Image Background Serves as Good Proxy for Out-of-distribution Data

Sen Pei

ByteDance Inc.

ByteDance

ICLR
International Conference On Learning Representations

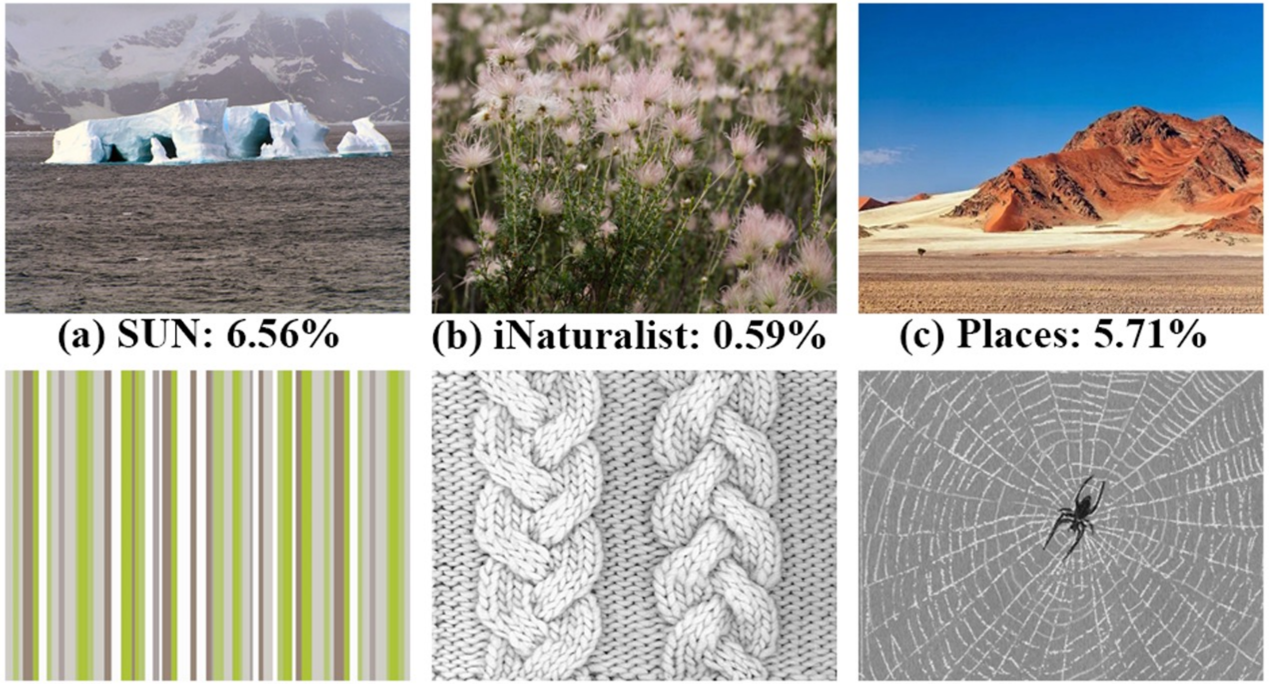
Outline

1. Background of Out-of-distribution detection
2. Motivation
3. The Detailed Methods
4. Evaluation Results and Visualization

Background of Out-of-distribution detection

Definition:

Out-of-distribution (OOD) data indicates the samples which are not included in the training images, i.e., both the category and domain are different from the training dataset.



(a) SUN: 6.56% (b) iNaturalist: 0.59% (c) Places: 5.71%

(d) Textures: 8.87% (e) Textures: 39.77% (f) Textures: 86.40%

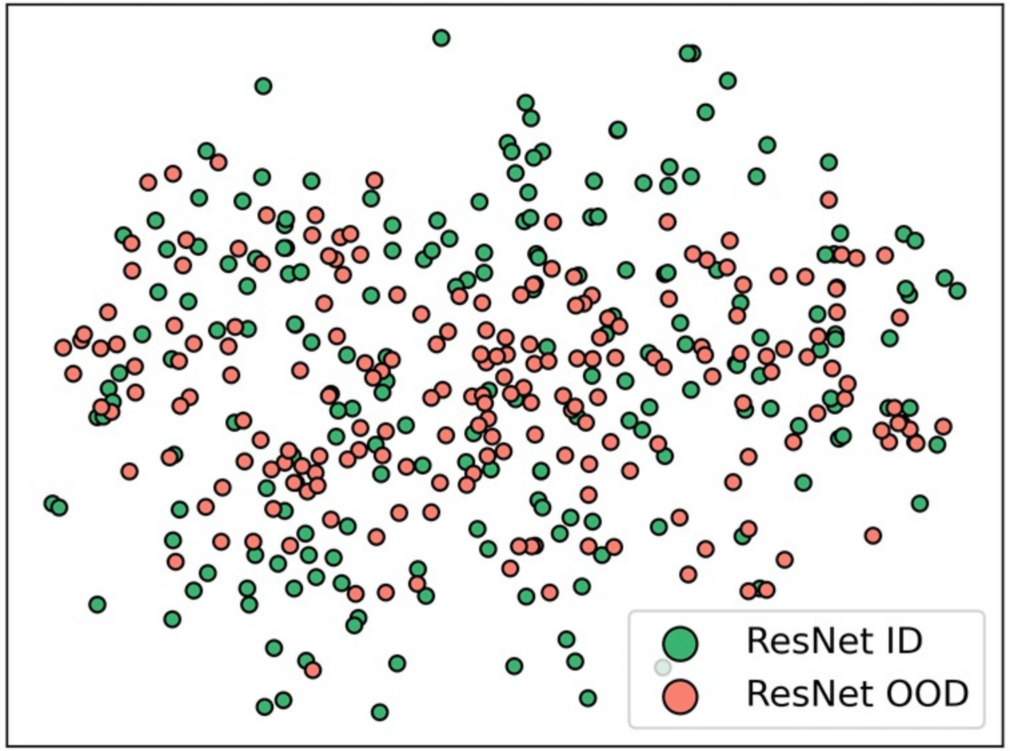
Left:

These images come from several different datasets, such as SUN, iNaturalist. Places, and Textures. They are treated as OOD data in ImageNet classification task.

Background of Out-of-distribution detection

Situation:

Conventional neural networks fail to distinguish the ID and OOD data. The following figure demonstrates the embedding space of ResNet-50 (trained on ImageNet).



Left:

Green: The feature of images sampled from ImageNet (ID).

Red: The feature of images sampled from iNaturalist (OOD).

The conventional classification models confuse the ID and OOD features.

Motivation

DS evidence theory:

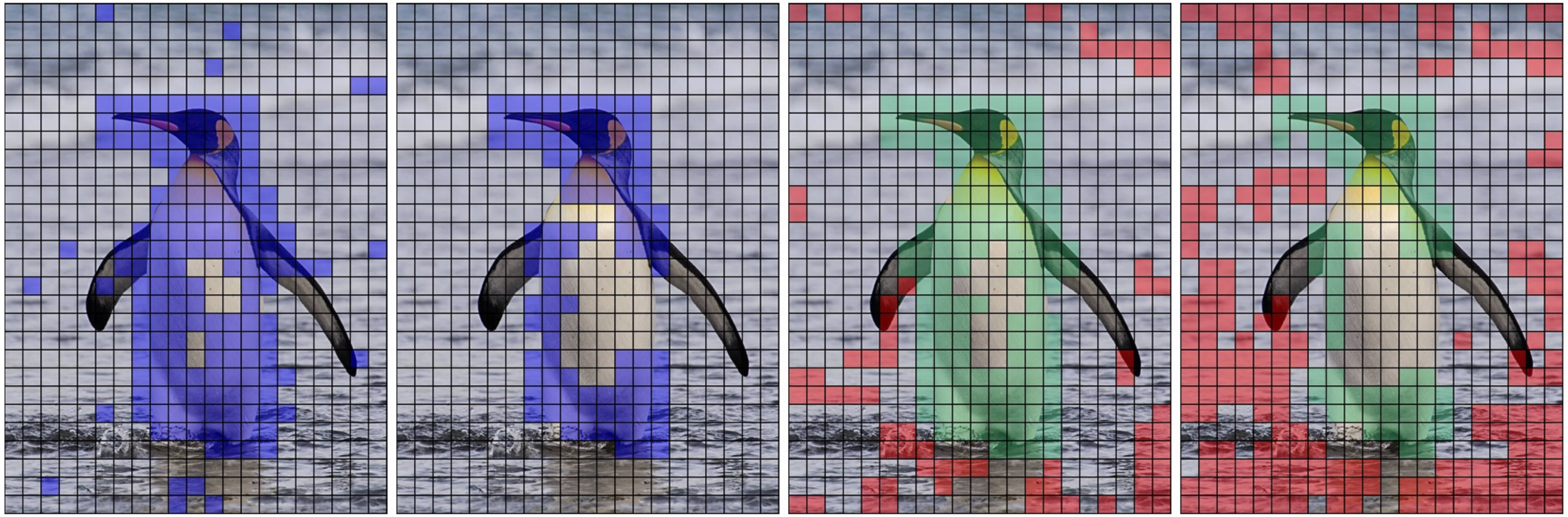
We omit the detailed derivation of our methodology, and present our conclusion directly in this part. Concretely, robust classification can be treated as conventional classification and OOD detection.

$$P(w_i|x) = \frac{e^{-s_i}}{\sum_{j=1}^{M} e^{-s_j}} \cdot \frac{\sum_{j=1}^{M} e^{-s_j}}{\sum_{j=1}^{M+1} e^{-s_j}} \triangleq \underbrace{P(w_i|x \in \mathcal{S}_{\mathbb{ID}}, x)}_{\text{ID factor}} \cdot \underbrace{P(x \in \mathcal{S}_{\mathbb{ID}}|x)}_{\text{OOD factor}}$$

Motivation

Locality:

Conventional neural networks can split the input images as object patches and background patches implicitly. We employ these background patches as proxy for OOD data. With these supervision, we train an auxiliary ID/OOD classifier.

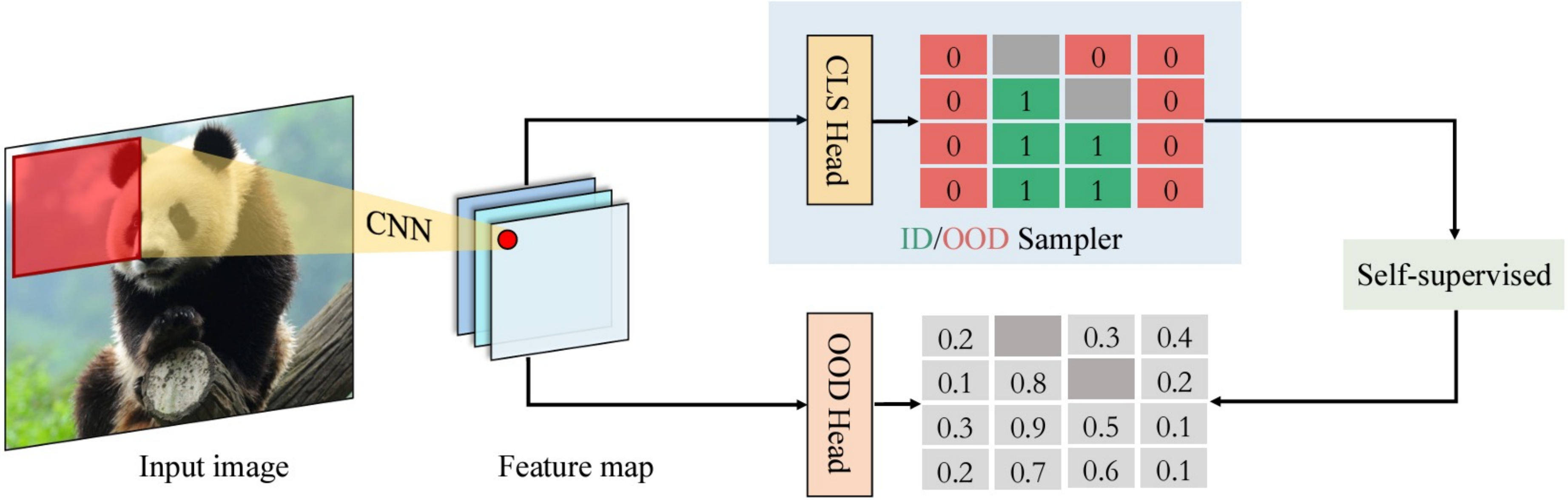


(a) Locality of ResNet-50. (b) Locality of SSOD. (c) Samplers of ResNet-50. (d) Samplers of SSOD.

The detailed method

SSOD:

Self-supervised sampling for out-of-distribution detection. The image patches identified as objects are treated as ID samples, while that identified as backgrounds are treated as OOD samples. These ID and OOD proxy provide supervision for training the OOD head.



Evaluation Results and Visualization

Results on ImageNet: SSOD achieves competitive OOD detection performance on ImageNet.

Method	*iNaturalist*		*SUN*		*Places*		*Texture*		*Average*	
	↓F	↑A	↓F	↑A	↓F	↑A	↓F	↑A	↓F	↑A
MSP	54.99	87.74	70.83	80.86	73.99	79.76	68.00	79.61	66.95	81.99
MSP (CLIP-B)	40.89	88.63	65.81	81.24	67.90	80.14	64.96	78.16	59.89	82.04
MSP (CLIP-L)	34.54	92.62	61.18	83.68	59.86	84.10	59.27	82.31	53.71	85.68
MaDist	97.00	52.65	98.50	42.41	98.40	41.79	55.80	85.01	87.43	55.47
ODIN	47.66	89.66	60.15	84.59	67.89	81.78	50.23	85.62	56.48	85.41
GODIN	61.91	85.40	60.83	85.60	63.70	83.81	77.85	73.27	70.43	82.02
KLM	27.36	93.00	67.52	78.72	72.61	76.49	49.70	87.07	54.30	83.82
Energy	55.72	89.95	59.26	85.89	64.92	82.86	53.72	85.99	58.41	86.17
KNN (w/o)	59.08	86.20	69.53	80.10	77.09	74.87	**11.56**	**97.18**	54.32	84.59
KNN (w/)	30.18	94.89	48.99	88.63	59.15	84.71	<u>16.97</u>	<u>94.45</u>	38.82	90.67
MOS	**9.28**	**98.15**	40.63	92.01	49.54	89.06	60.43	81.23	39.97	90.11
Fort (ViT-B)	15.07	96.64	54.12	86.37	57.99	85.24	53.32	84.77	45.12	88.25
Fort (ViT-L)	15.74	96.51	52.34	87.32	55.14	86.48	51.38	85.54	43.65	88.96
MCM (CLIP-B)	30.91	94.61	37.59	92.57	44.69	89.77	57.77	86.11	42.74	90.77
MCM (CLIP-L)	28.38	94.95	<u>29.00</u>	<u>94.14</u>	**35.42**	**92.00**	59.88	84.88	<u>38.17</u>	<u>91.49</u>
SSOD (Ours)	<u>14.80</u>	<u>96.91</u>	**28.52**	**94.33**	<u>38.92</u>	<u>90.78</u>	45.32	87.02	**31.89**	**92.26**

Evaluation Results and Visualization

Results on CIFAR-10: SSOD achieves competitive OOD detection performance on CIFAR-10.

OOD	Metrics	Methods							
		MSP	MaDist	ODIN	GODIN	Energy	CSI	KNN	SSOD
SVHN	↓ FPR95	59.66	9.24	20.93	15.51	54.41	37.38	24.53	**2.12**
	↑ AUROC	91.25	97.80	95.55	96.60	91.22	94.69	95.96	**99.44**
LSUN	↓ FPR95	45.21	67.73	7.26	4.90	10.19	5.88	25.29	**4.42**
	↑ AUROC	93.80	73.61	98.53	99.07	98.05	98.86	95.69	**99.11**
iSUN	↓ FPR95	54.57	**6.02**	33.17	34.03	27.52	10.36	25.55	10.06
	↑ AUROC	92.12	**98.63**	94.65	94.94	95.59	98.01	95.26	98.16
Texture	↓ FPR95	66.45	23.21	56.40	46.91	55.23	28.85	27.57	**1.91**
	↑ AUROC	88.50	92.91	86.21	89.69	89.37	94.87	94.71	**99.59**
Places	↓ FPR95	62.46	83.50	63.04	62.63	42.77	38.31	50.90	**7.44**
	↑ AUROC	88.64	83.50	86.57	87.31	91.02	93.04	89.14	**98.42**
Average	↓ FPR95	57.67	37.94	36.16	32.80	38.02	24.20	30.80	**5.19**
	↑ AUROC	90.90	89.29	92.30	93.52	93.05	95.90	94.15	**98.94**
	↑ ID ACC	94.21	94.21	94.21	93.96	94.21	**94.38**	94.21	94.17

Results on Hard-OOD-Dataset

Method	*ImageNet-O*		*OpenImage-O*		*Average*	
	↓F	↑A	↓F	↑A	↓F	↑A
MSP	93.85	56.13	63.53	84.50	78.69	70.32
Energy	90.10	53.95	76.83	75.95	83.47	64.95
ODIN	93.25	52.87	64.49	81.53	78.87	67.20
MaxL	92.65	54.39	65.50	81.50	79.08	67.95
KLM	88.50	67.00	60.58	87.31	74.54	77.16
ReAct	**72.85**	**81.15**	60.79	85.30	66.82	**83.23**
MaDist	78.45	68.02	55.91	89.52	67.18	78.77
ViM	76.00	74.80	**50.45**	**90.76**	**63.23**	82.78
SSOD	79.80	74.43	55.41	89.73	67.61	82.08

Evaluation Results and Visualization

Results on Hard-OOD-Dataset

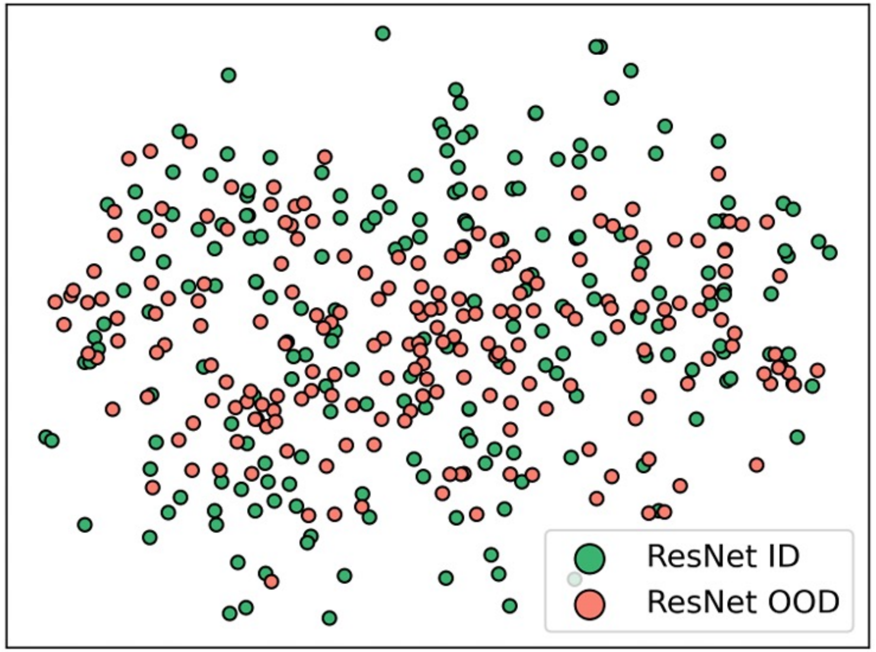
Method	*ImageNet-O*		*OpenImage-O*		*Average*	
	↓F	↑A	↓F	↑A	↓F	↑A
MSP	93.85	56.13	63.53	84.50	78.69	70.32
Energy	90.10	53.95	76.83	75.95	83.47	64.95
ODIN	93.25	52.87	64.49	81.53	78.87	67.20
MaxL	92.65	54.39	65.50	81.50	79.08	67.95
KLM	88.50	67.00	60.58	87.31	74.54	77.16
ReAct	**72.85**	**81.15**	60.79	85.30	<u>66.82</u>	**83.23**
MaDist	78.45	68.02	55.91	89.52	67.18	78.77
ViM	<u>76.00</u>	<u>74.80</u>	**50.45**	**90.76**	**63.23**	<u>82.78</u>
SSOD	79.80	74.43	<u>55.41</u>	<u>89.73</u>	67.61	82.08

Evaluation Results and Visualization

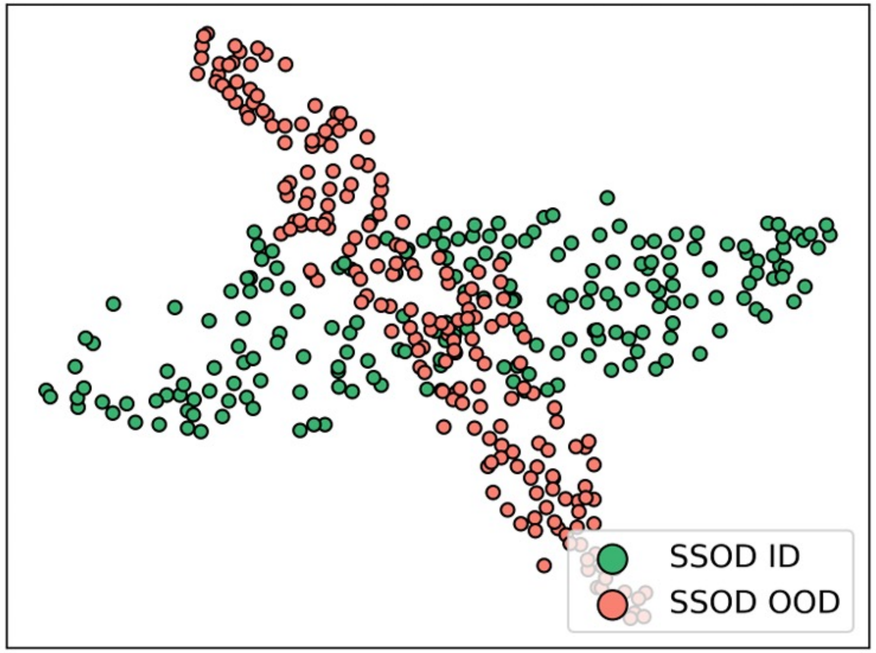
Visualization

Left: conventional neural networks confuses ID (ImageNet) and OOD (iNaturalist) images.

Right: SSOD can better distinguish the ID (ImageNet) and OOD (iNaturalist) images.



(a) ID *vs.* OOD features in ResNet-50.



(b) ID *vs.* OOD features in SSOD.

Evaluation Results and Visualization

Visualization

The synthetic ID and OOD features serve as good proxy for real data. In the following figure, the gray/blue points are authentic OOD/ID images, while the red/green points are synthetic OOD/ID data generated by our proposed SSOD.