Enlarging Instance-specific and Class-specific Information for Open-set Action Recognition

Jun Cen[1,2], Shiwei Zhang[2], Xiang Wang[3], Yixuan Pei[4],
Zhiwu Qing[3], Yingya Zhang[2], Qifeng Chen[1]

[1]The Hong Kong University of Science and Technology, [2]Alibaba Group
[3]Huazhong University of Science and Technology, [4]Xi'an Jiaotong University

WED-PM-280

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Problem Description

InD samples

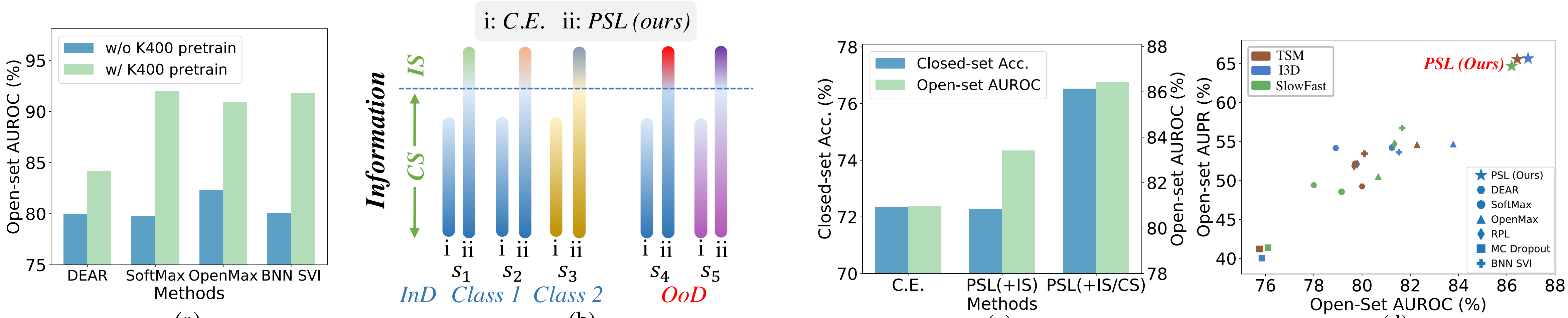
Samples whose classes are in the training set

OoD samples

Samples whose classes are not in the training set

- During the inference stage, not all samples are in the classes of the training set. These samples are called Out-of-Distribution (OoD) samples. In contrast, samples whose classes are witnessed during training are called In-Distribution (InD) samples.
- Detecting OoD actions while correctly classifying the InD actions are called Open-set Action Recognition (OSAR).

Summary



(a) (b) (c) (d)

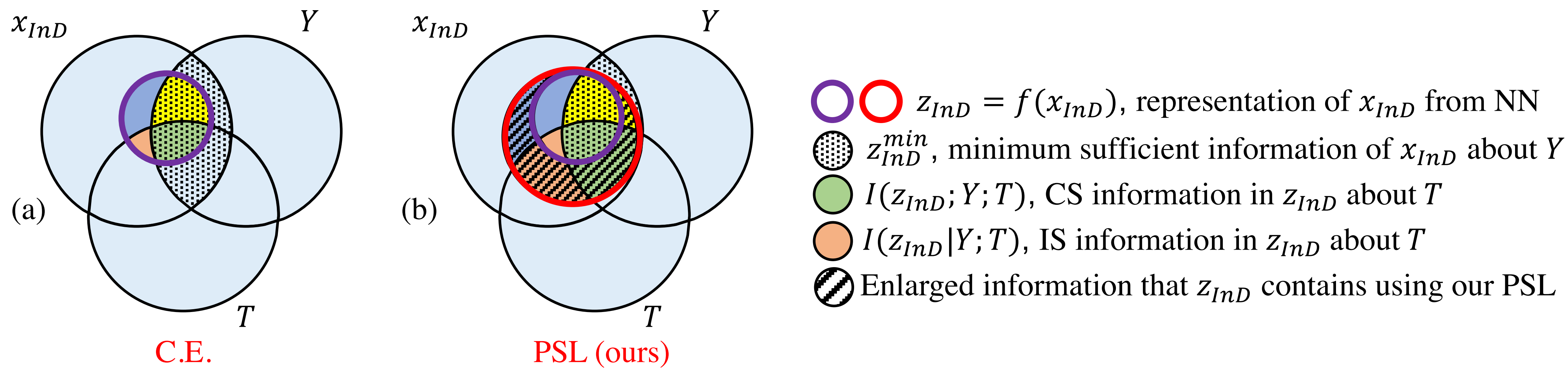
(a) Richer semantic features brought by the pretraining can significantly improve the open-set performance.

(b) Information in the feature is divided into IS and CS information. s4 can be identified as OoD since it has distinct IS information (IS bars in different colors) with s1 and s2, while s5 has distinct CS information (CS bars in different colors) with all InD samples so it may be OoD. Our PSL aims to learn more IS and CS information (bars in longer lengths) than Cross-Entropy (C.E.).

(c) Both enlarged IS and CS information boosts the open-set performance.

(d) Our PSL achieves the best OSAR performance.

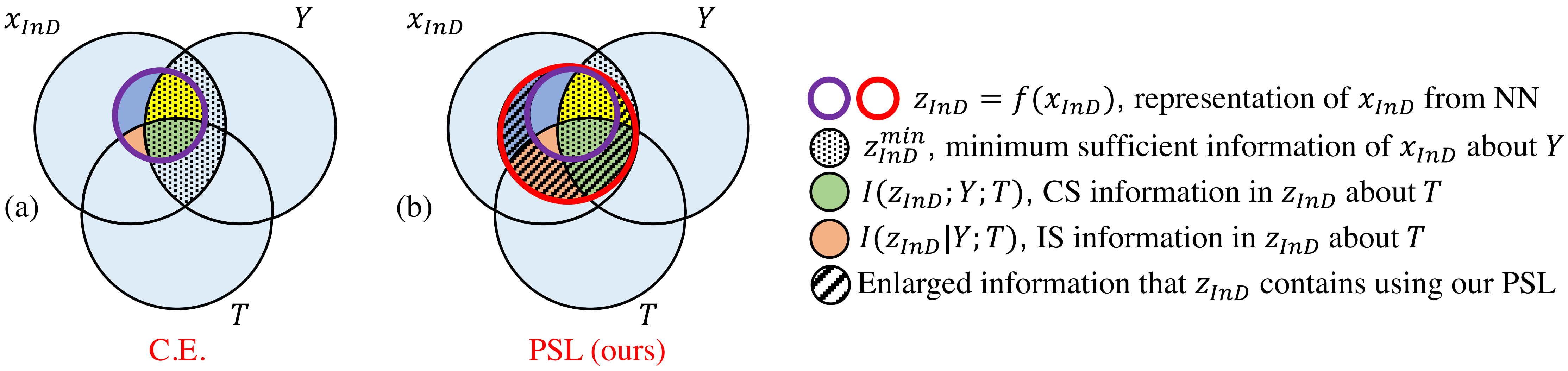
Information Analysis in OSAR



$$I(x_{InD}; z_{InD}) = \underbrace{I(x_{InD}; z_{InD}|Y)}_{IS} + \underbrace{I(z_{InD}; Y)}_{CS}, \qquad I(z_{InD}; T) = \underbrace{I(z_{InD}|Y; T)}_{IS \text{ about } T} + \underbrace{I(z_{InD}; Y; T)}_{CS \text{ about } T},$$

- The information is divided into IS and CS information. IS information is not related to the classification task Y while CS information is.
- IS and CS information could be related to the OSAR task T. Therefore, enlarging the IS and CS information is helpful for the OSAR task.

IS and CS Information Behavior under Cross-entropy

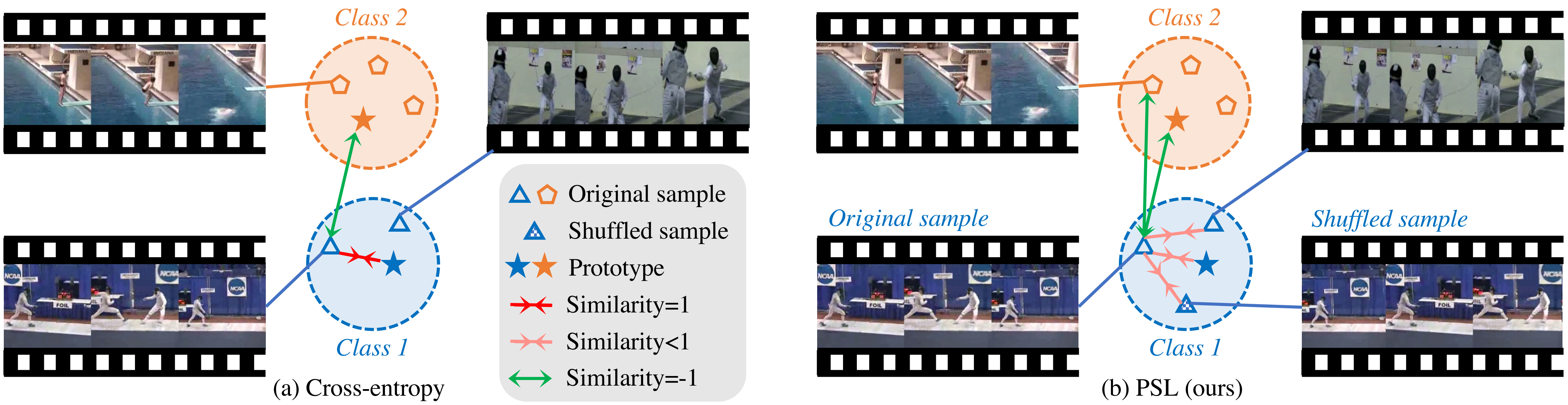


Proposition 1 *For two feature representations of samples in the same class, more CS information means these two feature representations are more similar, and more IS information decreases their feature similarity.*

- CS information is for the closed-set label prediction task Y , which is fully supervised by C.E. loss, so it is maximized during training.

- In contrast, C.E. encourages representations of the same class to be exactly same with the corresponding prototype, and such high similarity eliminates the IS information according to Proposition 1. Therefore, C.E. loss tends to maximize the CS information and eliminate the IS information in the feature representation.

Prototypical Similarity Learning



(a) Cross-entropy

(b) PSL (ours)

- We find that cross-entropy tends to eliminate IS information because it encourages the features of the same class to be exactly same. So we argue that the features of the same class should reach a less than 1 similarity rather than 1 to keep the IS information.

- We also encourage the similarity between the original sample and shuffled sample to be less than 1, since they share the same appearance information but different temporal information. We find this technique enlarges the CS information.

Results

Datasets	Methods	w/o K400 Pretrain				w/ K400 Pretrain			
		AUROC↑	AUPR↑	FPR95↓	Acc.↑	AUROC↑	AUPR↑	FPR95↓	Acc.↑
UCF101 (InD) HMDB51 (OoD)	OpenMax [8]	*82.28*	*54.59*	*50.69*	*73.92*	90.89	73.16	38.77	95.32
	MC Dropout [7]	75.75	41.21	54.78	73.63	88.23	67.62	38.12	95.06
	BNN SVI [24]	80.10	53.43	52.33	71.51	*91.81*	*79.65*	31.43	94.71
	SoftMax [6]	79.72	52.13	53.22	*73.92*	91.75	77.69	*28.60*	95.03
	RPL [27]	79.67	51.85	56.40	71.46	90.53	77.86	37.09	*95.59*
	DEAR [5]	80.00	49.23	53.28	71.33	84.16	75.54	89.40	94.48
	PSL(ours)	**86.43**	**65.54**	**41.67**	**76.53**	**94.05**	**86.55**	**23.18**	**95.62**
	Δ	(+4.15)	(+10.95)	(-9.02)	(+2.61)	(+2.24)	(+6.90)	(-5.42)	(+0.03)
UCF101 (InD) MiTv2 (OoD)	OpenMax [8]	*84.43*	*76.69*	*47.74*	*73.92*	*93.34*	88.14	*28.95*	95.32
	MC Dropout [7]	75.66	62.20	51.57	73.63	88.71	83.36	39.46	95.06
	BNN SVI [24]	79.48	71.73	52.52	71.51	91.86	*90.12*	36.21	94.71
	SoftMax [6]	80.55	73.17	50.49	*73.92*	91.95	89.16	32.00	95.03
	RPL [27]	80.21	72.04	52.83	71.46	90.64	88.79	38.43	*95.59*
	DEAR [5]	79.00	67.10	52.44	71.33	86.04	87.38	87.40	94.48
	PSL(ours)	**86.53**	**79.95**	**40.99**	**76.53**	**95.75**	**94.96**	**18.96**	**95.62**
	Δ	(+2.10)	(+3.26)	(-6.75)	(+2.61)	(+2.41)	(+4.84)	(-9.99)	(+0.03)

Table 1. Comparison with state-of-the-art methods on **HMDB51 and MiTv2 (OoD)** using TSM backbone. Acc. refers to closed-set accuracy. AUROC, AUPR and FPR95 are open-set metrics. Best results are in **bold** and second best results in *italic*. The gap between best and second best is in **blue**. DEAR and our methods contain video-specific operation.

					InD		OoD					
	s	Q_{ns}	Q_{sc}	Q_{shuf}	Mean	Variance	Mean	Variance	AUROC↑	AUPR↑	FPR95↓	Acc.↑
$\mathcal{L}_{PL}$	✗	✗	✗	✗	0.81	0.0015	0.63	0.0029	80.95	52.79	52.51	72.36
$\mathcal{L}_{PSL}$	✓	✗	✗	✗	0.79	0.0016	0.62	0.0028	81.79	54.16	52.33	72.33
$\mathcal{L}_{PSL}^{CT}$	✓	✓	✗	✗	0.71	0.0022	0.61	0.0036	82.60	57.36	50.03	72.17
	✓	✓	✓	✗	0.71	0.0023	0.49	0.0035	83.42	59.05	51.32	72.28
	✓	✓	✓	✓	0.74	0.0016	0.63	0.0029	86.43	65.58	41.75	77.19

Table 2. Abaltion results of different components in $\mathcal{L}_{PSL}^{CT}$.

Results

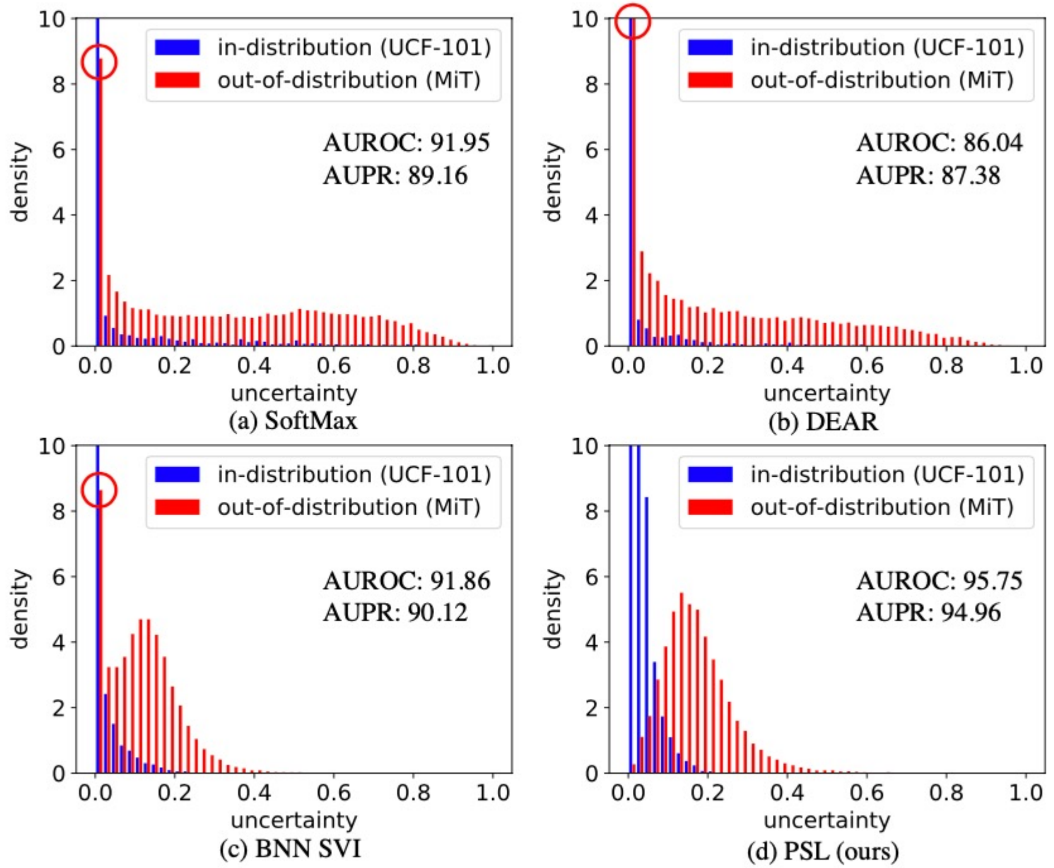


Figure 4. The uncertainty distribution of InD and OoD samples of (a) Softmax, (b) DEAR, (c) BNN SVI and (d) our PSL method.

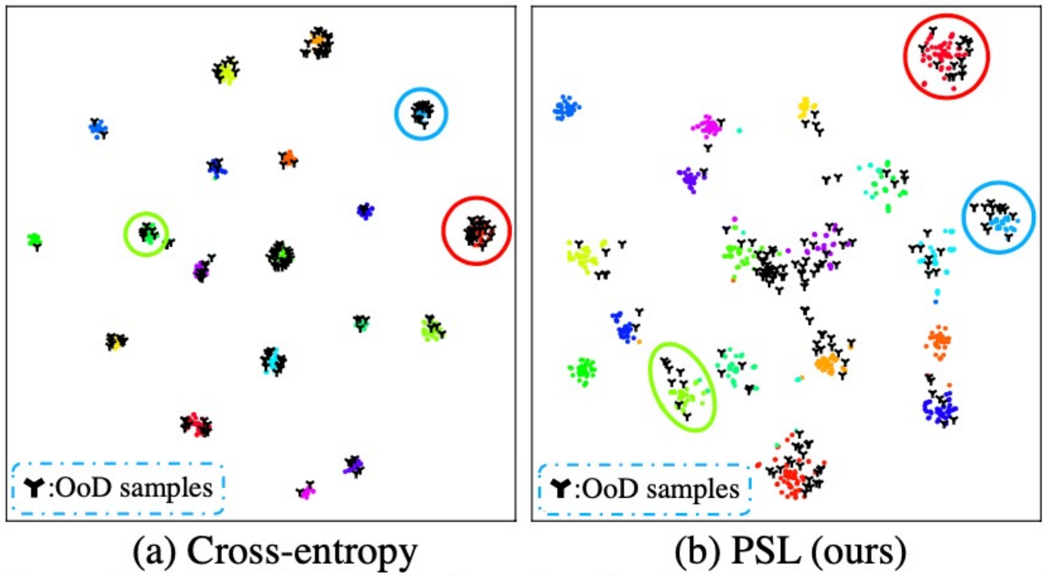


Figure 5. Feature representation visualization of cross-entropy and our PSL method. OoD samples are in black and InD samples are in other colors. In the red, blue and green circles, it is clear that OoD samples distribute at the edge of InD samples in our PSL, while greatly overlap with each other in the cross-entropy method.