Bidirectional Uncertainty-Based Active Learning for Open-Set Annotation

Chen-Chen Zong[1], Ye-Wen Wang[1], Kun-Peng Ning[2], Hai-Bo Ye[1], Sheng-Jun Huang[1*]

[1]Nanjing University of Aeronautics and Astronautics
[2]Peking University

*huangsj@nuaa.edu.cn

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

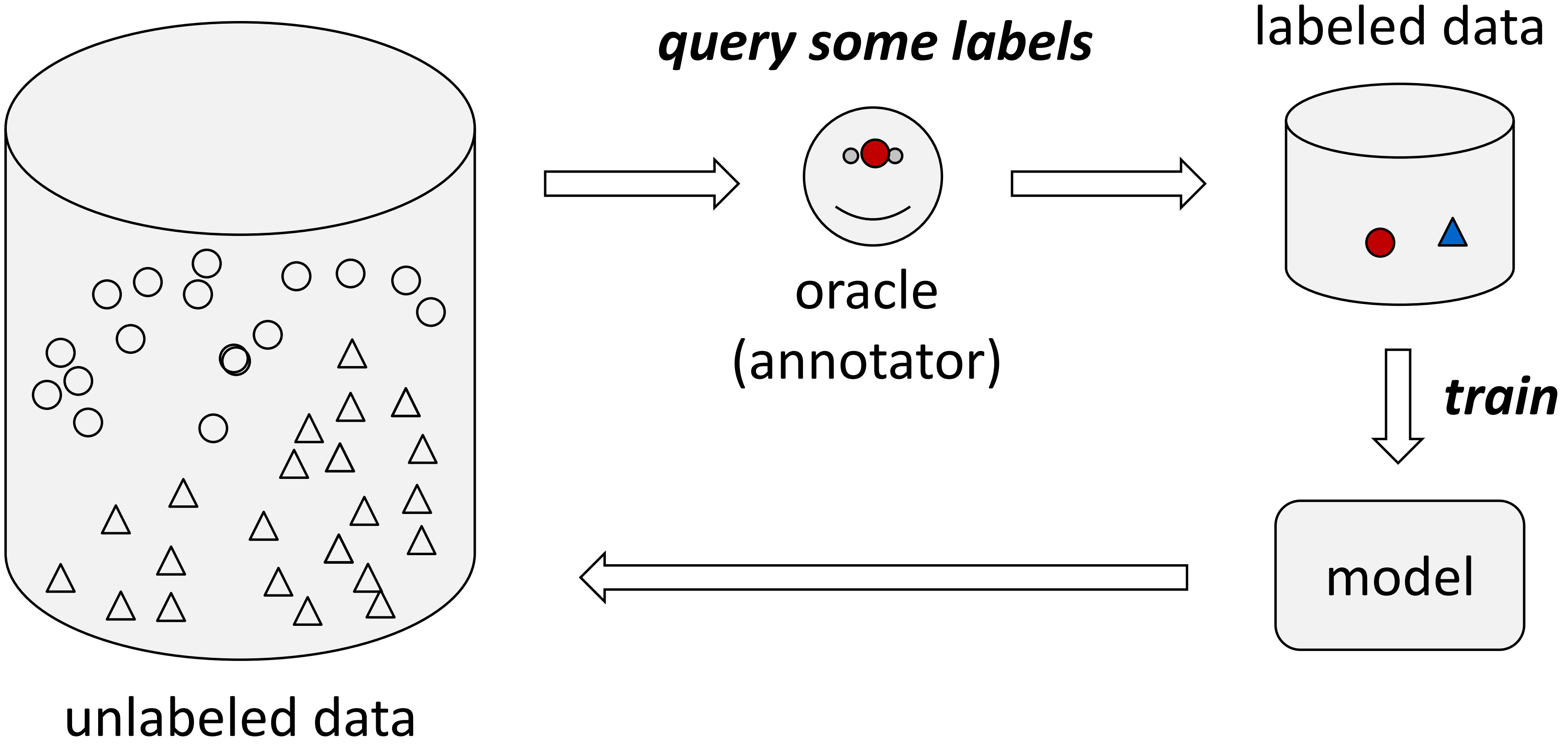
Outline

- Motivation
- The Framework of BUAL
- Random Label Negative Learning
- Bidirectional Sampling Strategy
- Experiments

Outline

- Motivation
- The Framework of BUAL
- Random Label Negative Learning
- Bidirectional Sampling Strategy
- Experiments

Active Learning

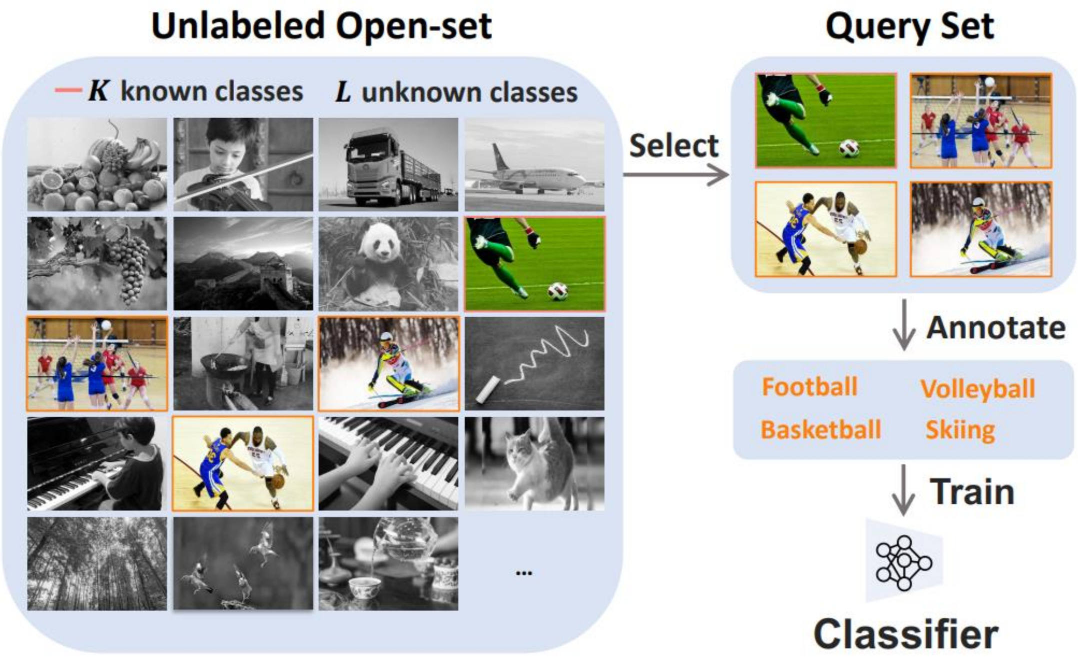


Goal: train an effective model with least labeling cost

Open-Set Annotation

Known class:
color images with border

Unknown class:
gray-scale images without border



- Active learning in open set scenarios presents a novel challenge of identifying the most valuable examples in an unlabeled data pool that comprises data from both known and unknown classes.

Motivation

- Traditional active learning methods prioritize selecting informative examples with low confidence, with the risk of *mistakenly selecting unknown-class examples* with similarly low confidence.

- Recent open-set annotation methods favor the most probable known class examples, with the risk of *picking simple already mastered examples*.

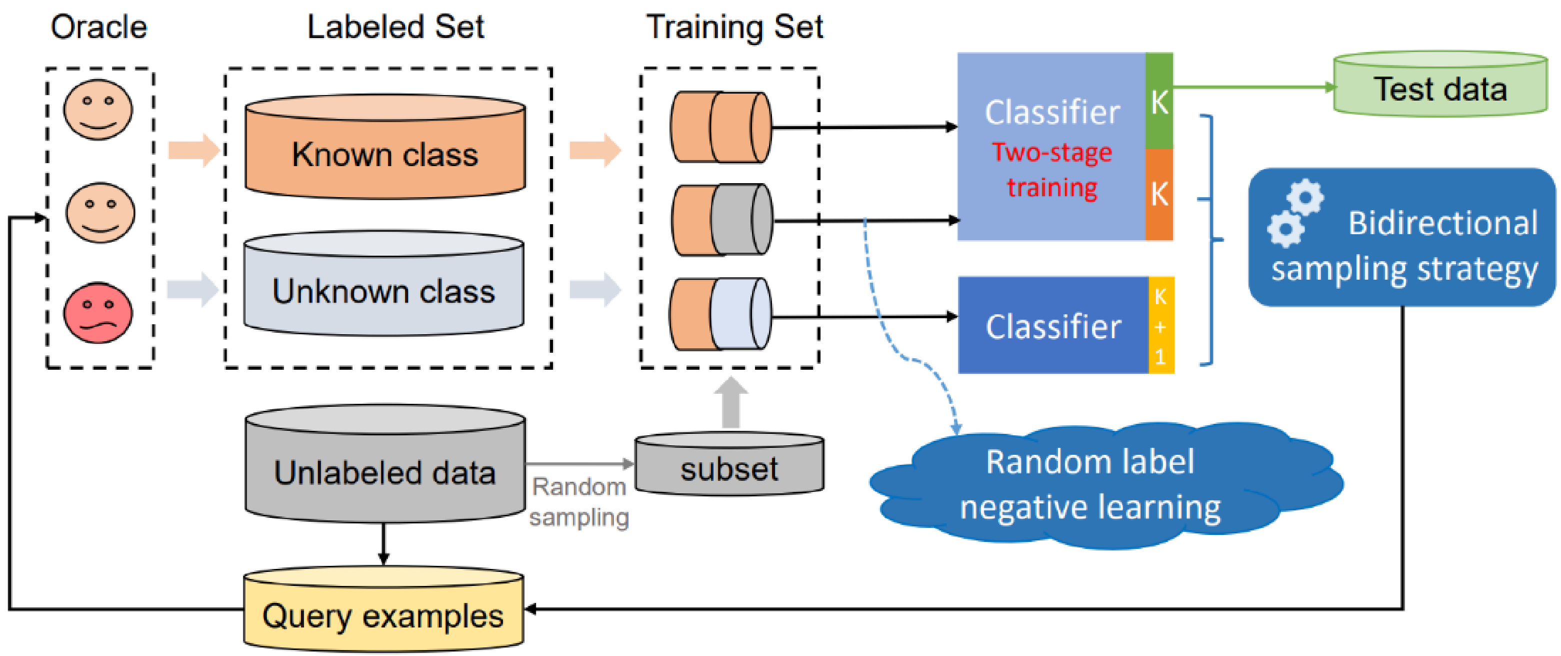
Can we effectively distinguish the "informative" examples of known classes from examples of unknown classes?

Outline

- Motivation
- The Framework of BUAL
- Random Label Negative Learning
- Bidirectional Sampling Strategy
- Experiments

The Framework of BUAL

BUAL: Bidirectional Uncertainty-based Active Learning framework



☐ Three components

- Model training
- Example selection
- Oracle labeling

☐ Two core contributions

- Random label negative learning (RLNL)
- Bidirectional sampling strategy

Outline

- Motivation
- The Framework of BUAL
- Random Label Negative Learning
- Bidirectional Sampling Strategy
- Experiments

Random Label Negative Learning

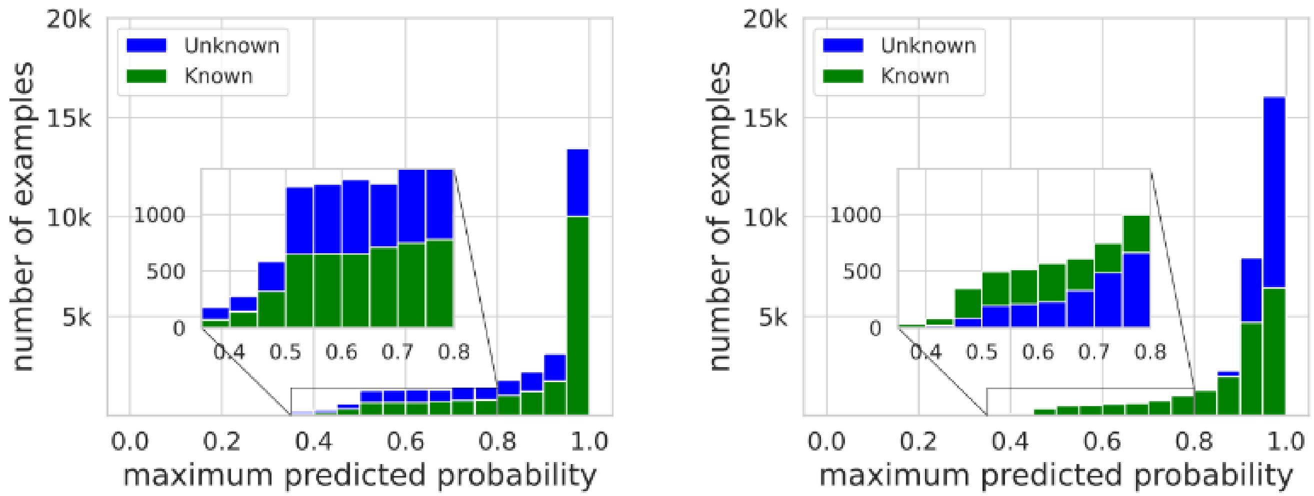
Can we effectively distinguish the "informative" examples of known classes from examples of unknown classes?

⇩

Pushing unknown class examples toward regions with high-confidence predictions.

Random Label Negative Learning

We achieve this!!!



(a) w/o NL (b) w/ NL

- ✓ We *randomly assign labels* to unlabeled examples in each training round and finetune the target model using the *negative learning* loss performed on them and already labeled examples.

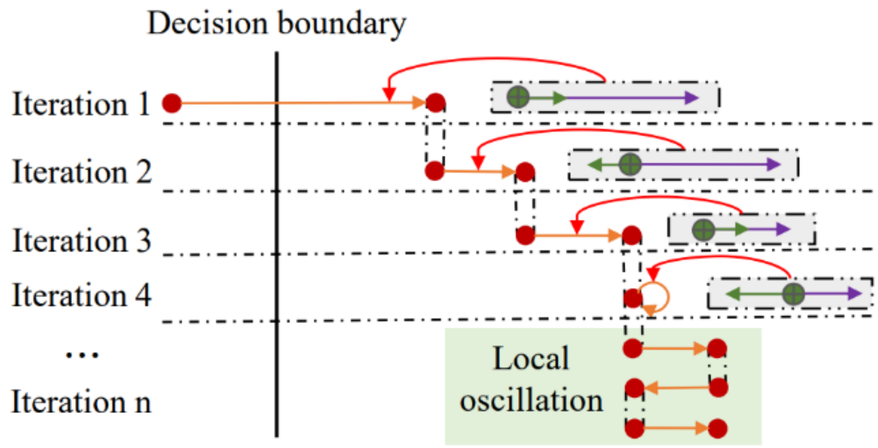
$$\ell_{NL}(f,\bar{y}) = -\sum_{k=1}^{K} \bar{y}_k \log(1-p_k)$$

Why RLNL works?

- ✓ Once unlabeled known class examples receive the correct labels, they suffer a larger penalty and are reduced confidence predictions by the model since they deviate from the distribution information obtained from labeled data.

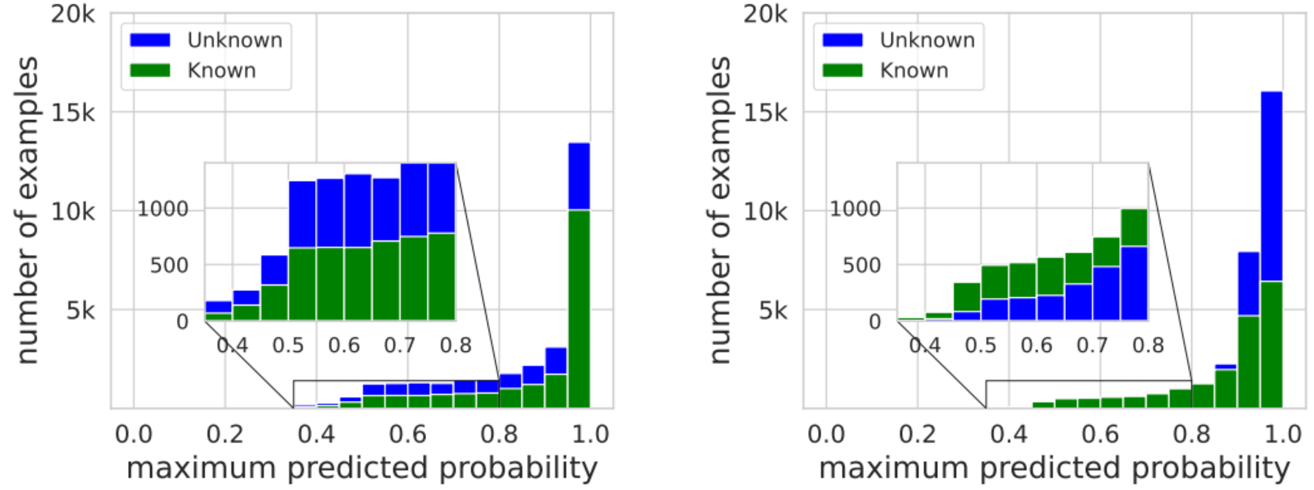
- ✓ In contrast, unlabeled unknown class examples will oscillate at uncharted away from the decision boundary to counteract the update gradient produced by the labeled data.

purple -->: gradient by labeled examples	green -->: gradient by the example itself

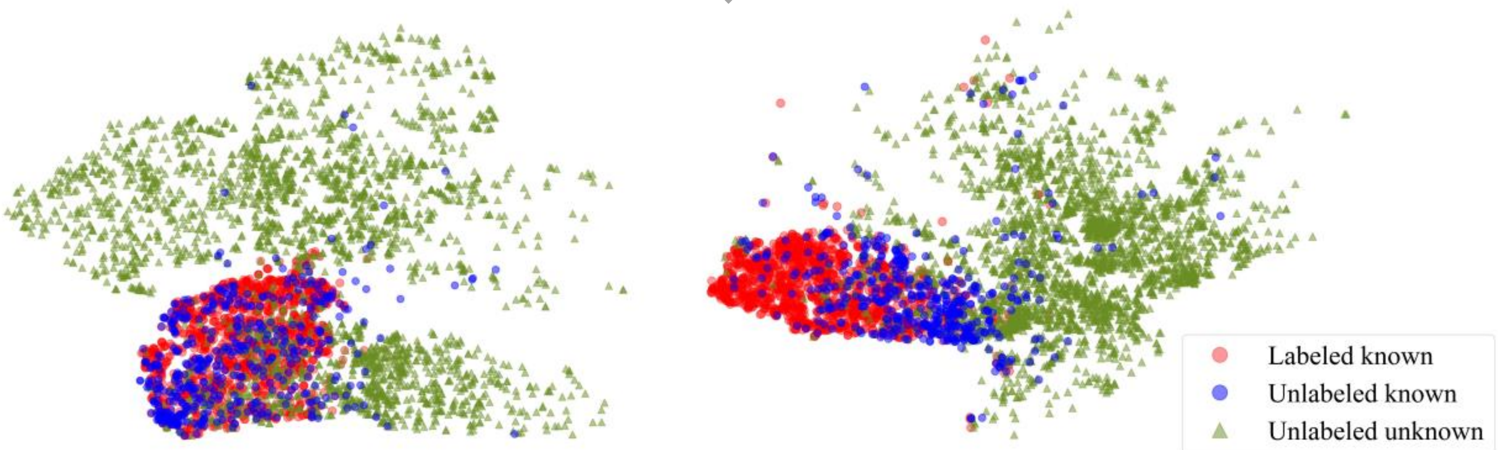


Possible update scenario for unlabeled unknown class examples

Why RLNL works?



(a) w/o NL **(b)** w/ NL



(a) Before RLNL **(b)** After RLNL

Outline

- Motivation
- The Framework of BUAL
- Random Label Negative Learning
- Bidirectional Sampling Strategy
- Experiments

Bidirectional Sampling Strategy

☐ Design Criteria

- The negative head is slightly biased for the measurement due to the unstable training. Thus, if an example is likelier to belong to known classes, we prefer to utilize the sample uncertainty obtained from positive head.

- Once an example has a higher risk of belonging to the unknown classes, the uncertainty obtained from positive head is unreliable, and thus the uncertainty produced by the negative head should be given a higher weight.

$$\boldsymbol{x}^* = arg\max_{\boldsymbol{x}} p_{K+1}^{aux}(\boldsymbol{x}) unc_n + r\left[1 - p_{K+1}^{aux}(\boldsymbol{x})\right] unc_p$$

r represent the precision of known classes p_{K+1}^{aux} is the probability on unknown class

We can expand the existing uncertainty-based active learning methods to complex and ever-changing open-set scenarios.

Bidirectional Sampling Strategy

- **Bidirectional Least Confident Sampling**

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} p^{aux}_{K+1}(\boldsymbol{x}) \left[1 - \boldsymbol{\mathcal{P}}^{-}_{y^-}(\boldsymbol{x})\right] + r\left[1 - p^{aux}_{K+1}(\boldsymbol{x})\right] \left[1 - \boldsymbol{p}^{+}_{y^+}(\boldsymbol{x})\right],$$

where $y^- = \arg\max_y \boldsymbol{\mathcal{P}}^{-}_{y}(\boldsymbol{x})$, $y^+ = \arg\max_y \boldsymbol{p}^{+}_{y}(\boldsymbol{x})$.

- **Bidirectional Margin-Based Sampling**

$$\begin{aligned} \boldsymbol{x}^* = & \arg\max_{\boldsymbol{x}} p^{aux}_{K+1}(\boldsymbol{x}) \left[\boldsymbol{\mathcal{P}}^{-}_{y_1^-}(\boldsymbol{x}) - \boldsymbol{\mathcal{P}}^{-}_{y_2^-}(\boldsymbol{x})\right] \\ & + r\left[1 - p^{aux}_{K+1}(\boldsymbol{x})\right] \left[\boldsymbol{p}^{+}_{y_1^+}(\boldsymbol{x}) - \boldsymbol{p}^{+}_{y_2^+}(\boldsymbol{x})\right], \end{aligned}$$

where $y_1^- = \arg\max_y \boldsymbol{\mathcal{P}}^{-}_{y}(\boldsymbol{x})$, $y_1^+ = \arg\max_y \boldsymbol{p}^{+}_{y}(\boldsymbol{x})$ $y_2^- = \arg\max_{y \backslash y_1^-} \boldsymbol{\mathcal{P}}^{-}_{y}(\boldsymbol{x})$, $y_2^+ = \arg\max_{y \backslash y_1^+} \boldsymbol{p}^{+}_{y}(\boldsymbol{x})$.

- **Bidirectional Entropy-Based Sampling**

$$\begin{aligned} \boldsymbol{x}^* = & \arg\max_{\boldsymbol{x}} p^{aux}_{K+1}(\boldsymbol{x}) \left[-\boldsymbol{\mathcal{P}}^{-}_{y^-}(\boldsymbol{x}) \log \boldsymbol{\mathcal{P}}^{-}_{y^-}(\boldsymbol{x})\right] \\ & + r\left[1 - p^{aux}_{K+1}(\boldsymbol{x})\right] \left[-\boldsymbol{p}^{+}_{y^+}(\boldsymbol{x}) \log \boldsymbol{p}^{+}_{y^+}(\boldsymbol{x})\right], \end{aligned}$$

where y^- and y^+ are consistent with the previous definition.

Bidirectional Sampling Strategy

- **Bidirectional Least Confident Sampling**

$$x^* = \arg\max_{x} p^{aux}_{K+1}(x)\left[1-\mathcal{P}^-_{y^-}(x)\right] + r\left[1-p^{aux}_{K+1}(x)\right]\left[1-p^+_{y^+}(x)\right],$$

where $y^- = \arg\max_y \mathcal{P}^-_y(x)$, $y^+ = \arg\max_y p^+_y(x)$.

- **Bidirectional Margin-Based Sampling**

$$x^* = \arg\max_{x} p^{aux}_{K+1}(x)\left[\mathcal{P}^-_{y_1^-}(x) - \mathcal{P}^-_{y_2^-}(x)\right] + r\left[1-p^{aux}_{K+1}(x)\right]\left[p^+_{y_1^+}(x) - p^+_{y_2^+}(x)\right],$$

where $y_1^- = \arg\max_y \mathcal{P}^-_y(x)$, $y_1^+ = \arg\max_y p^+_y(x)$ $y_2^- = \arg\max_{y \backslash y_1^-} \mathcal{P}^-_y(x)$, $y_2^+ = \arg\max_{y \backslash y_1^+} p^+_y(x)$.

- **Bidirectional Entropy-Based Sampling**

$$x^* = \arg\max_{x} p^{aux}_{K+1}(x)\left[-\mathcal{P}^-_{y^-}(x)\log \mathcal{P}^-_{y^-}(x)\right] + r\left[1-p^{aux}_{K+1}(x)\right]\left[-p^+_{y^+}(x)\log p^+_{y^+}(x)\right],$$

where y^- and y^+ are consistent with the previous definition.

Outline

- Motivation
- The Framework of BUAL
- Random Label Negative Learning
- Bidirectional Sampling Strategy
- Experiments

Experiments

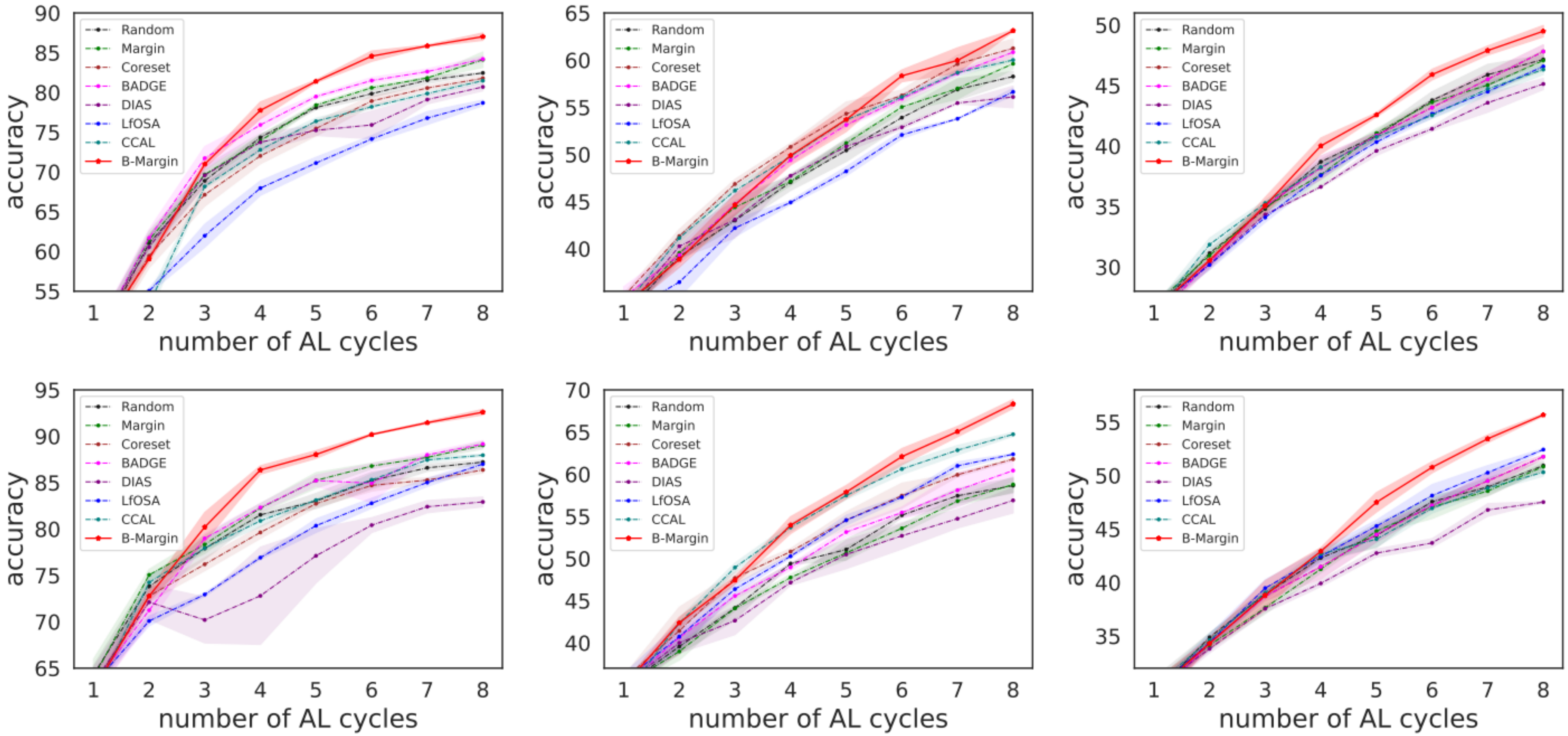


Figure 6: Accuracy comparison on CIFAR-10 (first column), CIFAR-100 (second column), and Tiny-Imagenet (third column). The ratio of unknown class examples to the total number of examples is fixed at 0.4 (first row) and 0.6 (second row) for each dataset. The results of 0.2 and 0.8 openness ratios are shown in the supplementary file.

Experiments

Table 1: The final round average accuracy of different methods on CIFAR-10, CIFAR-100, and Tiny-Imagenet. The best performance is highlighted in bold.

Datasets		CIFAR-10				CIFAR-100				Tiny-Imagenet			
Openness Ratio		0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8
(1)	Random	83.3	82.5	87.2	96.9	57.6	58.3	58.7	61.2	45.7	47.2	50.9	55.0
(2)	LC	84.3	81.6	87.5	96.2	55.8	54.6	54.0	56.2	44.8	45.9	48.4	51.6
	Margin	86.0	84.1	89.0	97.0	59.3	59.6	58.8	58.9	46.4	47.1	50.8	54.0
	Entropy	85.4	83.4	88.0	96.8	57.1	56.8	55.7	56.4	44.6	44.5	46.9	50.7
(3)	Coreset	85.0	81.8	86.4	97.4	60.2	61.2	61.8	64.2	46.2	47.8	51.8	54.0
(4)	BADGE	86.8	84.2	89.2	96.4	60.2	60.8	60.4	62.0	46.3	47.8	51.8	53.3
(5)	LfOSA	73.7	78.7	87.0	98.6	52.3	56.6	62.4	68.2	42.5	46.6	52.4	59.9
	CCAL	80.8	81.5	88.0	98.1	55.9	60.0	64.7	67.7	44.4	46.3	50.3	57.0
(6)	DIAS	81.8	80.7	83.0	94.0	55.7	56.1	56.9	57.2	43.1	45.1	47.5	54.4
Ours	B-LC	**87.0**	87.2	92.5	**99.1**	59.3	62.8	67.5	**72.1**	45.7	48.7	54.7	60.6
	B-Margin	86.5	87.0	**92.6**	98.9	**60.9**	**63.1**	**68.3**	71.5	**46.5**	**49.5**	**55.7**	**61.2**
	B-Entropy	86.9	**87.4**	**92.6**	**99.1**	58.9	61.7	66.9	71.4	45.4	47.5	55.2	61.0

Experiments

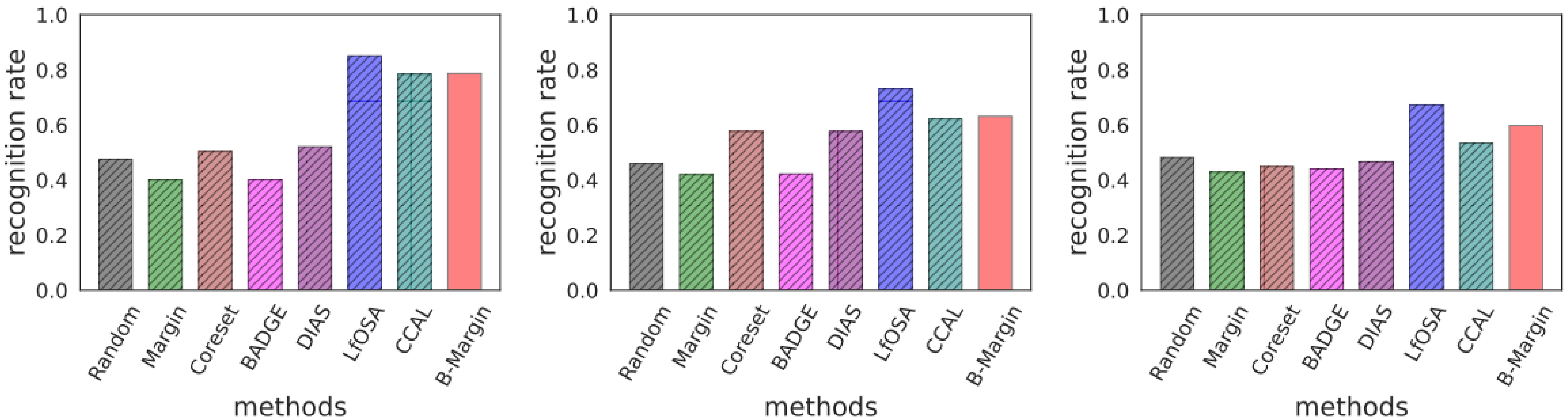


Fig. 7: The average recognition rate on CIFAR-10 (first column), CIFAR-100 (second column), and Tiny-Imagenet (third column).

Experiments

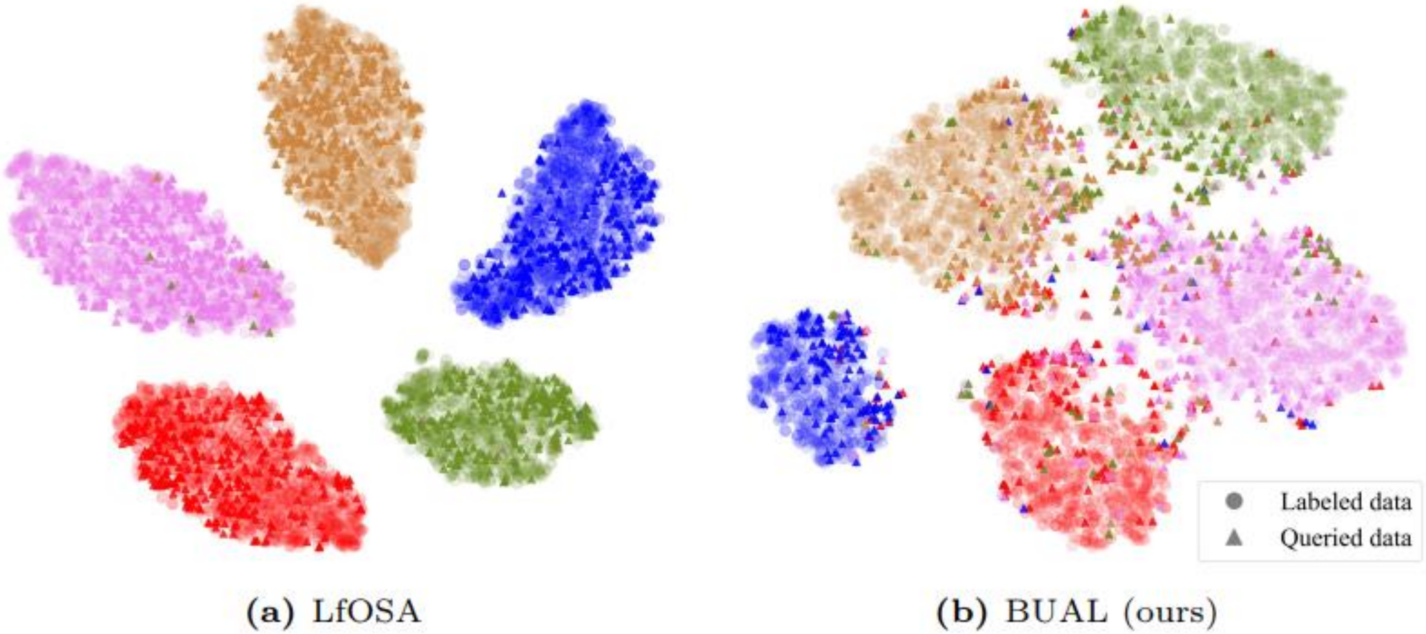


Fig. 8: The t-SNE feature visualization of data from one query and labeled pool on CIFAR-10 with an openness ratio of 0.5.

Experiments

Table 2: Final accuracy of each component in Equation 3 on CIFAR-10, CIFAR-100, and Tiny-Imagenet with an openness ratio of 0.6.

Dataset \ Method	unc_p	unc_n	$w/o\ w$	$w/o\ f_{aux}$	B-LC
CIFAR-10	87.5	89.4	90.8	91.3	**92.5**
CIFAR-100	54.0	63.5	62.4	65.0	**67.5**
Tiny-Imagenet	48.4	52.3	52.0	53.0	**54.7**

Thanks!