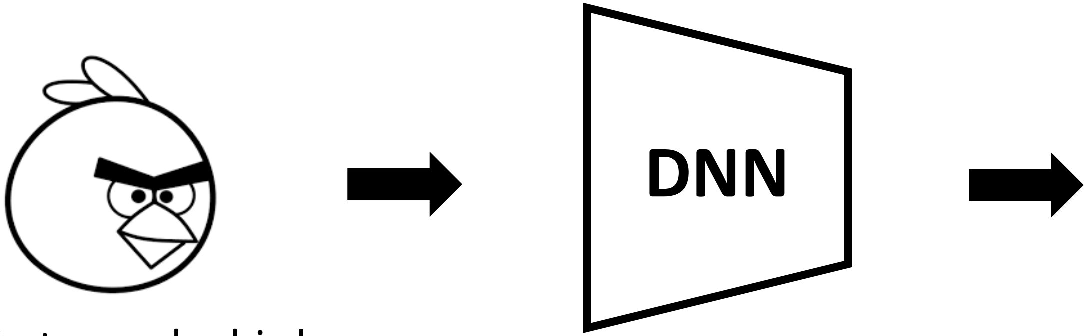
Test sample: bird
(unseen category)

Softmax probability

Cat: 0.95
Dog: 0.05

Over-confident wrong predictions!

Out-of-Distribution (OOD) Detection



Test sample: bird
(unseen category)

Softmax probability

$\boxed{\text{Cat: 0.51}}$ → *Reject low-confident predictions.*

Dog: 0.49

Previous solutions:

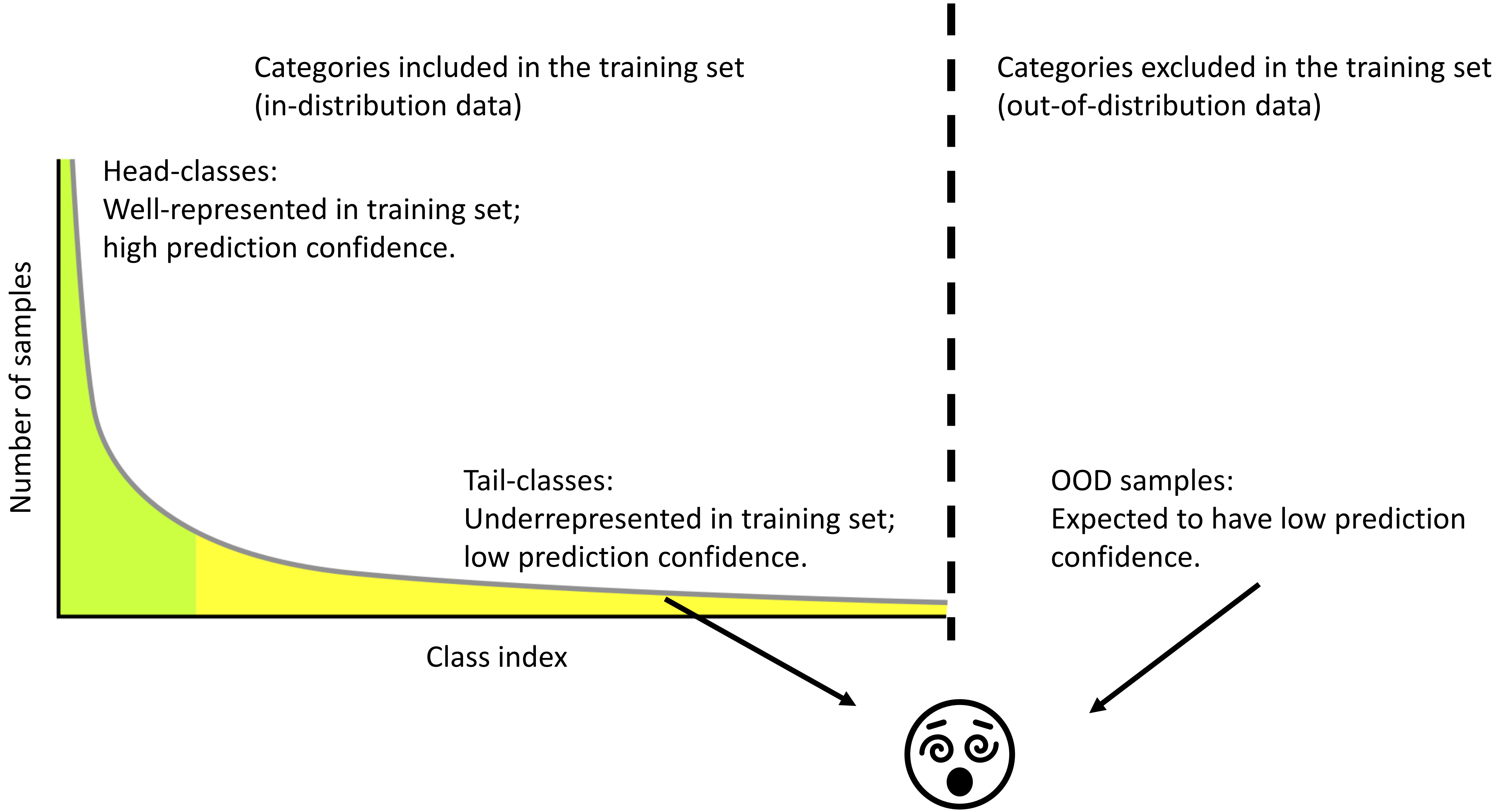
$$\mathcal{L} = \underbrace{\mathbb{E}_{x\sim\mathcal{D}_{\text{in}}}[\mathcal{L}_{\text{in}}(x)]}_{\text{In-distribution classification loss}} + \lambda\underbrace{\mathbb{E}_{x\sim\mathcal{D}_{\text{out}}}[\mathcal{L}_{\text{out}}(x)]}_{\text{Out-of-distribution detection loss}}$$

For example, in Outlier Exposure (OE)[1]:

$$\mathcal{L}_{\text{out}}(x) = \text{KL}(\underbrace{f(x)}_{\text{Softmax probability}} \| \underbrace{u}_{\text{Uniform distribution}})$$

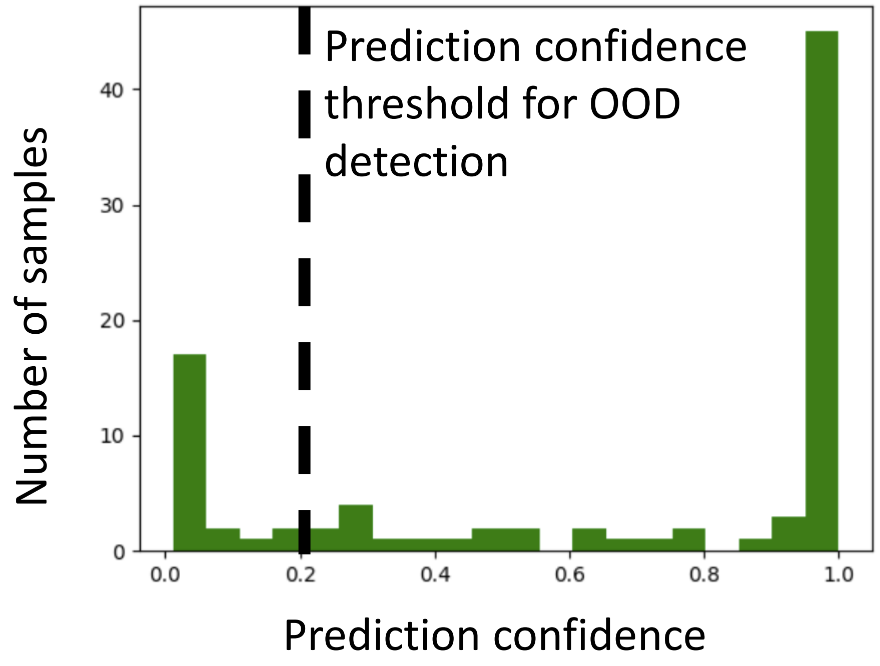
TEXAS The University of Texas at Austin amazon

[1] Deep anomaly detection with outlier exposure. In ICLR, 2019.

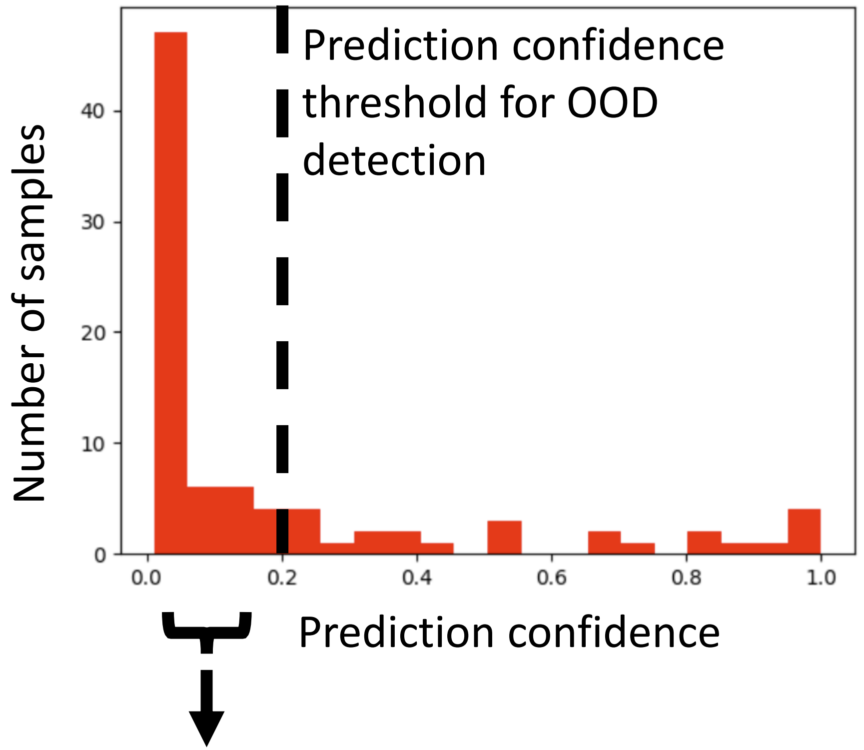


TEXAS
The University of Texas at Austin
amazon

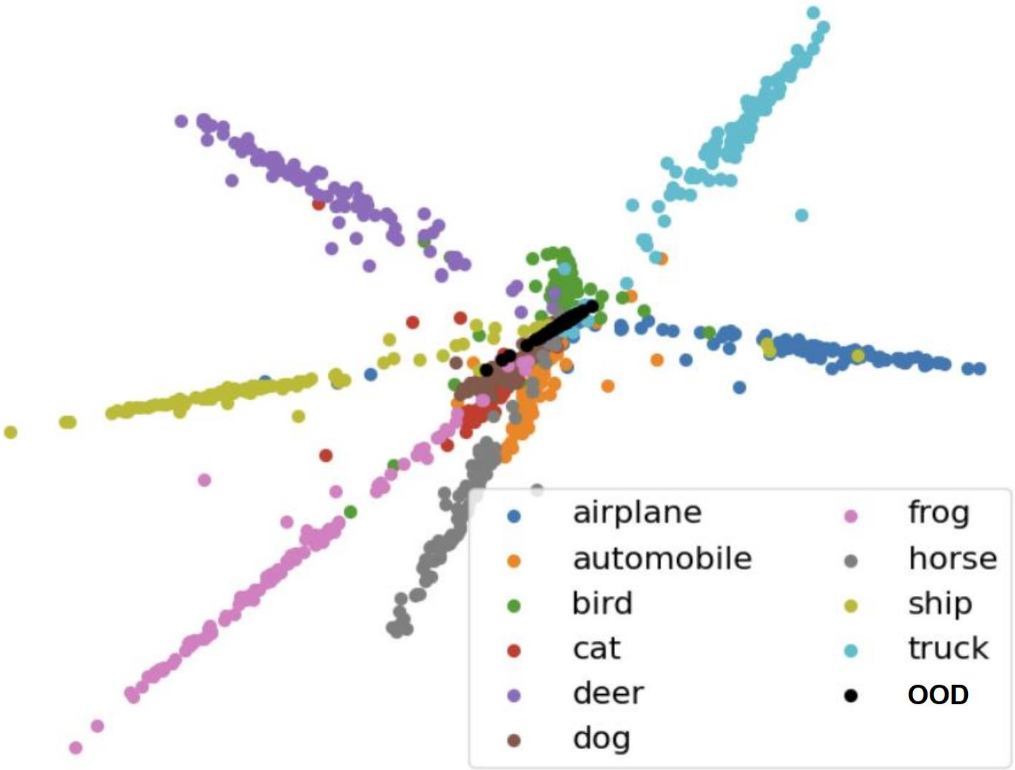
Results on a head-class in CIFAR100-LT.



Results on a tail-class in CIFAR100-LT.



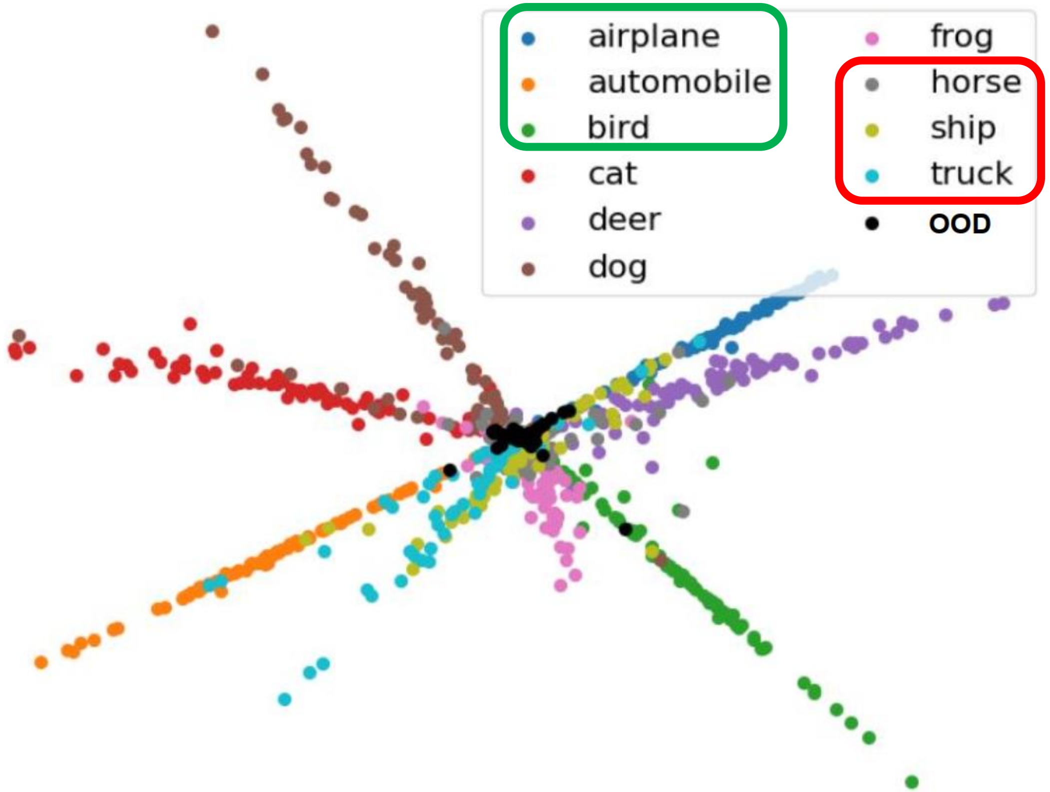
Tail-class samples have lower prediction confidence and tend to be misclassified as OOD samples.



Outlier Exposure (OE) on CIFAR10

Head classes are well-separated from OOD samples.

Tail classes heavily overlap with OOD samples.



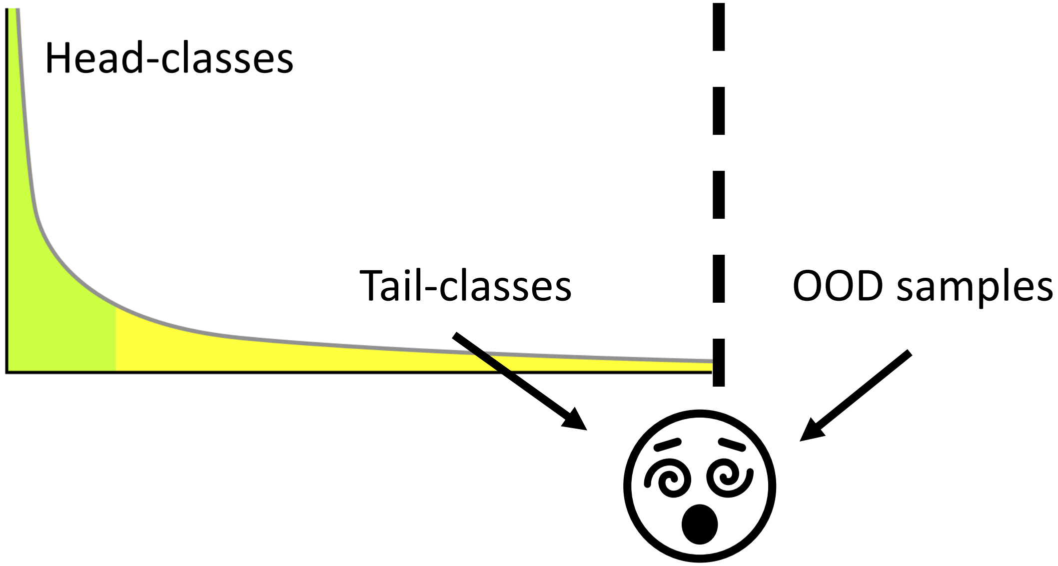
Outlier Exposure (OE) on CIFAR10-LT (long-tailed version of CIFAR10)

TEXAS The University of Texas at Austin amazon

Existing OOD detection methods suffer significant performance drop when trained on long-tailed datasets.

Method	Dataset	AUROC (↑)	AUPR (↑)	FPR95 (↓)	ACC (↑)
NT (MSP)	CIFAR10	85.86	84.37	52.52	93.45
	CIFAR10-LT	72.28 (-13.58)	70.27 (-14.10)	66.07 (+13.55)	72.34 (-21.11)
OE	CIFAR10	96.68	96.29	14.59	92.81
	CIFAR10-LT	89.92 (-6.75)	87.71 (-8.58)	34.80 (+20.21)	73.30 (-19.51)
EnergyOE	CIFAR10	96.59	96.37	14.80	93.07
	CIFAR10-LT	89.31 (-7.27)	88.92 (-7.45)	40.88 (+26.08)	74.68 (-18.39)
SOFL	CIFAR10	96.74	96.60	14.57	89.13
	CIFAR10-LT	91.13 (-5.61)	90.49 (-6.10)	34.98 (+20.41)	54.42 (-34.71)
OECC	CIFAR10	96.27	95.41	14.77	91.95
	CIFAR10-LT	87.28 (-8.99)	86.29 (-9.12)	45.24 (+30.47)	60.16 (-31.79)
NTOM	CIFAR10	96.92	96.95	14.95	91.44
	CIFAR10-LT	92.89 (-4.03)	92.31 (-4.65)	29.03 (+14.09)	66.41 (-25.03)

OOD Detection on Long-Tailed Dataset



The challenge: DNNs get confused between tail-class in-distribution samples and OOD samples.

Naively combining OOD detection methods with long-tail recognition (LTR) methods doesn't bring significant gains.

$\mathcal{D}_{\text{in}}$	**OOD Detection Method**		**LTR Method**	**AUROC (↑)**	**AUPR (↑)**	**FPR95 (↓)**	**ACC (↑)**
CIFAR 10-LT	OE	+	None	<u>89.92</u>	<u>87.71</u>	34.80	73.30
			Re-weighting	89.34	86.39	37.09	70.35
			τ-norm	89.58	85.88	<u>33.80</u>	73.33
			LA	89.46	86.39	34.94	<u>73.93</u>
	Our method			**90.99**±0.19	**89.24**±0.34	**33.36**±0.79	**77.08**±1.01
CIFAR 100-LT	OE	+	None	72.62	<u>66.73</u>	68.69	39.33
			Re-weighting	72.07	66.05	70.62	39.42
			τ-norm	<u>72.71</u>	66.59	<u>68.04</u>	40.87
			LA	72.56	66.48	68.24	<u>42.06</u>
	Our method			**73.32** ± 0.32	**67.18** ± 0.10	**67.44** ± 0.58	**43.10** ± 0.47

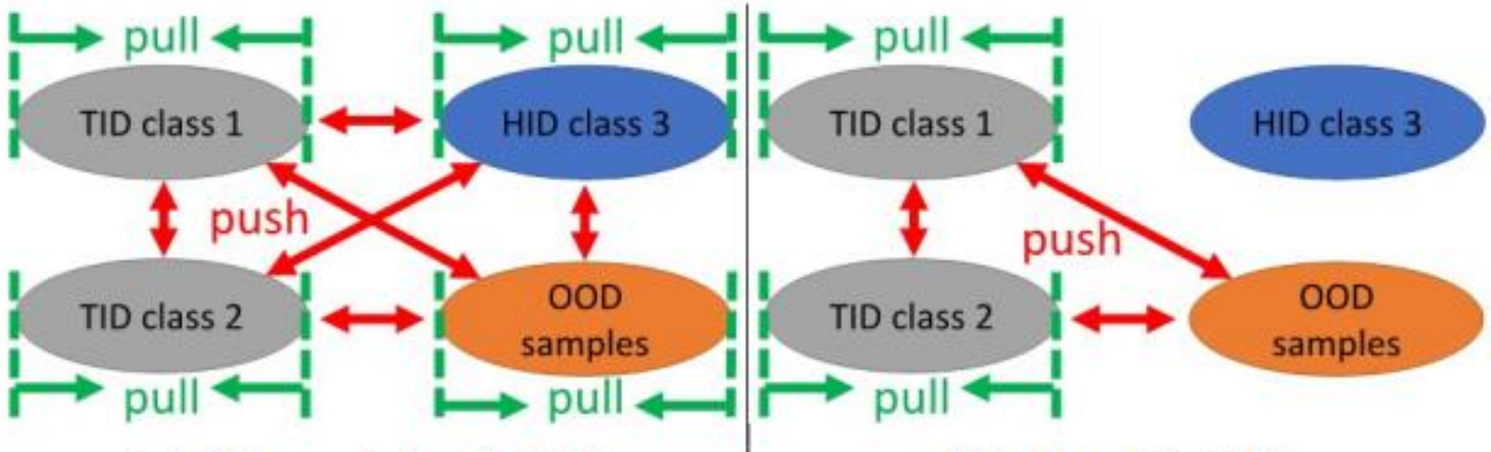
OOD detection methods:
Discourage over-confident predictions on rare samples.

LTR methods:
Encourage confident predictions on rare samples.

Our solution:

1. Explicitly distinguish tail-class in-distribution samples from OOD samples.
2. Disentangle the model to two branches (sharing most parameters): One for OOD detection, and the other for in-distribution classification (LTR).

Partial and Asymmetric Supervised Contrastive Learning (PASCL)

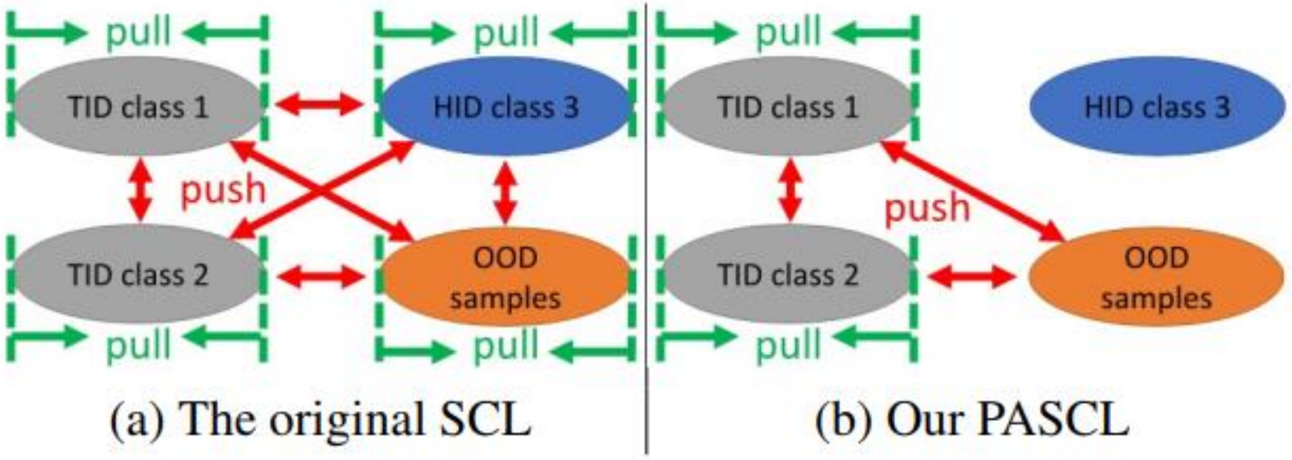


(a) The original SCL [1] (b) Our PASCL

[2] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D. Supervised contrastive learning. In *NeurIPS*, 2020.

Partiality

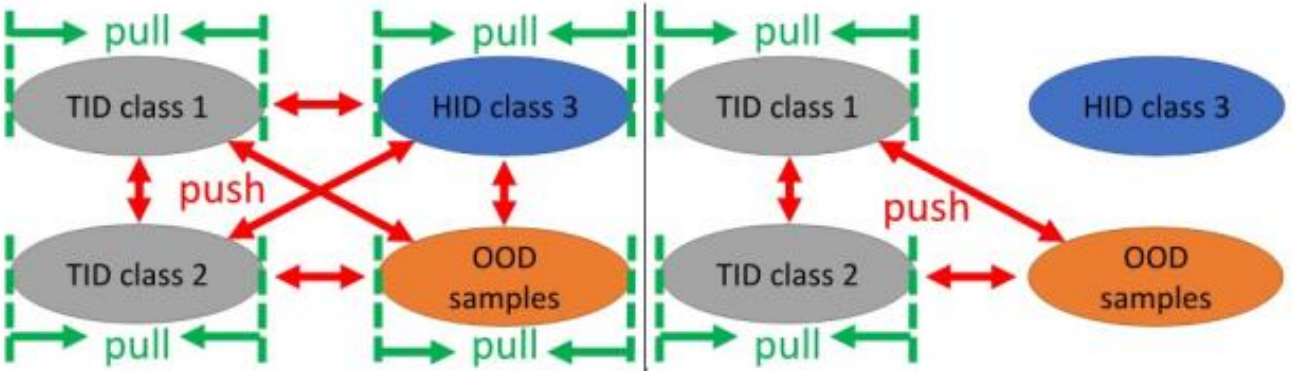
- We apply contrastive learning only on tail-class in-distribution and OOD data, not head-class in-distribution data.
- Intuition: As shown previously, head-class in-distribution samples can be easily separated from the OOD samples, so we do not use an extra contrastive learning loss to explicitly push them away.



TEXAS The University of Texas at Austin amazon

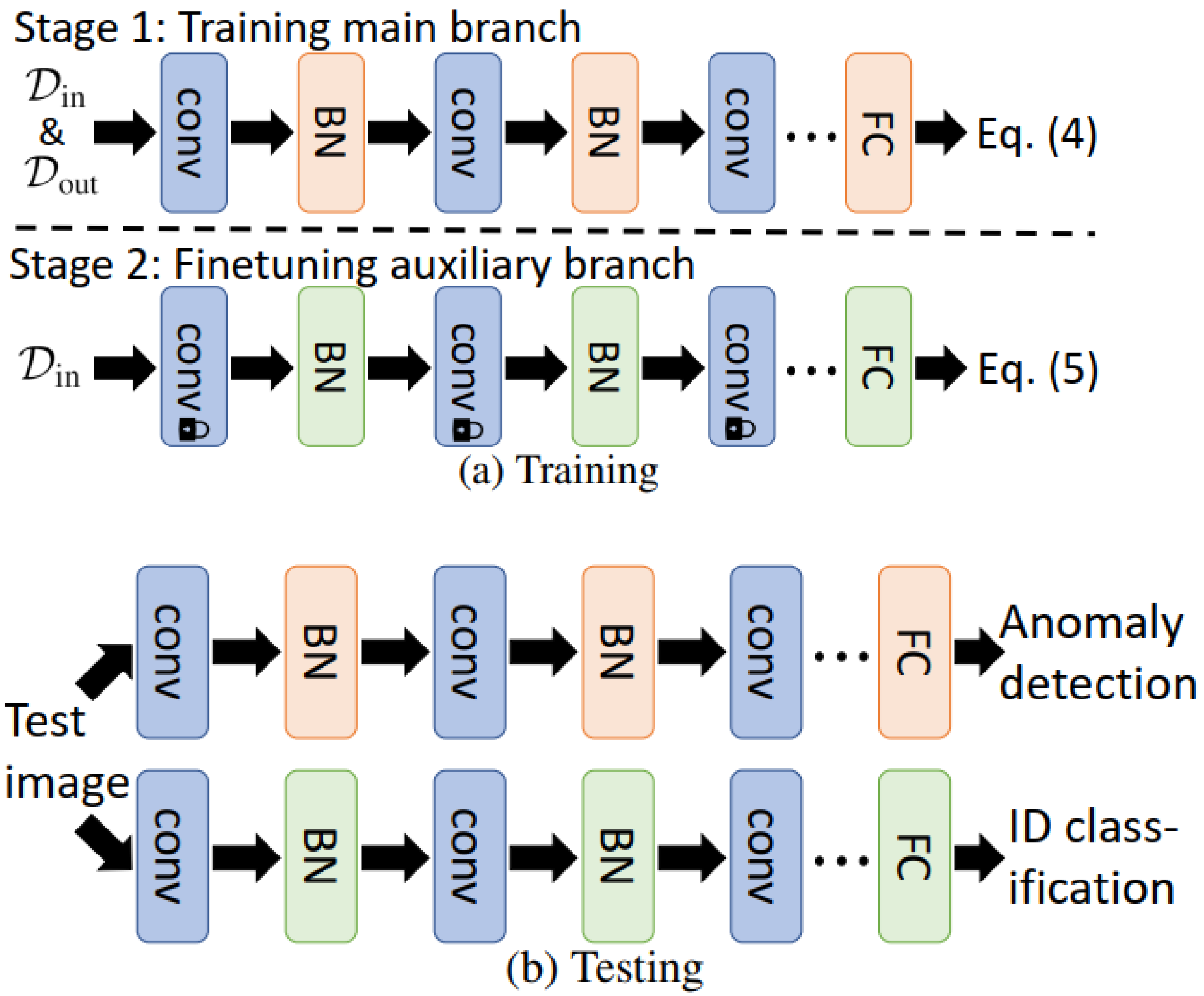
Asymmetry

- We pull in-distribution samples within the same class together, but do not pull OOD samples together.
- Intuition: OOD training set typically has huge diversity in order to be representative for the open visual world, and thus the OOD training samples are not necessarily from the same class.



(a) The original SCL (b) Our PASCL

Auxiliary Branch Finetuning (ABF)

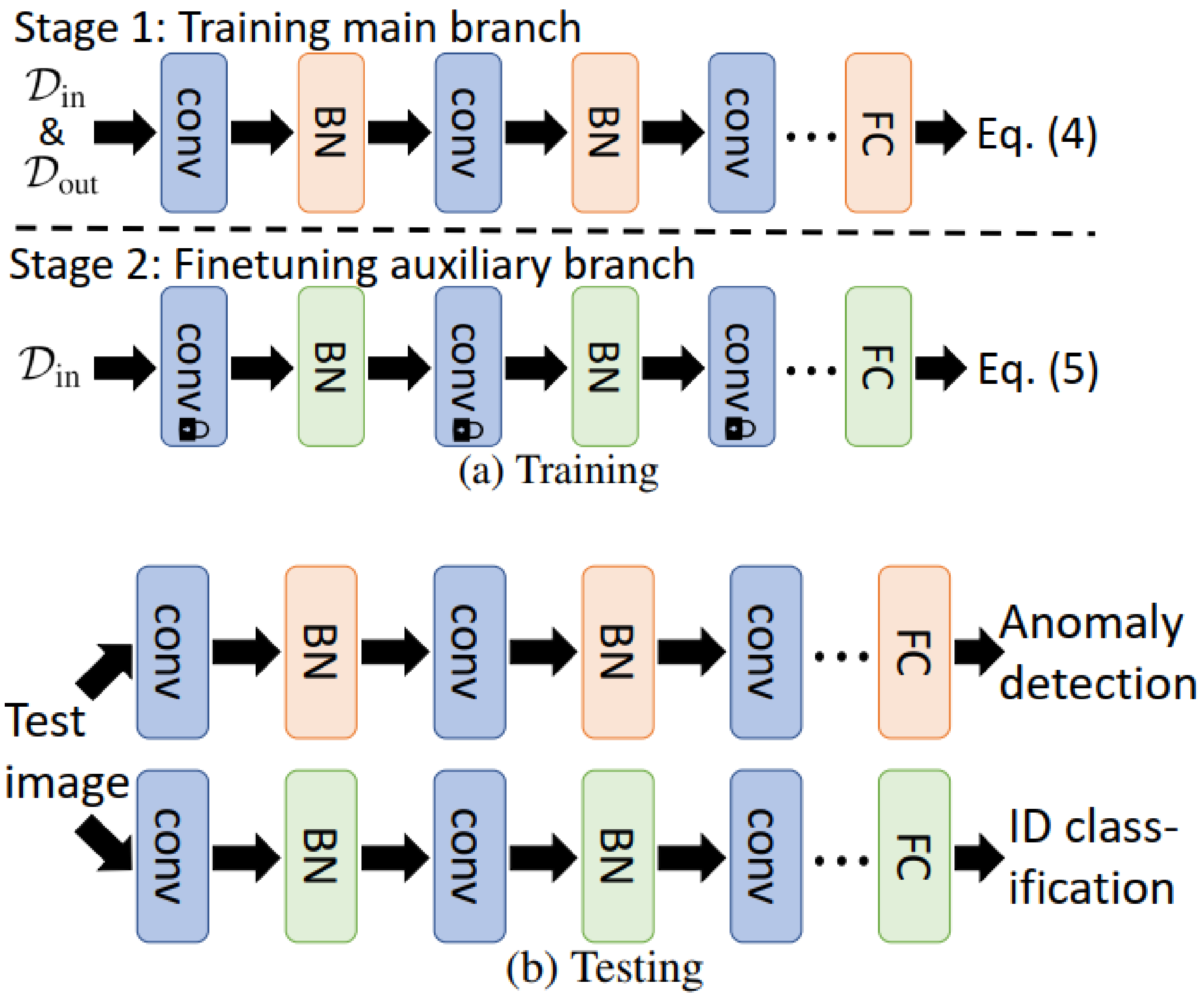


Intuition:

1. In-distribution and out-of-distribution samples have different underlying distributions.
2. It is challenging to achieve anomaly detection and long-tailed in-distribution classification using a shared classification head.

[3] Intriguing properties of adversarial training at scale. In *ICLR*, 2020.

Auxiliary Branch Finetuning (ABF)



(a) Training

(b) Testing

Cross-entropy loss

Reduce prediction confidence on OOD training samples.

$$\mathcal{L} = \mathbb{E}_{x\sim\mathcal{D}_{\text{in}}}[\mathcal{L}_{\text{in}}(x)] + \lambda_1 \mathbb{E}_{x\sim\mathcal{D}_{\text{out}}}[\mathcal{L}_{\text{out}}(x)] + \lambda_2 \mathbb{E}_{x\sim\mathcal{I}}[\mathcal{L}_{\text{c}}(x)], \quad (4)$$

Our PASCL loss function

$$\mathbb{E}_{x\sim\mathcal{D}_{\text{in}}}[\mathcal{L}_{\text{in}}^{LA}(x)]. \quad (5)$$

Stage 2: Finetune BN and FC layers in auxiliary branch using logit-adjustment (LA) cross-entropy loss [2] on in-distribution data.

TEXAS The University of Texas at Austin amazon

[4] Long-tail learning via logit adjustment. In *ICLR*, 2021.

Algorithm 1 Partial and Asymmetric Supervised Contrastive Learning (PASCL)

Input: in-distribution training set $\mathcal{D}_{\text{in}}$, OOD training set $\mathcal{D}_{\text{out}}$, main branch training iteration n_1, auxiliary branch finetuning iteration n_2.

#Stage 1: Train main branch.

for $i = 1$ **to** n_1 **do**

  Sample a batch of in-distribution and OOD training samples.

  Update the main branch model by minimizing Eq. (4)

end for

#Stage 2: Finetune auxiliary branch.

Fix all layers except the auxiliary BN and classification layers in the model.

for $i = 1$ **to** n_2 **do**

  Sample a batch of in-distribution training samples.

  Update the auxiliary BN and classification layers by minimizing Eq. (5).

end for

TEXAS
The University of Texas at Austin
amazon

Results on CIFAR10-LT

(a) OOD detection results and
in-distribution classification results in terms of ACC95.

$\mathcal{D}_{\text{out}}^{\text{test}}$	Method	AUROC (↑)	AUPR (↑)	FPR95 (↓)	ACC95 (↑)
Texture	OE	92.59 ± 0.42	83.32 ± 1.67	25.10 ± 1.08	84.52 ± 0.76
	Ours	**93.16** ± 0.37	**84.80** ± 1.50	**23.26** ± 0.91	**85.86** ± 0.72
SVHN	OE	95.10 ± 1.01	97.14 ± 0.81	16.15 ± 1.52	81.33 ± 0.81
	Ours	**96.63** ± 0.90	**98.06** ± 0.56	**12.18** ± 3.33	**82.72** ± 1.51
CIFAR100	OE	83.40 ± 0.30	80.93 ± 0.57	**56.96** ± 0.91	**94.56** ± 0.57
	Ours	**84.43** ± 0.23	**82.99** ± 0.48	57.27 ± 0.88	94.48 ± 0.31
Tiny ImageNet	OE	86.14 ± 0.29	79.33 ± 0.65	47.78 ± 0.72	91.19 ± 0.33
	Ours	**87.14** ± 0.18	**81.54** ±0.38	**47.69** ± 0.59	**91.20** ± 0.35
LSUN	OE	91.35 ± 0.23	87.62 ± 0.82	27.86 ± 0.68	85.49 ± 0.69
	Ours	**93.17** ± 0.15	**91.76** ± 0.53	**26.40** ± 1.00	**86.67** ± 0.90
Places365	OE	90.07 ± 0.26	95.15 ± 0.24	34.04 ± 0.91	87.07 ± 0.53
	Ours	**91.43** ± 0.17	**96.28** ± 0.14	**33.40** ± 0.88	**87.87** ± 0.71
Average	OE	89.77 ± 0.27	87.25 ± 0.61	34.65 ± 0.46	87.36 ± 0.51
	Ours	**90.99** ± 0.19	**89.24** ± 0.34	**33.36** ± 0.79	**88.13** ± 0.56

(b) In-distribution classification results in terms of ACC@FPRn.

Method	ACC@FPRn (↑)			
	0	0.001	0.01	0.1
OE	73.84 ± 0.77	73.90 ± 0.77	74.46 ± 0.81	78.88 ± 0.66
Ours	**77.08** ± 1.01	**77.13** ± 1.02	**77.64** ± 0.99	**81.96** ± 0.85

(c) Comparison with other methods.

$\mathcal{D}_{\text{out}}^{\text{test}}$	Method	AUROC (↑)	AUPR (↑)	FPR95 (↓)	ACC (↑)
Average	ST (MSP)	72.28	70.27	66.07	72.34
	OECC	87.28	86.29	45.24	60.16
	EnergyOE	89.31	88.92	40.88	74.68
	OE	89.77 ± 0.27	87.25 ± 0.61	34.65 ± 0.46	73.84 ± 0.77
	Ours	**90.99** ± 0.19	**89.24** ± 0.34	**33.36** ± 0.79	**77.08** ± 1.01

TEXAS
The University of Texas at Austin
amazon

Results on CIFAR100-LT

(a) OOD detection results and in-distribution classification results in terms of ACC95.

$\mathcal{D}_{\text{out}}^{\text{test}}$	**Method**	**AUROC (↑)**	**AUPR (↑)**	**FPR95 (↓)**	**ACC95 (↑)**
Texture	OE	**76.71** ± 1.20	**58.79** ± 1.39	68.28 ± 1.53	71.43 ± 1.58
	Ours	76.01 ± 0.66	58.12 ± 1.06	**67.43** ± 1.93	**73.11** ± 1.55
SVHN	OE	77.61 ± 3.26	86.82 ± 2.50	58.04 ± 4.82	64.27 ± 3.26
	Ours	**80.19** ± 2.19	**88.49** ± 1.59	**53.45** ± 3.60	**64.50** ± 1.87
CIFAR10	OE	62.23 ± 0.30	**57.57** ± 0.34	80.64 ± 0.98	**82.67** ± 0.99
	Ours	**62.33** ± 0.38	57.14 ± 0.20	**79.55** ± 0.84	82.30 ± 1.07
Tiny ImageNet	OE	68.04 ± 0.37	**51.66** ± 0.51	76.66 ± 0.47	76.22 ± 0.61
	Ours	**68.20** ± 0.37	51.53 ± 0.42	**76.11** ± 0.80	**77.56** ± 1.15
LSUN	OE	77.10 ± 0.64	**61.42** ± 0.99	63.98 ± 1.38	65.64 ± 1.03
	Ours	**77.19** ± 0.44	61.27 ± 0.72	**63.31** ± 0.87	**68.05** ± 1.24
Places365	OE	75.80 ± 0.45	**86.68** ± 0.38	65.72 ± 0.92	67.04 ± 0.49
	Ours	**76.02** ± 0.21	86.52 ± 0.29	**64.81** ± 0.27	**69.04** ± 0.90
Average	OE	72.91 ± 0.68	67.16 ± 0.57	68.89 ± 1.07	71.21 ± 0.84
	Ours	**73.32** ± 0.32	**67.18** ± 0.10	**67.44** ± 0.58	**72.43** ± 0.66

(b) in-distribution classification results in terms of ACC@FPRn.

Method	**ACC@FPRn (↑)**			
	0	0.001	0.01	0.1
OE	39.04 ± 0.37	39.07 ± 0.38	39.38 ± 0.38	42.40 ± 0.44
Ours	**43.10** ± 0.47	**43.12** ± 0.47	**43.39** ± 0.48	**46.14** ± 0.38

(c) Comparison with other methods.

$\mathcal{D}_{\text{out}}^{\text{test}}$	**Method**	**AUROC (↑)**	**AUPR (↑)**	**FPR95 (↓)**	**ACC (↑)**
Average	ST (MSP)	61.00	57.54	82.01	40.97
	OECC	70.38	66.87	73.15	32.93
	EnergyOE	71.10	67.23	71.78	39.05
	OE	72.91 ± 0.68	67.16 ± 0.57	68.89 ± 1.07	39.04 ± 0.37
	Ours	**73.32** ± 0.32	**67.18** ± 0.10	**67.44** ± 0.58	**43.10** ± 0.47

Results on ImageNet-LT

$\mathcal{D}_{out}^{test}$	Method	AUROC (↑)	AUPR (↑)	FPR@TPRn (↓)				ACC@TPRn (↑)				ACC@FPRn (↑)			
				0.98	0.95	0.90	0.80	0.98	0.95	0.90	0.80	0	0.001	0.01	0.1
ImageNet -1k-OOD	ST (MSP)	53.81	51.63	95.38	90.15	83.52	72.97	**96.67**	**92.61**	**87.43**	**77.52**	39.65	39.68	40.00	43.18
	OECC	63.07	63.05	**93.15**	**86.90**	78.79	65.23	94.25	88.23	80.12	68.36	38.25	38.28	38.56	41.47
	EnergyOE	64.76	64.77	94.15	87.72	78.36	63.71	80.18	74.38	67.65	59.68	38.50	38.52	38.72	40.99
	OE	66.33	68.29	95.11	88.22	78.68	65.28	95.46	88.22	78.68	65.28	37.60	37.62	37.79	40.00
	Ours	**68.00** (+1.67)	**70.15** (+1.86)	94.38 (-0.73)	87.53 (-0.69)	**78.12** (-0.56)	**62.48** (-2.80)	95.69 (+0.23)	89.55 (+1.33)	80.88 (+2.20)	69.60 (+4.32)	**45.49** (+7.89)	**45.51** (+7.89)	**45.62** (+7.83)	**47.49** (+7.49)

Ablation study on each component in PASCL

$\mathcal{D}_{\text{in}}$	Asymmetry	Partiality	ABF	**AUROC** (↑)	**AUPR** (↑)	**FPR95** (↓)	**ACC95** (↑)	**ACC@FPRn** (↑)			
								0	0.001	0.01	0.1
CIFAR10-LT	No contrastive loss (OE)			95.10 ± 1.01	97.14 ± 0.81	16.15 ± 1.52	81.33 ± 0.81	73.84 ± 0.77	73.90 ± 0.77	74.46 ± 0.81	78.88 ± 0.66
	✗	✗	✗	95.34 ± 1.58	97.30 ± 1.20	15.12 ± 3.07	81.94 ± 1.28	75.03 ± 1.46	75.09 ± 1.45	75.60 ± 1.44	80.02 ± 1.10
	✗	✓	✗	95.01 ± 1.25	96.74 ± 0.78	15.31 ± 4.35	82.34 ± 1.56	74.46 ± 1.80	74.52 ± 1.80	75.04 ± 1.76	80.21 ± 0.99
	✓	✗	✗	94.91 ± 1.43	96.86 ± 1.47	15.57 ± 1.19	82.08 ± 0.47	75.24 ± 0.99	75.29 ± 0.98	75.77 ± 0.98	79.85 ± 0.77
	✓	✓	✗	**96.63** ± 0.90	**98.06** ± 0.56	**12.18** ± 3.33	81.70 ± 1.21	76.20 ± 0.79	76.26 ± 0.79	76.85 ± 0.81	81.07 ± 0.58
	✓	✓	✓	**96.63** ± 0.90	**98.06** ± 0.56	**12.18** ± 3.33	**82.72** ± 1.51	**77.08** ± 1.01	**77.13** ± 1.02	**77.64** ± 0.99	**81.96** ± 0.85
CIFAR100-LT	No contrastive loss (OE)			77.61 ± 3.26	86.82 ± 2.50	58.04 ± 4.82	64.27 ± 3.26	39.04 ± 0.37	39.07 ± 0.38	39.38 ± 0.38	42.40 ± 0.44
	✗	✗	✗	78.05 ± 2.12	87.18 ± 0.87	59.10 ± 5.03	**66.44** ± 3.90	40.21 ± 0.43	40.25 ± 0.43	40.56 ± 0.45	43.71 ± 0.42
	✗	✓	✗	79.46 ± 1.83	88.01 ± 1.90	54.59 ± 3.34	63.86 ± 2.52	40.24 ± 0.53	40.28 ± 0.53	40.60 ± 0.55	43.93 ± 0.57
	✓	✗	✗	79.54 ± 2.38	87.68 ± 1.51	54.27 ± 3.69	63.33 ± 2.87	40.00 ± 0.42	40.04 ± 0.41	40.36 ± 0.42	43.60 ± 0.42
	✓	✓	✗	**80.19** ± 2.19	**88.49** ± 1.59	**53.45** ± 3.60	63.10 ± 1.87	40.33 ± 0.20	40.36 ± 0.20	40.66 ± 0.18	43.79 ± 0.22
	✓	✓	✓	**80.19** ± 2.19	**88.49** ± 1.59	**53.45** ± 3.60	64.50 ± 1.87	**43.10** ± 0.47	**43.12** ± 0.47	**43.39** ± 0.48	**46.14** ± 0.38

Thank you!

- Code and pre-trained models are available at https://github.com/amazon-research/long-tailed-ood-detection

TEXAS
The University of Texas at Austin
amazon