WED-PM-318

Speaker: Hyunjun Choi

Balanced Energy Regularization Loss for Out-of-distribution Detection

Hyunjun Choi[1,2], Hawook Jeong[2], Jin Young Choi[1]

numb7315@snu.ac.kr hawook@rideflux.com jychoi@snu.ac.kr

[1]ASRI, ECE., Seoul National University, [2]RideFlux Inc

Perception and Intelligence Laboratory

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Previous work

- "Observer" methods
  - ❖ model the prediction uncertainty of a pre-trained network, without modifying its architecture or parameters
  - ❖ Hendrycks and Gimpel,2016; Liang et al., 2017; Lee et al., 2018
    - Baseline for OOD : maximum softmax score (MSP)
    - ODIN
    - Mahalanobis distance based OOD

- "Mutators" methods
  - ❖ modify the network structure or loss, and depend on training its parameters to provide a confidence measure
  - ❖ Hendrycks et al., 2018; Hsu et al., 2020
    - Deep anomaly detection with outlier exposure
    - Generalized odin: Detecting out-of distribution image without learning from out-of-distribution data

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Previous works

- Using additional data (auxiliary dataset)

  - ❖ Use auxiliary data as outlier data has shown promising performance
  - ❖ Do not overlap with classes of OOD test data, in-distribution data
  - ❖ Image classification task
    - Deep anomaly detection with outlier exposure (OE) (ICLR 2018)
    - Energy-based OOD detection (EnergyOE) (Neurips 2020)
  - ❖ Semantic segmentation task
    - Entropy Maximization and Meta Classification for Out-of-Distribution Detection in Semantic Segmentation (Meta OOD) (ICCV 2021)
    - Pixel-wise Energy-biased Abstention Learning for Anomaly Segmentation on Complex Urban Driving Scenes (PEBAL) (ECCV 2022)

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Previous works

- Deep anomaly detection with outlier exposure (OE) (ICLR 2018)

  ❖ Using auxiliary data as outlier data in training for the first time
  ❖ Superior performance compared to baseline (MSP)
  ❖ Leverage the regularization loss for outlier data
  -Fine-tuning (main theme)
  ❖ Proposed loss : Outlier exposure loss (OE loss)

$$\min_{\theta} \boxed{\mathbb{E}_{(\mathbf{x},y)\sim D_{in}^{train}}[-\log F_y(\mathbf{x})]} + \lambda \boxed{L_{OE}}, \qquad (1)$$

$$\boxed{L_{OE} = \mathbb{E}_{\mathbf{x}_{out}\sim D_{out}^{train}}[H(\mathbf{u}; F(\mathbf{x}))],}$$

H: cross entropy loss
u: uniform distribution

  ❖ Meta OOD (ICCV 2021) : Using OE loss in semantic segmentation OOD task

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Previous works

- Energy-based OOD detection (EnergyOE) (Neurips 2020)

  ❖ Instead of MSP, Energy score is proposed
  ❖ Leverage the energy regularization loss in fine-tuning
  ❖ Proposed loss : Energy regularization loss

$$\boxed{\begin{aligned} L_{energy} &= L_{in,hinge} + L_{out,hinge} \\ &= \mathbb{E}_{(\mathbf{x}_{in},y)\sim D_{in}^{train}}[(\max(0, E(\mathbf{x}) - m_{in}))^2] \\ &+ \mathbb{E}_{\mathbf{x}_{out}\sim D_{out}^{train}}[(\max(0, E(\mathbf{x}) - m_{out}))^2], \end{aligned}} \quad (2)$$

$$\boxed{E(\mathbf{x}; f) = -T \cdot \log(\textstyle\sum_{j=1}^{K} e^{f_j(\mathbf{x}))/T})}$$

T=1 fixed,
double hinge loss for energy

  ❖ PEBAL (ECCV 2022): Using energy regularization loss in semantic segmentation OOD task

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Motivation

Previous auxiliary data based fine-tuning methods

$$L_{OE} = \boxed{\mathbb{E}_{\mathbf{x}_{out}\sim D_{out}^{train}}[H(\mathbf{u}; F(\mathbf{x}))],}$$

Deep anomaly detection with outlier exposure (OE) (ICLR 2018)

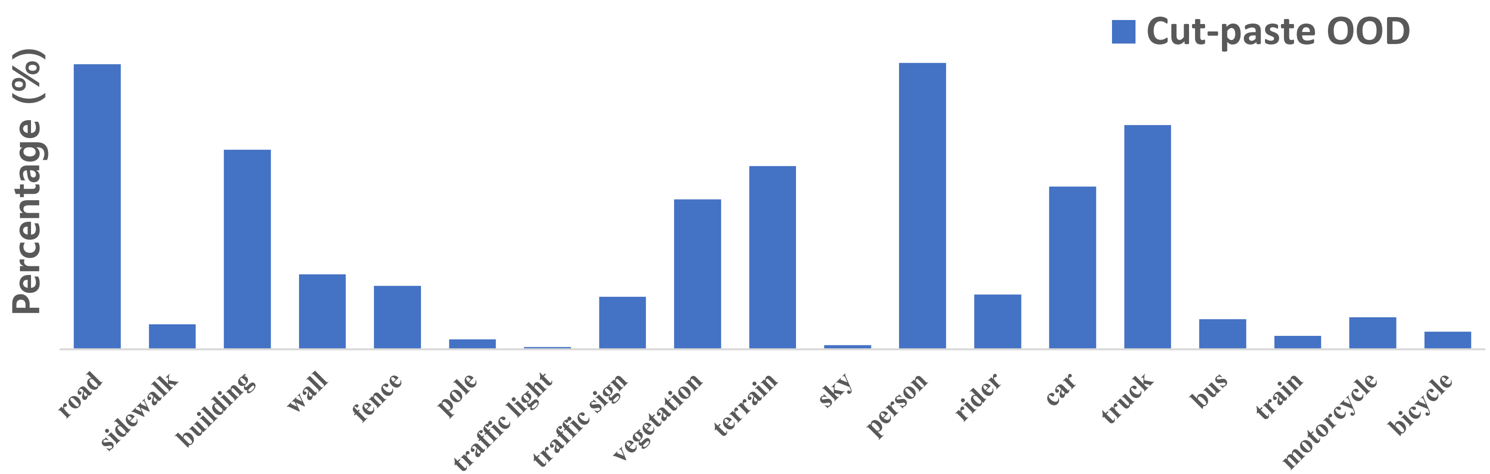
$$\begin{aligned} L_{energy} &= L_{in,hinge} + L_{out,hinge} \\ &= \mathbb{E}_{(\mathbf{x}_{in},y)\sim D_{in}^{train}}[(\max(0, E(\mathbf{x}) - m_{in}))^2] \\ &+ \boxed{\mathbb{E}_{\mathbf{x}_{out}\sim D_{out}^{train}}[(\max(0, E(\mathbf{x}) - m_{out}))^2],} \end{aligned}$$

Energy-based OOD detection (EnergyOE) (Neurips 2020)

Previous approach
: Equal regularization loss for all x_{out} data

Motivation

why x_{out} data should have different regularization loss?



OOD Inference result on pretrained network

: x_{out} data tends to have an imbalance in the distribution of the auxiliary OOD data across classes

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Key method

$$
\begin{aligned}
L_{energy} &= L_{in,hinge} + L_{out,hinge} \\
&= \mathbb{E}_{(\mathbf{x}_{in},y)\sim D_{in}^{train}}[(\max(0, E(\mathbf{x}) - m_{in}))^2] \\
&+ \mathbb{E}_{\mathbf{x}_{out}\sim D_{out}^{train}}[(\max(0, E(\mathbf{x}) - m_{out}))^2],
\end{aligned}
$$

$$
\begin{aligned}
&L_{energy,bal} = L_{in,hinge} + L_{out,bal} \\
&= \mathbb{E}_{(\mathbf{x}_{in},y)\sim D_{in}^{train}}[(\max(0, E(\mathbf{x}) - m_{in}))^2] \\
&+ \mathbb{E}_{\mathbf{x}\sim D_{out}^{train}}[(\max(0, E(\mathbf{x}) - m_{out} - \alpha \boxed{Z_\gamma}))^2 \boxed{Z_\gamma}],
\end{aligned}
$$

Different regularization loss for x_{out} data

Key method

$$L_{energy,bal} = L_{in,hinge} + L_{out,bal}$$
$$= \mathbb{E}_{(\mathbf{x}_{in},y)\sim D_{in}^{train}}[(\max(0, E(\mathbf{x}) - m_{in}))^2]$$
$$+ \mathbb{E}_{\mathbf{x}\sim D_{out}^{train}}[(\max(0, E(\mathbf{x}) - m_{out} - \alpha \boxed{Z_\gamma}))^2 \boxed{Z_\gamma}],$$

Posterior probability
(softmax output of network)

$$P(y = i|\mathbf{x}, o) = \frac{e^{f_i(\mathbf{x})}}{\sum_{j=1}^{K} e^{f_j(\mathbf{x})}}.$$

Prior probability
(OOD inference result on pretrained network)

$$P(y = i|o) = \frac{N_i}{N_1 + N_2 + \cdots + N_K}.$$

Z term: measure whether a sample is of the majority or minority class.

$$Z = \sum_{j=1}^{K} P(y = j|\mathbf{x}, o) P(y = j|o).$$

Extended version of Z:
By power (gamma) on prior probability

$$P_\gamma(y = i|o) = L^1 norm\{P^\gamma(y = i|o)\}$$

$$\boxed{Z_\gamma = \sum_{j=1}^{K} P(y = j|\mathbf{x}, o) P_\gamma(y = j|o),}$$

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Training procedure

Algorithm 1: Balanced Energy Learning

Input: f:Pre-trained model
Data: D_{in}:in-distribution training set,
D_{out}:OOD training set
Step1: Inference on OOD training set
Load the weight of pre-trained model f;
$N_j \longleftarrow 0$, for all j=1 to K
for $t = 1$ *to* T_1 **do**
    Sample a mini batch $D_{mini,o}$ from D_{out}
    Inference on the mini batch $f(D_{mini,o})$
    for $j = 1$ *to* K **do**
        $n_j \longleftarrow \text{count}(\max_i f(D_{mini,o}), j)$
        $N_j \longleftarrow N_j + n_j$
Compute prior probability of OOD as Eq. (3).
Step2: Fine-tuning the pre-trained model
for $t = T_1 + 1$ *to* T_2 **do**
    Sample mini-batches $D_{mini,i}$ and $D_{mini,o}$ from D_{in} and D_{out}, respectively.
    Update unfrozen classification layers of f by minimizing Eq. (8).

Step1:
Prior probability calculation
(OOD inference result on pretrained network)

Step2:
Finetuning the pretrained model based on the balanced energy regularization loss

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

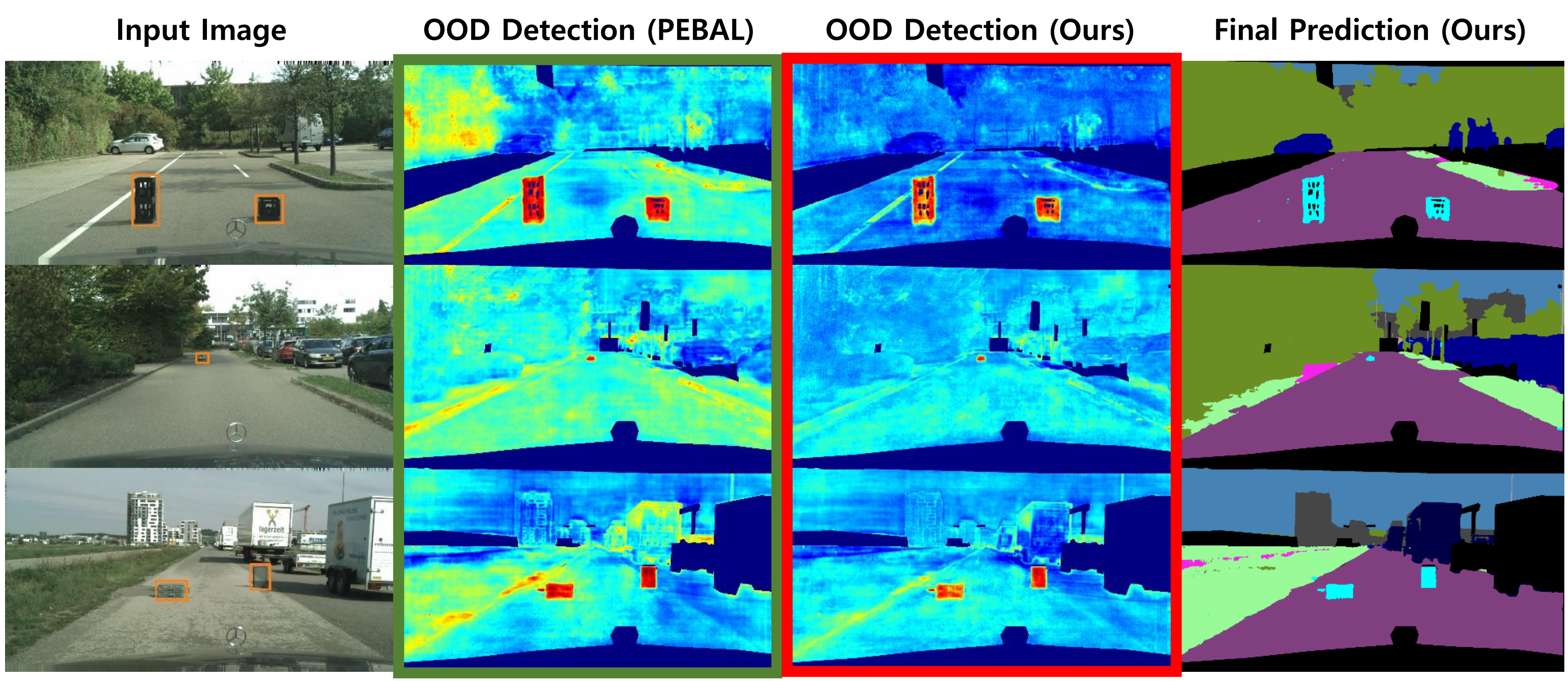
Experimental result

- OOD detection in semantic segmentation
- OOD detection in Long-tailed Image Classification
- OOD detection in Image Classification

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Qualitative result

- OOD in semantic segmentation



JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Quantitative result

- OOD in semantic segmentation
  - Fishyscapes test sets

$R^\dagger$: Re-training, $E^\dagger$: Extra Network, $O^\dagger$: OoD Data.

Method	$R^\dagger$	$E^\dagger$	$O^\dagger$	FS Lost & Found		FS Static	
				AP↑	FPR↓	AP↑	FPR↓
MSP [16]	✗	✗	✗	1.77	44.85	12.88	39.83
En† [17]	✗	✗	✗	2.93	44.83	15.41	39.75
kNN† [3]	✗	✗	✗	3.55	30.02	44.03	20.25
SML [19]	✗	✗	✗	31.05	21.52	53.11	19.64
BD† [36]	✓	✗	✗	9.81	38.46	48.70	15.05
DSN† [3]	✗	✓	✗	3.01	32.90	40.86	21.29
DMN† [3]	✗	✓	✗	4.25	47.15	62.14	17.43
IR† [28]	✗	✓	✗	5.70	48.05	29.60	27.13
DLR† [3]	✗	✓	✓	4.65	24.36	57.16	13.39
SB† [10]	✗	✓	✓	43.22	15.79	72.59	18.75
DODH† [2]	✓	✓	✓	31.31	19.02	96.76	0.29
OTVC†	✓	✗	✓	10.29	22.11	45.00	19.40
DD† [33]	✓	✗	✓	34.28	47.43	31.30	84.60
DH† [12]	✓	✗	✓	47.06	3.97	80.23	5.95
PEBAL [41]	✓	✗	✓	44.17	7.58	92.38	1.73
Ours	✓	✗	✓	**51.83**	**3.76**	**94.62**	**0.99**

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Quantitative result

- OOD in semantic segmentation
  - Fishyscapes validation sets
  - Road Anomaly test set

Method	FS Lost & Found			FS Static			Road Anomaly		
	AUC↑	AP↑	FPR↓	AUC↑	AP↑	FPR↓	AUC↑	AP↑	FPR↓
MSP [16]	89.29	4.59	40.59	92.36	19.09	23.99	67.53	15.72	71.38
Max Logit [16]	93.41	14.59	42.21	95.66	38.64	18.26	72.78	18.98	70.48
Entropy [17]	90.82	10.36	40.34	93.14	26.77	23.31	68.80	16.97	71.10
Energy [29]	93.72	16.05	41.78	95.90	41.68	17.78	73.35	19.54	70.17
Mahalanobis [24]	96.75	56.57	11.24	96.76	27.37	11.7	62.85	14.37	81.09
Meta-OOD [5]	93.06	41.31	37.69	97.56	72.91	13.57	-	-	-
Synboost [10]	96.21	60.58	31.02	95.87	66.44	25.59	81.91	38.21	64.75
SML [19]	94.97	22.74	33.49	97.25	66.72	12.14	75.16	17.52	70.70
Deep Gambler [31]	97.82	31.34	10.16	98.88	84.57	3.39	78.29	23.26	65.12
PEBAL [41]	98.96	58.81	4.76	**99.61**	92.08	1.52	87.63	**45.10**	44.58
Balanced Energy PEBAL (Ours)	**99.03**	**67.07**	**2.93**	99.55	**92.49**	**1.17**	**88.36**	43.58	**41.54**
EnergyOE [29]	98.14	45.61	8.21	99.32	89.12	2.62	83.32	32.59	53.01
Balanced EnergyOE (Ours)	**98.42**	**54.58**	**6.70**	**99.43**	**91.77**	**1.63**	**85.50**	**34.90**	**46.60**

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Quantitative result

- OOD in Long-tailed Image Classification
  - CIFAR10

(a)

Dataset	Method	AUC↑	AP↑	FPR↓
Texture	OE (tune)	87.98	80.05	45.54
	EnergyOE (tune)	95.53	**92.93**	23.26
	Ours	**95.69**	92.38	**21.26**
SVHN	OE (tune)	92.10	95.52	27.37
	EnergyOE (tune)	96.63	98.46	14.52
	Ours	**97.74**	**98.89**	**9.87**
CIFAR100	OE (tune)	78.24	76.35	65.28
	EnergyOE (tune)	84.44	84.63	59.92
	Ours	**85.20**	**84.98**	**57.95**
Tiny ImageNet	OE (tune)	81.47	75.79	58.68
	EnergyOE (tune)	88.40	84.95	45.17
	Ours	**88.92**	**84.98**	**42.38**
LSUN	OE (tune)	86.19	85.85	54.49
	EnergyOE (tune)	94.00	**93.70**	26.96
	Ours	**94.48**	93.15	**23.88**
Places365	OE (tune)	84.27	93.84	59.08
	EnergyOE (tune)	92.51	97.14	32.88
	Ours	**93.35**	**97.23**	**28.25**
Average	OE (tune)	85.04	84.57	51.74
	EnergyOE (tune)	91.92	**91.97**	33.79
	Ours	**92.56**	91.94	**30.60**

(b)

Dataset	Method	AUC↑	AP↑	FPR↓	ACC↑
Average	MSP [17](ST)	70.96	69.35	67.37	69.83
	Energy [29](ST)	75.93	72.91	61.00	69.83
	OECC [38]	87.28	86.29	45.24	60.16
	EnergyOE [29](scratch)	89.31	88.92	40.88	74.68
	OE [18](scratch)	89.77	87.25	34.65	73.84
	PASCL [43]	90.99	89.24	33.36	77.08
	Open-Sampling [45]	90.24	85.44	31.00	77.06
	OE [18](tune)	85.04	84.57	51.74	69.79
	EnergyOE [29](tune)	91.92	**91.97**	33.79	74.53
	Ours	**92.56**	91.94	**30.60**	76.22
	Ours+AdjLogit [34]	**92.56**	91.94	**30.60**	**81.37**

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Quantitative result

- OOD in Long-tailed Image Classification
  - CIFAR100

(a)

Dataset	Method	AUC↑	AP↑	FPR↓
Texture	OE (tune)	66.29	51.98	84.04
	EnergyOE (tune)	79.56	70.88	68.60
	Ours	**82.10**	**73.09**	**64.19**
SVHN	OE (tune)	74.93	85.41	63.94
	EnergyOE (tune)	86.19	91.74	42.27
	Ours	**88.66**	**92.88**	**33.79**
CIFAR10	OE (tune)	59.44	56.34	84.70
	EnergyOE (tune)	**61.15**	**56.66**	**82.60**
	Ours	59.40	54.97	85.16
Tiny ImageNet	OE (tune)	66.24	51.07	80.04
	EnergyOE (tune)	70.78	55.90	74.43
	Ours	**71.42**	**56.52**	**74.22**
LSUN	OE (tune)	73.46	59.07	73.05
	EnergyOE (tune)	81.61	69.16	57.37
	Ours	**83.83**	**71.23**	**52.04**
Places365	OE (tune)	71.70	85.08	74.62
	EnergyOE (tune)	79.12	89.09	61.96
	Ours	**81.10**	**89.94**	**57.52**
Average	OE (tune)	68.68	64.83	76.73
	EnergyOE (tune)	76.40	72.24	64.54
	Ours	**77.75**	**73.10**	**61.15**

(b)

Dataset	Method	AUC↑	AP↑	FPR↓	ACC↑
Average	MSP [17](ST)	60.26	57.58	84.00	38.74
	Energy [29](ST)	63.22	59.06	81.12	38.74
	OECC [38]	70.38	66.87	73.15	32.93
	EnergyOE [29](scratch)	71.10	67.23	71.78	39.05
	OE [18](scratch)	72.91	67.16	68.89	39.04
	PASCL [43]	73.32	67.18	67.44	43.10
	Open-Sampling [45]	74.46	69.49	66.82	39.86
	OE [18](tune)	68.68	64.83	76.73	38.93
	EnergyOE [29](tune)	76.40	72.24	64.54	40.65
	Ours	**77.75**	**73.10**	**61.15**	41.05
	Ours+AdjLogit [34]	**77.75**	**73.10**	**61.15**	**45.66**

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Quantitative result

- OOD in Image Classification
  - CIFAR10, CIFAR100

(a)

Dataset	Method	AUC↑	AP↑	FPR↓	ACC↑
Average	MSP [17](ST)	89.25	86.63	31.32	**93.69**
	Energy [29](ST)	91.55	89.88	29.07	**93.69**
	OECC [38]	96.33	95.38	**14.36**	91.57
	OE [18](tune)	95.68	95.36	18.20	93.37
	EnergyOE [29](tune)	96.77	**96.72**	14.82	93.30
	Ours	**96.83**	96.70	14.51	93.00

(b)

Dataset	Method	AUC↑	AP↑	FPR↓	ACC↑
Average	MSP [17](ST)	76.14	71.29	62.78	**75.70**
	Energy [29](ST)	79.78	73.31	57.59	**75.70**
	OECC [38]	84.03	77.94	45.26	69.55
	OE [18](tune)	82.76	77.93	51.72	74.33
	EnergyOE [29](tune)	85.84	**80.99**	43.02	74.95
	Ours	**85.85**	80.91	**42.93**	74.83

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Summary

- For OOD detection, we focus on the fine-tuning methodology using auxiliary data
- we propose a new balanced energy regularization loss
- The main idea of our loss is to apply large regularization to auxiliary samples of majority classes, compared to those of minority
- We show the effectiveness of our novel loss through extensive experiments on semantic segmentation, long-tailed image classification, and image classification datasets

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA