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

Open Set Learning with Counterfactual Images

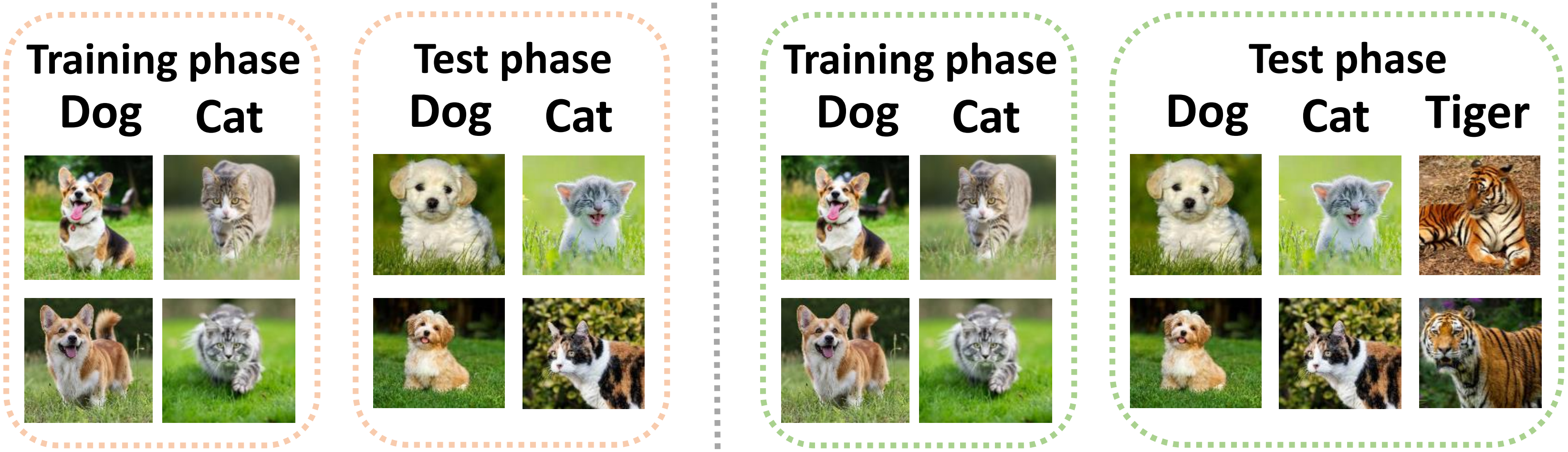
Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, Fuxin Li

Collaborative Robotics and Intelligent Systems Institute
Oregon State University

ECCV 2018

Open Set Recognition

- **Open set classification** vs. closed set classification



Closed set classification

Open set classification

Motivation

- Thresholding

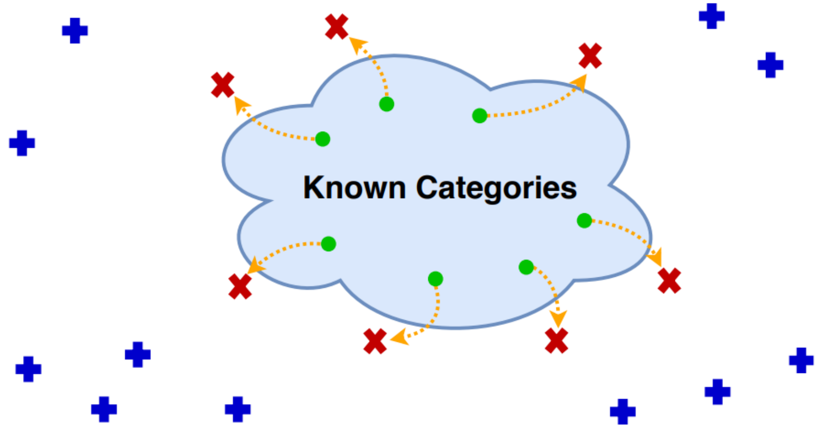
$$y^* = \begin{cases} \arg\max_{y_i} P(y_i|x) & \text{if } \max_{y_i} P(y_i|x) > \delta \\ unknown & \text{else} \end{cases}$$

- Potential limitation
  - It is difficult to find the optimal **global** threshold.
  - DNNs can output **incorrect high-confidence predictions** when faced with test data from outside the training distribution.

- OpenMax
  - Using Extreme Value Theory (EVT) to calibrate the model outputs.

Main Idea

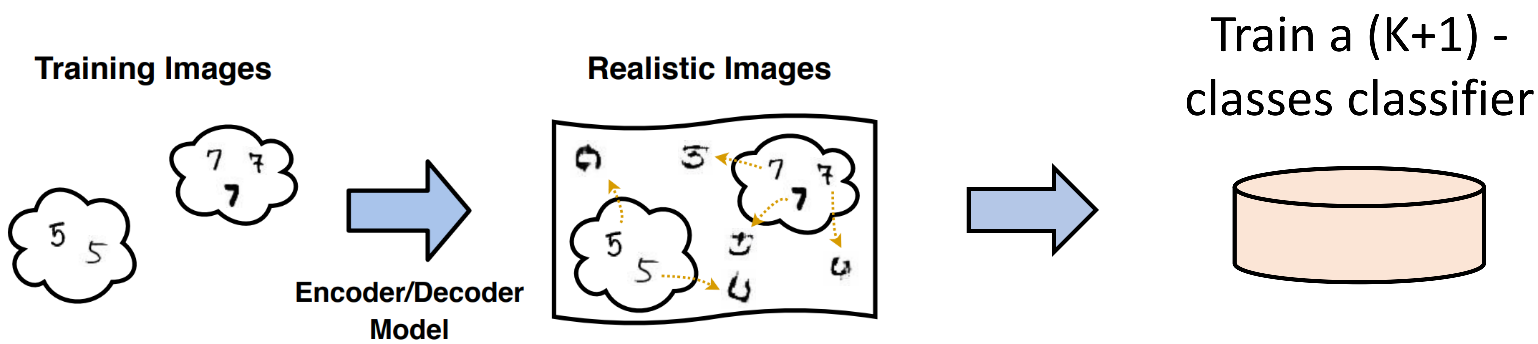
- Counterfactual images



$$\begin{aligned} \text{minimize} \quad & ||z_0 - z^*||_2 \\ \text{subject to} \quad & C_p(z^*) = 1 \end{aligned}$$

The Proposed Method

- Overview



- The generative model

$$\mathbf{L_D} = \sum_{x \in \mathbf{X}} D(G(E(x))) - D(x) + P(D)$$

$$\mathbf{L_G} = \sum_{x \in \mathbf{X}} ||x - G(E(x))||_1 - D(G(E(x)))$$

The Proposed Method

- Generating Counterfactual Open Set Examples

$$\begin{array}{ll} \text{minimize} & ||E(x) - z^*||_2 \\ \text{subject to} & G(z^*) \text{ is an open set example} \end{array}$$

$$z^* = \min_z ||z - E(x)||_2^2 + \log\left(1 + \sum_{i=1}^{K} \exp C_K(G(z))_i\right)$$

  - Randomly sample input seed image x from training set.
  - Each counterfactual image $G(z^*)$ is augmented to the training set with class label $K+1$.

Experiments

- Open set detection

Method	CIFAR-10	SVHN	MNIST
Softmax Threshold	$.677 \pm .038$	$.886 \pm .014$	$.978 \pm .006$
OpenMax	$.695 \pm .044$	$.894 \pm .013$	$.981 \pm .005$
G-OpenMax*	$.675 \pm .044$	$.896 \pm .017$	$.984 \pm .005$
Ours	$\mathbf{.699 \pm .038}$	$\mathbf{.910 \pm .010}$	$\mathbf{.988 \pm .004}$

- Closed set accuracy

Method	CIFAR-10	SVHN	MNIST
Softmax/OpenMax	$.801 \pm .032$	$.947 \pm .006$	$.995 \pm .002$
G-OpenMax*	$.816 \pm .035$	$.948 \pm .008$	$.996 \pm .001$
Ours	$\mathbf{.821 \pm .029}$	$\mathbf{.951 \pm .006}$	$.996 \pm .001$

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

Learning Placeholders for Open-Set Recognition

Da-Wei Zhou Han-Jia Ye† De-Chuan Zhan
State Key Laboratory for Novel Software Technology, Nanjing University
{zhoudw, yehj}@lamda.nju.edu.cn, zhandc@nju.edu.cn

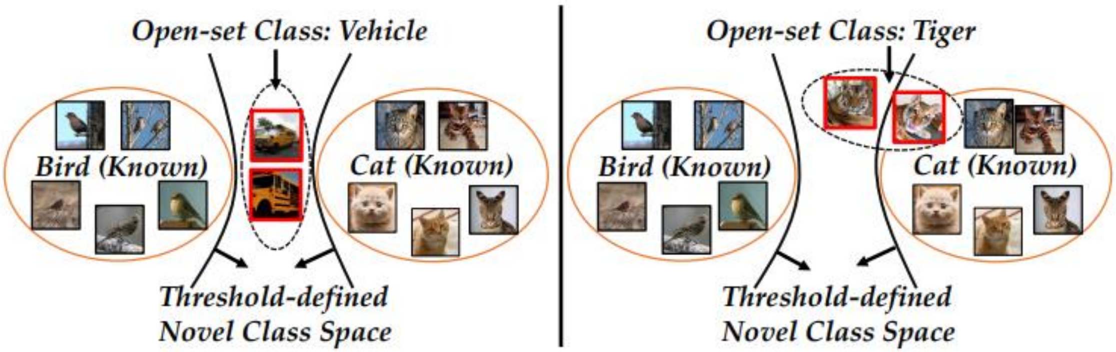
CVPR 2021

Motivation

- Thresholding

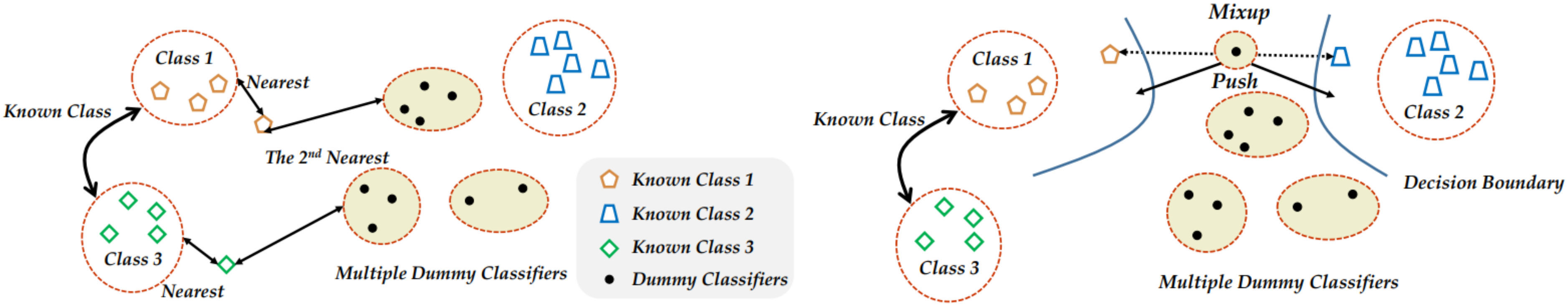
$$y^* = \begin{cases} \arg\max_{y_i} P(y_i|x) & \text{if } \max_{y_i} P(y_i|x) > \delta \\ unknown & \text{else} \end{cases}$$

- Potential limitation
  - It is difficult to find the optimal **global** threshold.
  - DNNs can output incorrect high-confidence predictions when faced with test data from outside the training distribution.



Learning placeholders for OS recognition

- Classifier placeholder
  - Augmenting the closed-set classifier with a **dummy classifier**, which adaptively outputs the **class-specific threshold** to separate known and unknown
- Data placeholder
  - Mimicking the emergence of novel classes, and transforming open-set training into closed-set training



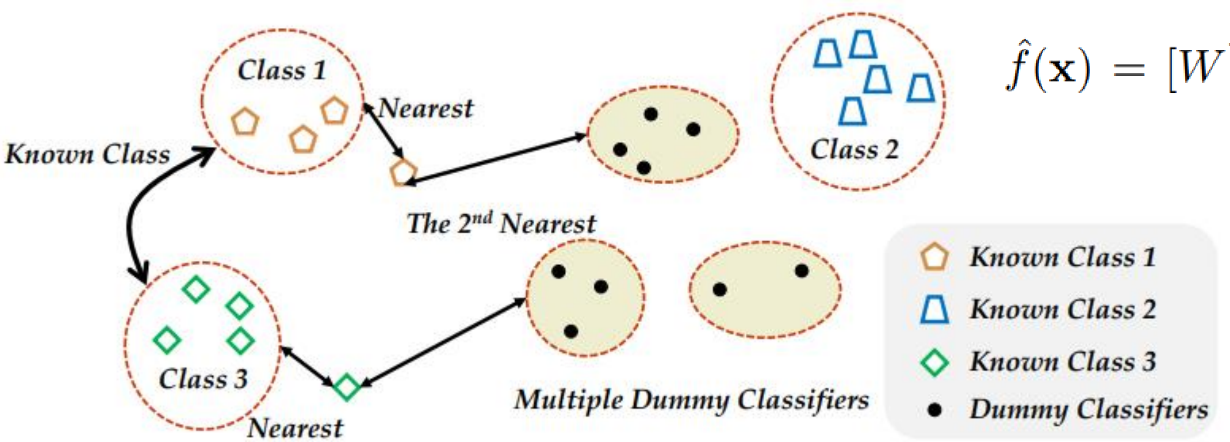
(a) Learning classifier placeholders (b) Learning data placeholders

Learning placeholders for OS recognition

- Classifier placeholder
  - A dummy classifier $\hat{f}(\mathbf{x}) = [W^\top \phi(\mathbf{x}), \boxed{\hat{\boldsymbol{w}}^\top \phi(\mathbf{x})}]$
  - Making the dummy classifier to output the **second-largest** probability for known instances.

$$l_1 = \sum_{(\mathbf{x},y)\in\mathcal{D}_{tr}} \ell(\hat{f}(\mathbf{x}), y) + \beta\ell(\boxed{\hat{f}(\mathbf{x}) \setminus y}, K+1)$$

Removing the probability of ground truth label.



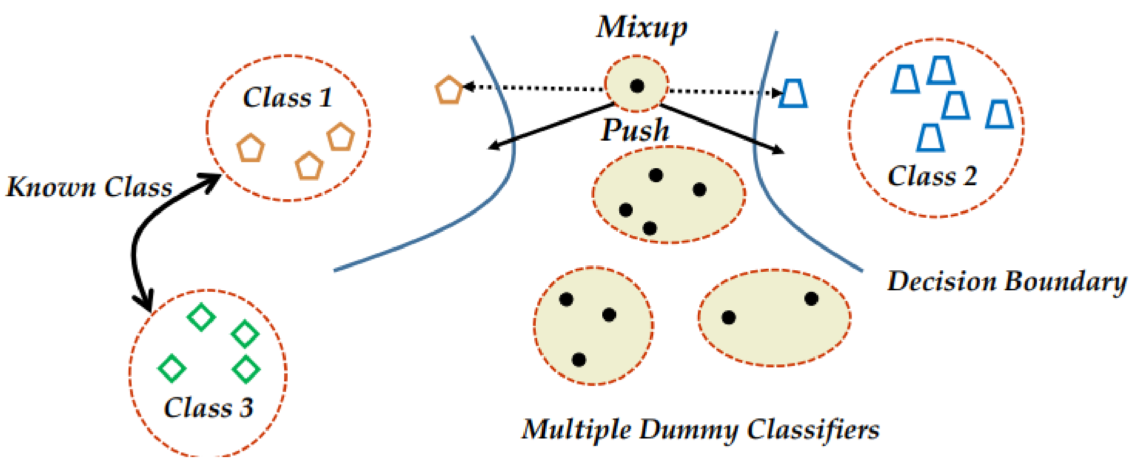
$$\hat{f}(\mathbf{x}) = [W^\top \phi(\mathbf{x}), \max_{k=1,\cdots,C} \hat{\boldsymbol{w}}_k^\top \phi(\mathbf{x})],$$

Learning placeholders for OS recognition

- Data placeholder
  - The distribution of these instances seems novel, and the generating process should be quick.
  - Manifold mixup

$$\tilde{\mathbf{x}}_{pre} = \lambda\phi_{pre}(\mathbf{x}_i) + (1-\lambda)\phi_{pre}(\mathbf{x}_j), y_i \neq y_j$$

$$l_2 = \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}_{tr}} \ell([W, \hat{\boldsymbol{w}}]^\top \phi_{post}(\tilde{\mathbf{x}}_{pre}), K+1)$$



Experiments

- Open set detection

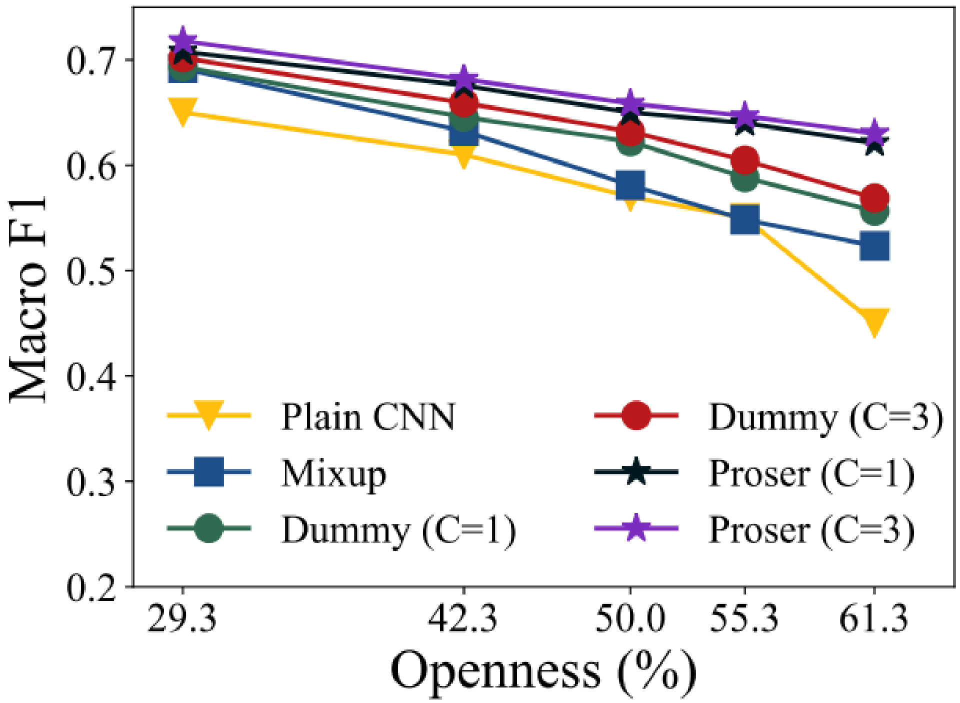
Methods	SVHN	CIFAR10	CIFAR+10	CIFAR+50	Tiny-ImageNet
Softmax	88.6	67.7	81.6	80.5	57.7
OpenMax [3]	89.4	69.5	81.7	79.6	57.6
G-OpenMax [6]	89.6	67.5	82.7	81.9	58.0
OSRCI [22]	91.0	69.9	83.8	82.7	58.6
C2AE [24]	89.2	71.1	81.0	80.3	58.1
GFROSR [26]	93.5	83.1	91.5	91.3	64.7
PROSER	**94.3**	**89.1**	**96.0**	**95.3**	**69.3**

- Closed set accuracy

Methods	SVHN	CIFAR10	Tiny-ImageNet
Plain CNN	96.5	92.8	52.2
PROSER	96.4	92.6	52.1

Experiments

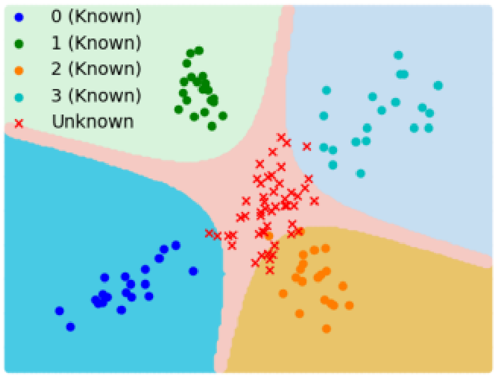
- Ablation studies



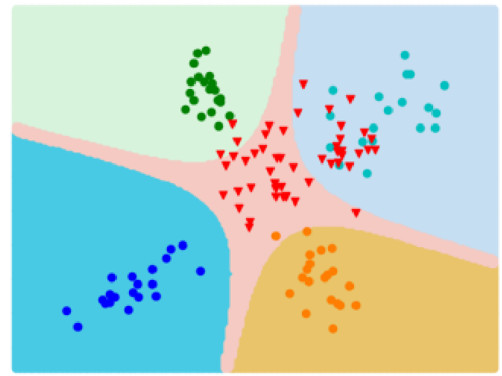
$$\text{Openness} = 1 - \sqrt{\frac{N_{train}}{N_{test}}}$$

Experiments

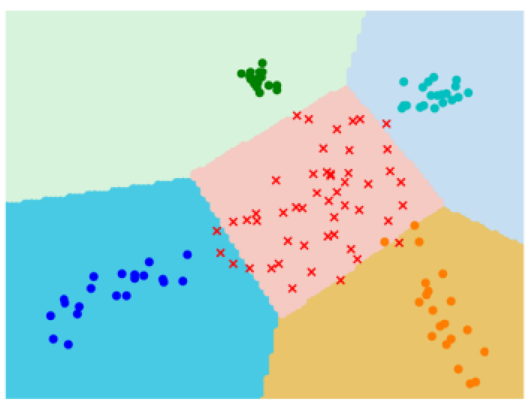
- Visualization of Decision Boundaries



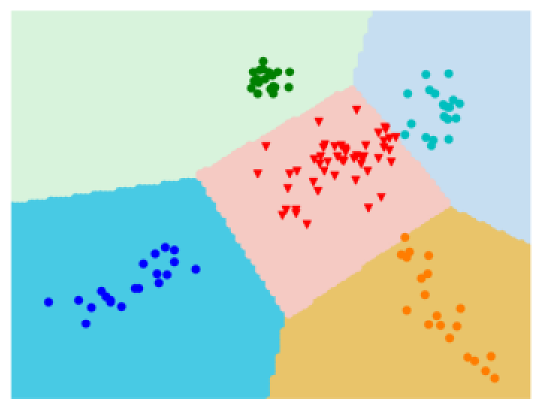
(a) Thresholding when Class 4 is novel



(b) Thresholding when Class 5 is novel



(c) PROSER when Class 4 is novel



(d) PROSER when Class 5 is novel

THANKS