南京航空航天大學
Nanjing University of Aeronautics and Astronautics

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

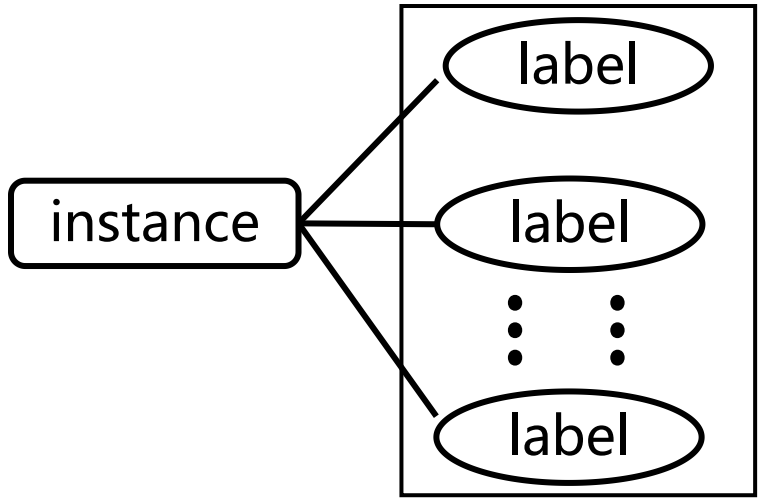
Can multi-label classification networks know what they don't know?

NeurIPS 2021

Multi-label Classification

- **ordinary supervised learning** vs. **multi-label learning**

Ordinary **supervised Learning**
(only one ground-truth label)



Multi-label learning
(a series of labels)

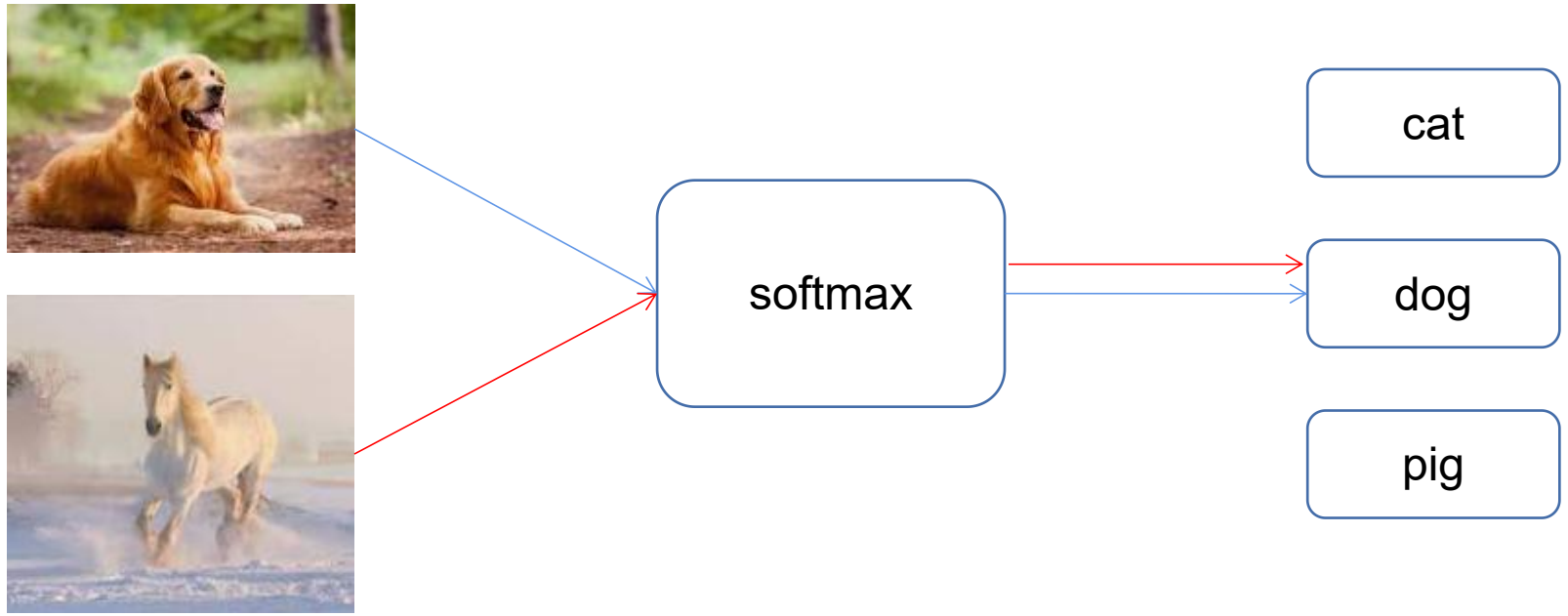
Multi-label Classification

cloud

sky

leaf

Out-of-distribution Detection （OOD）



OOD <- no label in the in-distribution data

decision function G：

$$G(\mathbf{x}; f) = \begin{cases} 0 & \text{if } \mathbf{x} \sim \mathcal{D}_{\text{out}}, \\ 1 & \text{if } \mathbf{x} \sim \mathcal{D}_{\text{in}}. \end{cases}$$

Out-of-distribution Detection

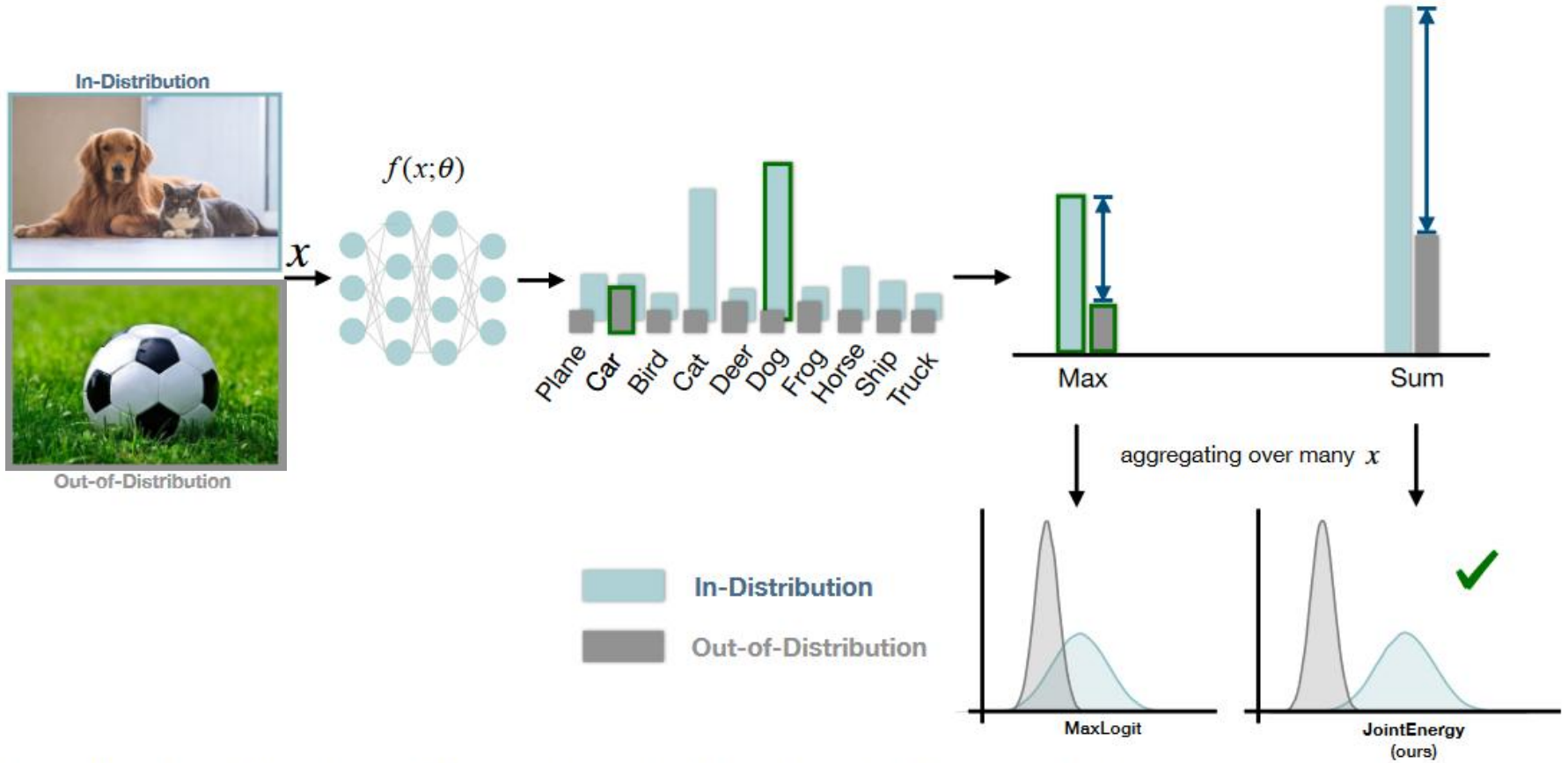


Figure 1: Out-of-distribution detection for multi-label classification networks. During inference time, input $\mathbf{x}$ is passed through classifier f, and label-wise scores are computed for each label. OOD indicator scores are either the maximum-valued score (denoted by green outlines) or the sum of all scores. Taking the sum results in a larger difference in scores and more separation between in-distribution and OOD inputs (denoted by red lines), resulting in better OOD detection. Plots in the bottom right depict the probability densities of MaxLogit [13] versus *JointEnergy* (ours).

Problem

OOD uncertainty estimation in the multi-label classification setting

estimate the joint likelihood using generative models

combine label-wise energy over all labels

Energy Function

a scoring function for OOD uncertainty estimation in the multi-class setting

$$f(\mathbf{x}) : \mathcal{X} \to \mathbb{R}^K$$

logits (softmax)

$$p(y_i = 1 \mid \mathbf{x}) = \frac{e^{f_{y_i}(\mathbf{x})}}{\sum_{j=1}^{K} e^{f_{y_j}(\mathbf{x})}}. \qquad (1)$$

by the Boltzmann distribution

probability distribution :

$$p(y_i = 1 \mid \mathbf{x}) = \frac{e^{-E(\mathbf{x},y_i)}}{\int_{y'} e^{-E(\mathbf{x},y')}} = \frac{e^{-E(\mathbf{x},y_i)}}{e^{-E(\mathbf{x})}}.$$

Energy Function

from a energy-based perspective by viewing **logit** as **energy function** :

$$E(\mathbf{x}, y_i) = -f_{y_i}(\mathbf{x})$$

$$p(y_i = 1 \mid \mathbf{x}) = \frac{e^{f_{y_i}(\mathbf{x})}}{\sum_{j=1}^{K} e^{f_{y_j}(\mathbf{x})}}. \quad = \quad p(y_i = 1 \mid \mathbf{x}) = \frac{e^{-E(\mathbf{x}, y_i)}}{\int_{y'} e^{-E(\mathbf{x}, y')}} = \frac{e^{-E(\mathbf{x}, y_i)}}{e^{-E(\mathbf{x})}}.$$

free energy function :

$$E(\mathbf{x}) = -\log \sum_{i=1}^{K} e^{f_{y_i}(\mathbf{x})}. \qquad (2)$$

Label-wise Free Energy

a standard pre-trained multi-label neural classifier

a logit output for the i-th class :

$$f_{y_i}(\mathbf{x}) = h(\mathbf{x};\theta)\cdot \mathbf{w}^i_{\text{cls}},$$

predictive probabilty :

$$p(y_i = 1 \mid \mathbf{x}) = \frac{e^{f_{y_i}(\mathbf{x})}}{1 + e^{f_{y_i}(\mathbf{x})}},$$

label-wise free energy :

$$E_{y_i}(\mathbf{x}) = -\log(1 + e^{f_{y_i}(\mathbf{x})}),$$

Unfortunately does not capture uncertainty jointly across labels

Label-wise Free Energy

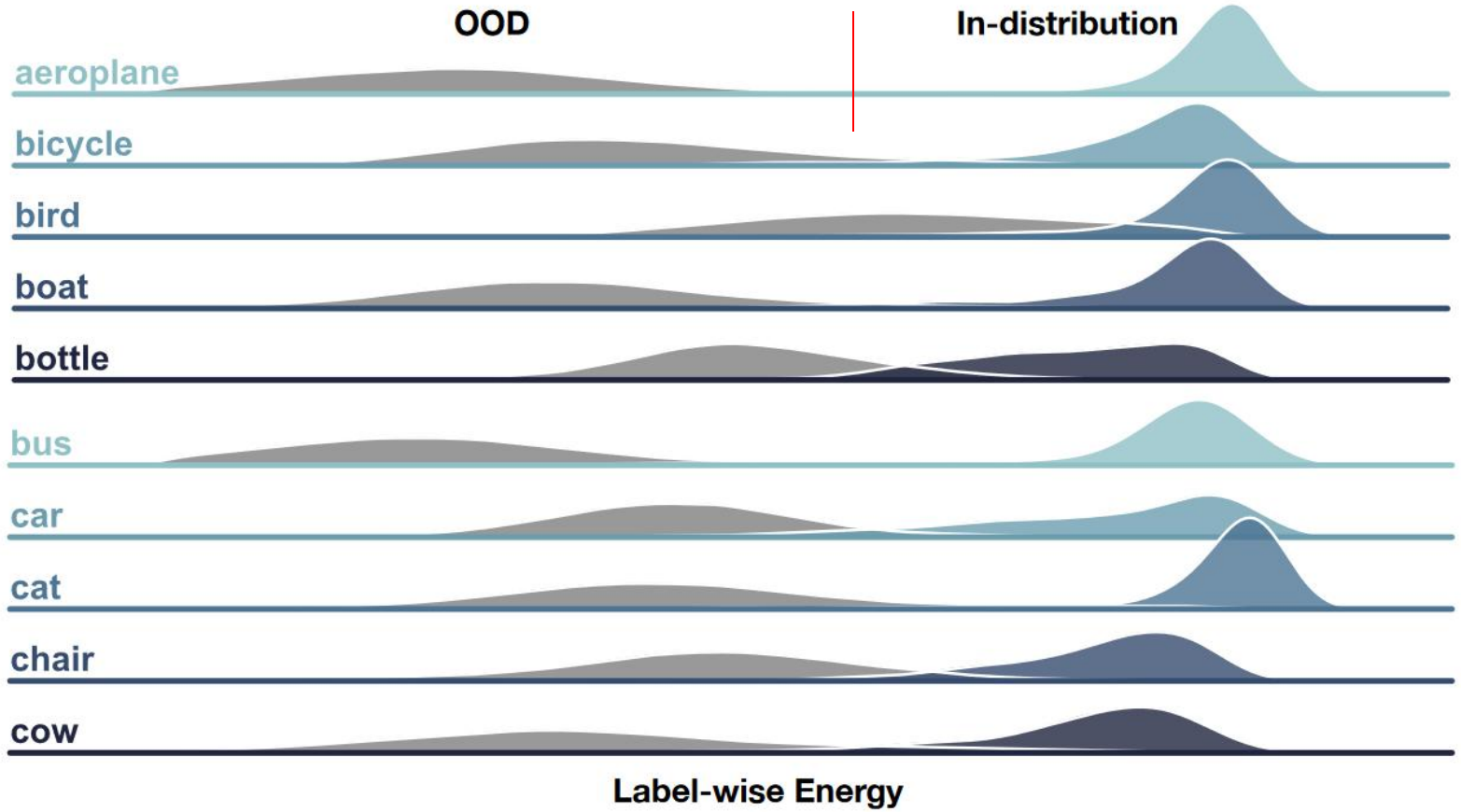


Figure 2: Label-wise energy scores $-E_{y_i}(\mathbf{x})$ distribution. The in-distribution classes (each per row) are a subset from PASCAL-VOC (green). OOD test data is from ImageNet (gray), which is the same for all labels. x-axis is in log scale for visibility.

Unfortunately does not capture uncertainty jointly across labels

JointEnergy

consider joint uncertainty across labels :

$$E_{\text{joint}}(\mathbf{x}) = \sum_{i=1}^{K} -E_{y_i}(\mathbf{x})$$

for OOD

$$G(\mathbf{x};\tau) = \begin{cases} \text{out} & \text{if } E_{\text{joint}}(\mathbf{x}) \leq \tau, \\ \text{in} & \text{if } E_{\text{joint}}(\mathbf{x}) > \tau, \end{cases}$$

→ energy threshold

Dataset

in-distribution

MS-COCO ,
PASCAL-VOC,
NUS-WIDE

To evaluate the models trained on the in-distribution datasets above, **OOD** :

ImageNet
Textures datasets

Table 3: Ablation study on the effect of aggregation methods: max vs summation. Values are AUROC.

$\mathcal{D}_{\text{in}}$	MaxEnergy	JointEnergy
MS-COCO	89.11	**92.70**
PASCAL-VOC	89.22	**91.10**
NUS-WIDE	83.58	**88.30**

Results

The sensitivity analysis on τ

the **larger** AOC, the **better**

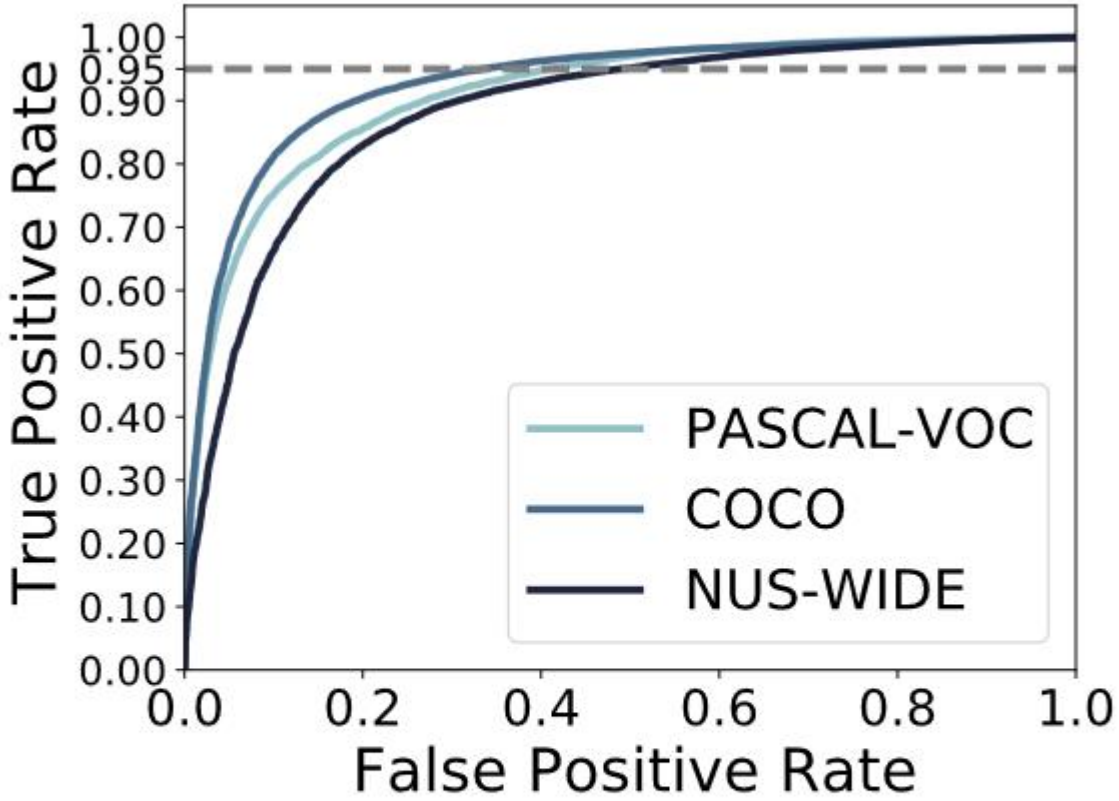


Figure 3: AUROC curves for OOD detector obtained from three in-distribution multi-label classification datasets.

Results

Table 1: OOD detection performance comparison using JointEnergy vs. competitive baselines. We use DenseNet [19] to train on the in-distribution datasets. We use a subset of ImageNet classes as OOD test data, as described in Section 4.1. All values are percentages. ↑ indicates larger values are better, and ↓ indicates smaller values are better. **Bold** numbers are superior results. Description of baseline methods, additional evaluation results on different OOD test data, and different architecture (*e.g.*, ResNet [14]) can be found in the Appendix.

$\mathcal{D}_{\text{in}}$	MS-COCO	PASCAL-VOC	NUS-WIDE
OOD Score	**FPR95 ↓ / AUROC ↑ / AUPR ↑**		
MaxLogit [15]	43.53 / 89.11 / 93.74	45.06 / 89.22 / 83.14	56.46 / 83.58 / 94.32
MSP [16]	79.90 / 73.70 / 85.37	74.05 / 79.32 / 72.54	88.50 / 60.81 / 87.00
ODIN [28]	43.53 / 89.11 / 93.74	45.06 / 89.22 / 83.16	56.46 / 83.58 / 94.32
Mahalanobis [27]	46.86 / 88.59 / 93.85	41.74 / 88.65 / 81.12	62.67 / 84.02 / 95.25
LOF [3]	80.44 / 73.95 / 86.01	86.34 / 69.21 / 58.93	85.21 / 67.75 / 89.61
Isolation Forest [31]	94.39 / 49.04 / 66.87	93.22 / 50.67 / 35.78	95.69 / 53.12 / 83.32
JointEnergy	**33.48 / 92.70 / 96.25**	**41.01 / 91.10 / 86.33**	**48.98 / 88.30 / 96.40**

Results

Table 2: Ablation study on the effect of summation for prior approaches. We use DenseNet [19] to train on the in-distribution datasets. We use ImageNet as OOD test data as described in Section 4.1. Note that *Sum* does not apply to tree-based or KNN-based approaches (e.g., LOF and Isolation Forest).

	$\mathcal{D}_{\text{in}}$	MS-COCO	PASCAL	NUS-WIDE
		FPR95 / AUROC / AUPR		
OOD Score	**Aggregation**	↓ / ↑ / ↑		
Logit	Sum	95.46 / 61.81 / 80.39	87.18 / 72.68 / 61.24	96.53 / 51.75 / 82.55
Prob	Sum	45.04 / 89.32 / 94.40	**38.57** / 86.53 / 79.10	50.84 / 83.82 / 95.15
ODIN	Sum	56.56 / 84.62 / 92.24	50.35 / 79.45 / 70.19	56.26 / 81.04 / 94.34
Mahalanobis	Sum	53.43 / 87.52 / 93.35	44.43 / 87.76 / 79.86	69.05 / 80.46 / 94.09
LOF	Sum	N/A	N/A	N/A
Isolation Forest	Sum	N/A	N/A	N/A
JointEnergy (ours)	Sum	**33.48 / 92.70 / 96.25**	41.01 / **91.10 / 86.33**	**48.98 / 88.30 / 96.40**

Thanks