JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

OpenMix: Exploring Outlier Samples for Misclassification Detection

Fei Zhu, Zhen Cheng, Xu-Yao Zhang, Cheng-Lin Liu

MAIS, Institute of Automation of Chinese Academy of Sciences
School of Artificial Intelligence, University of Chinese Academy of Sciences

WED-AM-367

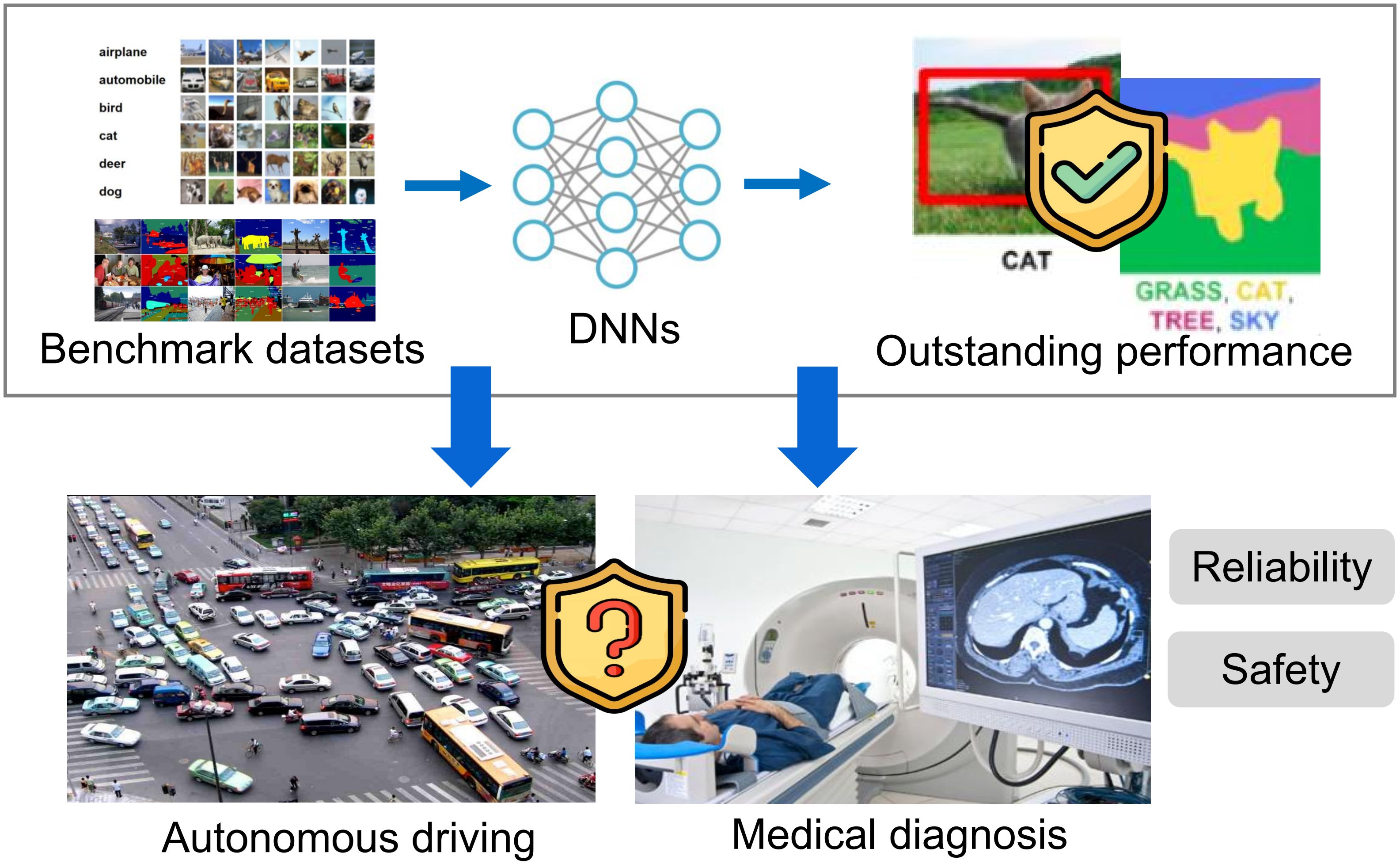
OpenMix Failure Detection

code and useful paper list

Overview

- Deep Neural Networks tend to be overconfident for their false predictions
- We propose to leverage outlier data for misclassification detection
- We provide analysis of why the well-known out-of-distribution (OOD) detection method OE is harmful for detecting misclassification errors
- The proposed OpenMix, including learning with reject class and outlier transformation, is a unified method for misclassification and OOD detection

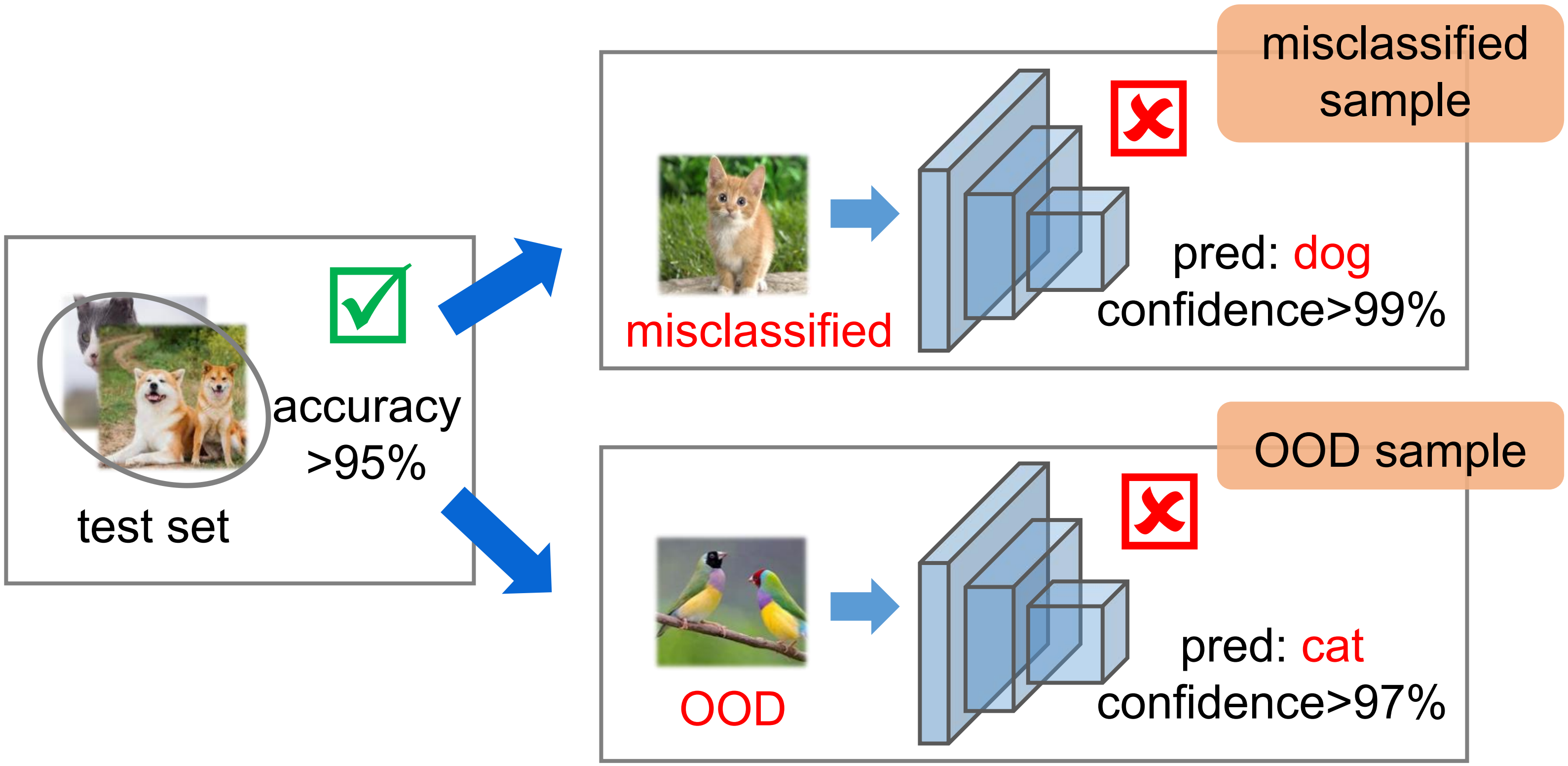
Background



- In risk-sensitive applications like autonomous driving, it is important to provide reliable confidence to avoid using wrong predictions

Background

- Deep Neural Networks tend to be overconfident for their false predictions：
  ① misclassified samples from known classes
  ② out-of-distribution (OOD) samples from unknown classes



- Recently, many works focus on out-of-distribution detection, ignoring detecting misclassified errors

Motivation

- In this paper, we focus on MisD, and propose a simple approach that can detect misclassified and OOD samples in a unified manner

Why are human beings good at confidence estimation?

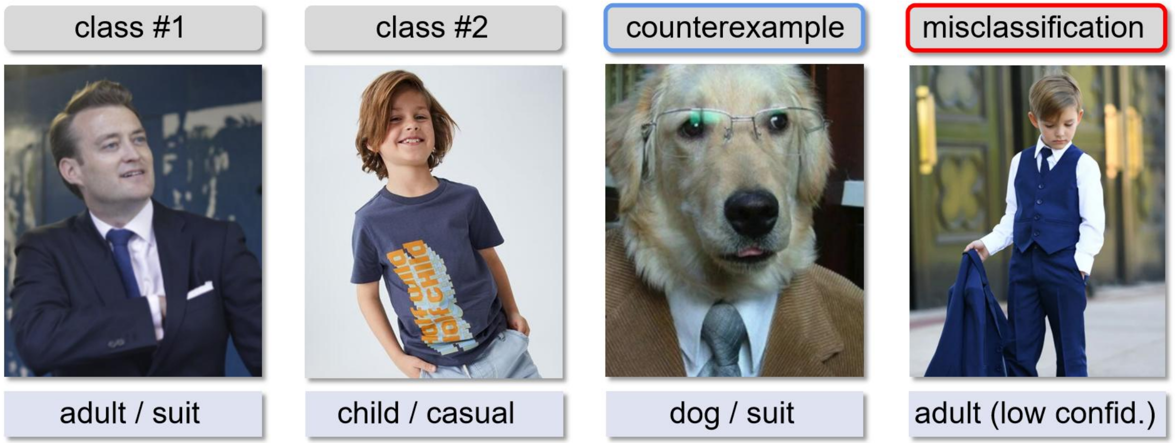


Illustration of advantages of counterexample data for reliable prediction.
Counterexample could help reduce model's confidence on wrong predictions

- Humans learn and predict in context, where we have abundant prior knowledge about other entities in the open world
- We propose to leverage outlier data, i.e., unlabeled random samples from non-target classes, as counterexamples for overconfidence mitigation

Motivation

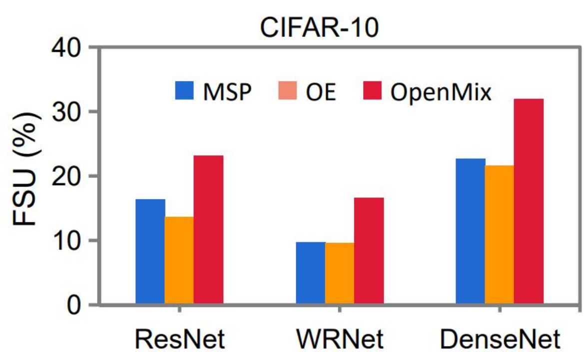
- Understanding the effect of OE

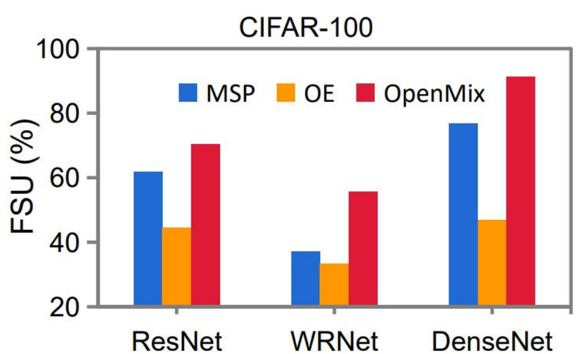
Dateset	Method	AURC ↓			AUROC ↑			FPR95 ↓		
		ResNet110	WRNet	DenseNet	ResNet110	WRNet	DenseNet	ResNet110	WRNet	DenseNet
CIFAR-10	MSP [25]	**9.52±0.49**	**4.76±0.62**	**5.66±0.45**	**90.13±0.46**	**93.14±0.38**	**93.14±0.65**	**43.33±0.59**	**30.15±1.98**	**38.64±4.70**
	+ OE [26]	10.10±0.54	4.83±0.13	8.23±0.95	90.02±0.36	93.09±0.15	91.44±0.15	46.89±1.78	38.78±2.59	45.86±2.30
CIFAR-100	MSP [25]	**89.05±1.39**	**46.84±0.90**	**66.11±1.56**	**84.91±0.13**	**88.50±0.44**	**86.20±0.04**	**65.65±1.72**	**56.64±1.33**	**62.79±0.83**
	+ OE [26]	103.06±2.50	58.05±1.21	86.96±2.27	83.81±0.49	86.36±0.20	84.25±0.50	71.11±0.77	62.96±0.38	70.39±0.65

MisD performance can not be improved with OE

$$\pi_{\text{inter}} = \frac{1}{Z_{inter}} \sum_{y_l, y_k, l \neq k} d(\boldsymbol{\mu}(Z_{y_l}), \boldsymbol{\mu}(Z_{y_k})) \qquad \pi_{\text{intra}} = \frac{1}{Z_{intra}} \sum_{y_l \in y} \sum_{\boldsymbol{z}_i, \boldsymbol{z}_j \in Z_{y_l}, i \neq j} d(\boldsymbol{z}_i, \boldsymbol{z}_j)$$

feature space uniformity (FSU): $\pi_{\text{fsu}} = \pi_{\text{intra}} / \pi_{\text{inter}}$





By forcing the outliers to be uniformly distributed over original classes, OE leads to over-compressed distributions, making it harder to separate misclassified samples from correct ones

Proposed Method: OpenMix

- How to use outliers for MisD?

On learning objective

uniform distribution for outliers
$\ell_{\mathrm{OE}}(f(\boldsymbol{x})) = \mathrm{KL}(\mathcal{U}(y) \| f(\boldsymbol{x}))$

learning with reject class
$\ell(f(\widetilde{\boldsymbol{x}}), \widetilde{y})$ where $\widetilde{y} = k+1$

On outlier data

original outliers: large distribution gap with ID misclassified samples

outlier transformation
$\breve{\boldsymbol{x}} = \lambda \boldsymbol{x} + (1-\lambda)\widetilde{\boldsymbol{x}}, \quad \mathbb{I}^{\breve{y}} = \lambda \mathbb{I}^{y} + (1-\lambda)\mathbb{I}^{\widetilde{y}}$

$\widetilde{y} = k+1$

- Why OpenMix is beneficial for MisD?

overconfidence: low density regions with rich uncertainty are largely ignored

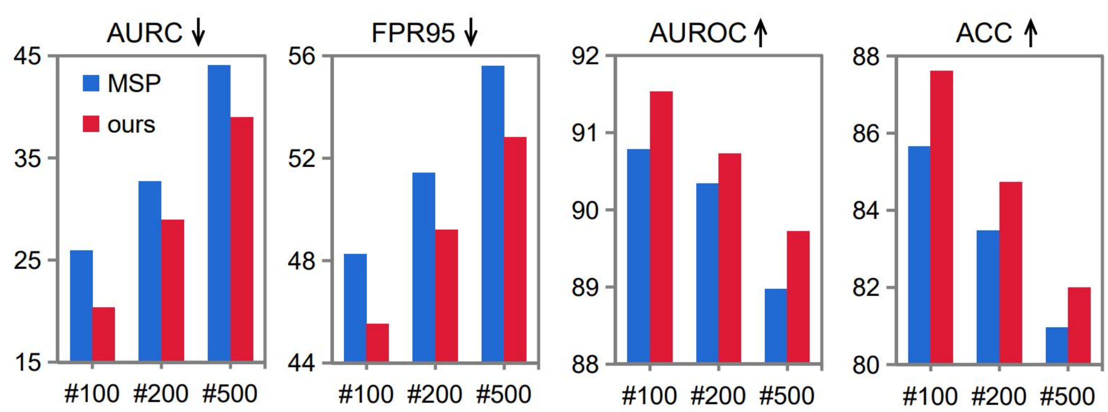
OpenMix: mix of the outlier and ID regions could reflect the property of low density regions, increasing the exposure of low density regions

Experiments

- Evaluation metrics: AURC, AUROC, FPR95, ACC

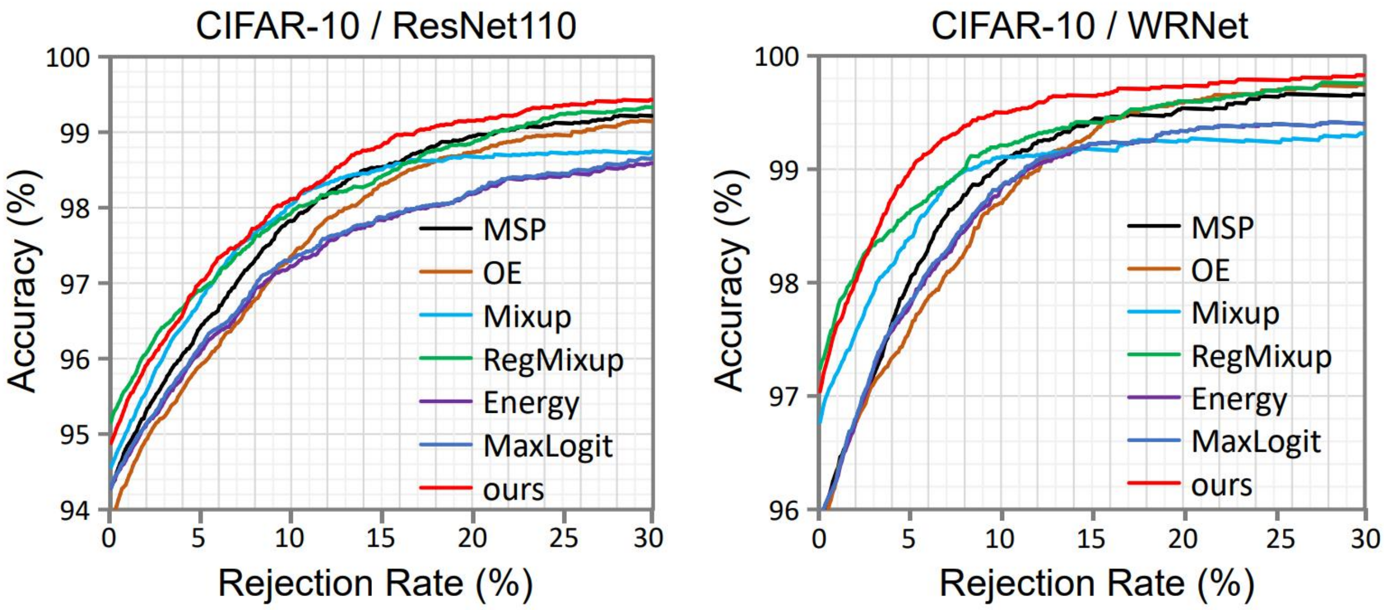
Network	Method	CIFAR-10				CIFAR-100			
		AURC ↓	AUROC ↑	FPR95 ↓	ACC ↑	AURC ↓	AUROC ↑	FPR95 ↓	ACC ↑
ResNet110	MSP [ICLR17] [25]	9.52±0.49	90.13±0.46	43.33±0.59	94.30±0.06	89.05±1.39	84.91±0.13	65.65±1.72	73.30±0.25
	Doctor [NeurIPS21] [19]	9.51±0.49	90.15±0.44	42.95±0.78	94.30±0.06	89.84±1.12	84.94±0.09	64.75±1.37	73.30±0.25
	ODIN [ICLR18] [38]	20.82±1.09	79.45±0.75	59.32±1.08	94.30±0.06	167.53±9.93	68.95±1.95	79.64±1.43	73.30±0.25
	Energy [NeurIPS20] [39]	15.13±0.85	84.72±0.80	53.89±0.65	94.30±0.06	128.66±5.05	76.80±1.07	73.54±0.73	73.30±0.25
	MaxLogit [ICML22] [23]	14.93±0.87	85.00±0.80	53.01±1.13	94.30±0.06	125.38±4.54	77.73±0.96	70.61±0.70	73.30±0.25
	LogitNorm [ICML22] [58]	12.57±1.32	88.82±0.84	56.27±2.61	92.64±0.23	118.00±3.17	79.56±0.16	73.09±0.18	71.68±0.34
	Mixup [NeurIPS18] [61]	16.27±1.33	86.21±0.83	40.71±0.88	94.69±0.31	87.39±1.83	84.60±0.88	64.95±3.28	75.08±0.30
	RegMixup [NeurIPS22] [48]	7.88±0.64	89.40±0.64	50.91±1.47	**95.10±0.23**	75.76±2.00	84.80±0.48	64.75±1.16	**76.15±0.14**
	OpenMix (ours)	**6.31±0.32**	**92.09±0.36**	**39.63±2.36**	94.98±0.20	**73.84±1.31**	**85.83±0.22**	**64.22±1.35**	75.77±0.35
WRNet	MSP [ICLR17] [25]	4.76±0.62	93.14±0.38	30.15±1.98	95.91±0.07	46.84±0.90	88.50±0.44	56.64±1.33	80.76±0.18
	Doctor [NeurIPS21] [19]	4.75±0.61	93.13±0.38	30.46±1.90	95.91±0.07	47.34±1.31	88.41±0.23	57.64±0.64	80.76±0.18
	ODIN [ICLR18] [38]	20.37±3.36	74.70±2.67	62.04±2.86	95.91±0.07	72.58±0.69	81.02±0.37	65.22±0.53	80.76±0.18
	Energy [NeurIPS20] [39]	6.91±0.66	90.47±0.51	39.13±2.07	95.91±0.07	57.30±1.24	85.05±0.34	64.15±0.26	80.76±0.18
	MaxLogit [ICML22] [23]	6.85±0.66	90.60±0.52	37.01±2.38	95.91±0.07	56.07±1.24	85.62±0.32	61.57±0.56	80.76±0.18
	LogitNorm [ICML22] [58]	5.81±0.45	91.06±0.26	46.06±2.24	95.50±0.33	72.05±1.32	82.23±0.28	66.32±0.11	79.11±0.09
	Mixup [NeurIPS18] [61]	5.30±2.02	90.79±2.64	29.68±3.26	96.71±0.05	46.91±2.43	87.61±0.46	56.05±2.50	82.51±0.18
	RegMixup [NeurIPS22] [48]	3.36±0.27	92.31±0.34	37.48±4.96	97.10±0.14	40.36±1.71	88.33±0.35	56.44±0.95	82.50±0.30
	OpenMix (ours)	**2.32±0.15**	**94.81±0.34**	**22.08±1.86**	**97.16±0.10**	**39.61±0.54**	**89.06±0.11**	**55.00±1.29**	**82.63±0.06**

- OOD detection methods failed in detecting misclassification errors
- OpenMix improves the reliability of confidence
- Large-scale experiments on ImageNet



Experiments

- Accuracy-rejection curves analysis



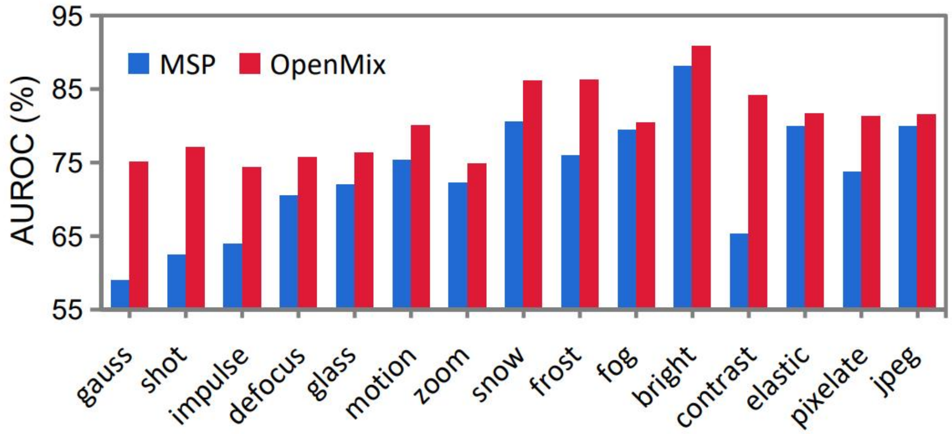
① If the desired accuracy is known, select the model with the lowest rejection rate

② If the acceptable rejection rate is known, select the model with the highest accuracy

our method as the best in both cases

Experiments

- MisD under distribution shift



- MisD in long-tailed recognitzion

Method	CIFAR-10-LT				CIFAR-100-LT			
	AURC	AUROC	FPR95	ACC	AURC	AUROC	FPR95	ACC
LA [42]	62.13	84.52	69.77	79.02	347.43	78.46	76.47	41.69
+ CRL	63.81	85.30	63.05	78.50	345.05	78.74	76.19	41.58
+ ours	**38.07**	**87.21**	**64.14**	**83.60**	**284.77**	**81.22**	**73.80**	**46.52**
VS [33]	58.45	84.47	70.15	80.11	343.48	78.20	77.25	42.20
+ CRL	62.06	83.98	67.19	79.69	345.06	78.29	77.44	41.88
+ ours	**41.52**	**87.12**	**63.31**	**83.02**	**277.34**	**81.42**	**72.93**	**47.16**

Experiments

- OpenMix improves OOD detection: averaged over six OOD test datasets

Method	FPR95 ↓			AUROC ↑			AUPR ↑		
	ResNet	WRN	DenseNet	ResNet	WRN	DenseNet	ResNet	WRN	DenseNet
	ID: CIFAR-10								
MSP [5]	51.69	40.83	48.60	89.85	92.32	91.55	97.42	97.93	98.11
LogitNorm [14]	29.72	12.97	19.72	94.29	97.47	96.19	98.70	99.47	99.11
ODIN [8]	35.04	26.94	30.67	91.09	93.35	93.40	97.47	97.98	98.30
Energy [9]	33.98	25.48	30.01	91.15	93.58	93.45	97.49	98.00	98.35
MaxLogit [3]	34.61	26.72	30.99	91.13	93.14	93.44	97.46	97.78	98.35
OE [6]	**5.28**	**3.49**	**5.25**	**98.04**	**98.59**	**98.20**	**99.55**	**99.71**	**99.62**
CRL [10]	51.18	40.83	47.28	91.21	93.67	92.37	98.11	98.67	98.35
FMFP [17]	39.50	26.83	35.12	93.83	96.22	94.88	98.73	99.23	98.95
OpenMix (ours)	39.72	16.86	32.75	93.22	96.92	94.85	98.46	99.34	98.84
	ID: CIFAR-100								
MSP [5]	81.68	77.53	77.03	74.21	77.96	76.79	93.34	94.36	93.94
LogitNorm [14]	63.49	57.38	61.56	82.50	86.60	82.10	95.43	96.80	95.16
ODIN [8]	74.30	76.03	69.44	76.55	79.57	80.53	93.54	94.59	94.78
Energy [9]	74.42	74.93	68.36	76.43	79.89	80.87	93.59	94.66	94.86
MaxLogit [3]	74.45	75.27	69.85	76.61	79.75	80.48	93.66	94.67	94.77
OE [6]	**59.85**	**49.02**	**53.03**	**86.33**	**90.07**	**88.51**	**96.47**	**97.67**	**97.25**
CRL [10]	81.67	79.08	75.77	72.72	76.81	76.41	92.69	94.22	93.85
FMFP [17]	80.19	70.98	72.87	72.92	81.54	77.56	92.94	95.71	94.19
OpenMix (ours)	74.66	68.87	66.63	75.95	84.88	81.23	93.56	96.55	95.30

Conclusive Remarks

- Misclassification detection is an important yet far less explored problem for safety-critical applications
- We propose to leverage outlier data for overconfidence mitigation
- We analysize why OE is harmful for MisD from the perspective of feature space uniformity
- The proposed OpenMix is a unified method for MisD and OOD detection

- Code is available at https://github.com/Impression2805/OpenMix
- Useful paper list https://github.com/Impression2805/Awesome-Failure-Detection

OpenMix

Failure Detection

Thanks for your attention!