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], and Sheng-Jun Huang[1]†

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

Abstract. Active learning (AL) 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. Traditional methods prioritize selecting informative examples with low confidence, with the risk of mistakenly selecting unknown-class examples with similarly low confidence. Recent methods favor the most probable known-class examples, with the risk of picking simple already mastered examples. In this paper, we attempt to query examples that are both likely from known classes and highly informative, and propose a *Bidirectional Uncertainty-based Active Learning* (BUAL) framework. Specifically, we achieve this by first pushing the unknown class examples toward regions with high-confidence predictions, *i.e.*, the proposed *Random Label Negative Learning* method. Then, we propose a *Bidirectional Uncertainty sampling* strategy by jointly estimating uncertainty posed by both positive and negative learning to perform consistent and stable sampling. BUAL successfully extends existing uncertainty-based AL methods to complex open-set scenarios. Extensive experiments on multiple datasets with varying openness demonstrate that BUAL achieves state-of-the-art performance. The code is available at this link.

Keywords: Active learning · Open-set annotation · Negative learning · Uncertainty estimation

1 Introduction

Labeling data can be costly and time-consuming, often requiring high levels of expertise from annotators [26]. This expense poses a significant challenge when dealing with insufficient labeled data in deep learning tasks. Recently, active learning (AL) has emerged as a prominent approach to tackle this issue and has gained widespread attention [9,17,21]. It iteratively selects the most informative examples from the unlabeled data pool and queries their labels from an oracle, enabling the learning of an effective model with reduced labeling costs.

Existing AL methods [5,8,19,22,27,29,30] typically operate under the closed-set assumption, assuming that the label categories in the unlabeled data pool

† Corresponding Author: huangsj@nuaa.edu.cn

match those of the target task. However, this assumption often does not hold in practical scenarios. For example, consider a task that involves classifying images into two target categories, "Dog" and "Cat". Collecting training examples through keyword-based image search inevitably introduces irrelevant images from other categories (*i.e.*, unknown classes), alongside the two target categories (*i.e.*, known classes).

In such open-set scenarios, many previous AL methods, which prefer querying examples with less confident predictions, may lead to failure since examples from unknown classes often receive uncertain predictions. To mitigate the impact of unknown class examples, some AL methods designed specifically for open-set scenarios attempt to query examples that are more likely to belong to known classes based on sample similarity [3] or model-predicted max activation value (MAV) [20]. However, since examples similar to the labeled ones may be already mastered by the model, they do not significantly benefit the target model. These methods only perform well when the proportion of unknown class examples is high. When the proportion is low, they tend to perform poorly than traditional AL methods, even inferior to random sampling (please refer to Figure 6). However, determining the proportion of unknown class examples in practical scenarios is often challenging. This further limits the usability of these methods.

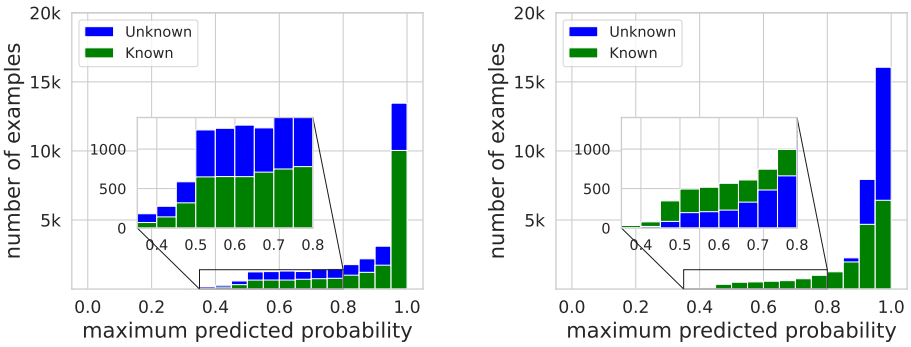
In this paper, we start with a specific question: **can we effectively distinguish the "informative" examples of known classes from examples of unknown classes?** Intuitively, if we can push the unknown class examples toward regions with high-confidence predictions, existing uncertainty-based AL methods can be applied directly in open-set scenarios. To achieve this, we propose to fine-tune the model by performing negative learning (NL) [11,14,16,32] on unlabeled examples. NL is an indirect learning manner that explores the utility of complementary labels, *i.e.*, the label categories that an instance does not belong to. For a K-classification problem, the NL loss is defined as:

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

where f is the model we want to optimize, $\bar{y}$ and $\bar{\boldsymbol{y}} = [\bar{y}_1, \ldots, \bar{y}_k, \ldots, \bar{y}_K]$ represent a complementary label and its corresponding one-hot form, respectively, and p_k denotes the probability of the k-th category.

Specifically, the fine-tuning process comprises two parts. On one hand, for already labeled examples, we train them directly using Equation 1. On the other hand, for unlabeled examples, we first randomly assign labels to them in each training round and then train the model using Equation 1. Notably, the unlabeled known class example has a relatively higher chance of receiving the correct label, whereas the unknown class example will never be assigned the correct label. Once unlabeled known class data 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 data are not constrained to move towards the high-confidence region for counteracting the update gradient produced by the labeled data.

To validate this, we conducted preliminary experiments on CIFAR-10 [12] with 4 known and 6 unknown classes. Figure 1 illustrates the confidence statis-



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

Fig. 1: The statistics of prediction confidence before and after fine-tuning the model. In the zoomed-in area of Figure 1b, we swapped the display order of the two to prevent occlusion, allowing for a more intuitive view of how the distribution has changed.

tics for all unlabeled examples in a fixed query round before and after fine-tuning the model. As expected, known class examples are more prevalent in the low-confidence region, while unknown class ones are more common in high-confidence regions. This distribution characteristic presents a potential solution to the aforementioned question and offers a promising approach to AL for open-set scenarios.

Based on this, we propose a Bidirectional Uncertainty-based Active Learning framework (BUAL). On the one hand, we propose the Random Label Negative Learning (RLNL) method to fine-tune the model and leverage information from the unlabeled data pool. Specifically, we first train a model using all labeled data as a positive classifier. Then, negative learning is performed to fine-tune the model as a negative classifier by randomly assigning labels to data from the unlabeled pool in each training iteration. This effectively distinguishes known class examples with lower confidence predictions from unknown class ones. On the other hand, we propose a Bidirectional Uncertainty (BU) sampling strategy for active selection, which estimates prediction uncertainty from both positive and negative classifiers. By selecting examples with the highest uncertainty, we expect to identify the most informative instances from the known classes. Experiments are performed on multiple datasets with different unknown class ratios. The results demonstrate that BUAL can query more informative known class examples and that the model performance obtained by BUAL is substantially improved compared with existing state-of-the-art methods.

2 Related Work

Active learning (AL) is a prominent approach aimed at reducing label costs by selecting a batch of examples that are most valuable for model training. Existing AL methods can be broadly categorized into three groups based on sample selection strategies: uncertainty-based, representative-based, and hybrid strategies

which combine both aspects. Uncertainty-based strategies focus on sampling informative instances to reduce model uncertainty. Typical methods include Least Confident Sampling [15], Margin-based Sampling [2], and Entropy-based Sampling [7], *etc*. Representative-based strategies start from the sample distribution and aim to select representative instances that match the overall distribution. A typical method is Coreset [25]. Hybrid strategies combine uncertainty and representativeness by incorporating sample distribution information and the model's specific needs. Notable methods in this category include QUIRE [8] and BADGE [1], *etc*.

Open-set annotation (OSA) involves active learning under open-set scenarios [3,20]. Existing methods primarily focus on selecting examples that are most likely to belong to known classes. For instance, CCAL [3] employs contrastive learning to extract semantic and distinctive features of examples, facilitating discrimination of known class examples. LfOSA [20] introduces an auxiliary network to model the per-example max activation value (MAV) distribution and dynamically selects examples with the highest probability from known classes. However, these methods exhibit sensitivity to the openness of the dataset and may not consistently perform well. Open-set recognition (OSR) [18, 23, 24] is a related problem setting to OSA, aiming to predict correct labels for known class examples while simultaneously detecting examples from unknown classes. Nevertheless, direct use of OSR methods often falls short of expected effectiveness [20], mainly due to limited training examples and the inability to identify highly informative instances.

Complementary labels are labels other than the ground-truth label assigned to an example. Complementary label learning (CLL) [4,10,31] was initially introduced in [10], where the authors leveraged the lower acquisition cost and higher correct labeling rate of complementary labels to address multi-class classification problems. However, complementary labels contain less information compared to the correct labels, which can result in slow convergence of the model and challenges in achieving the desired performance when directly using complementary labels for training. Authors in [11] made the first attempt to combine CLL with the noise labeling problem [33] and proposed an indirect and robust training schema called negative learning (NL). In this paper, we exploit the properties of NL and extend it to open-set and label-free settings for the first time.

3 Methodology

3.1 Preliminaries

Notations. In open-set annotation (OSA) settings, there are two labeled data pools: $D_l^{kno} = \left\{(x_i^l, y_i^l)\right\}_{i=1}^{n_l^{kno}}$ for known classes and $D_l^{unk} = \left\{(x_i^l)\right\}_{i=1}^{n_l^{unk}}$ for unknown classes, as well as an unlabeled data pool: $D_u = \{(x_i^u)\}_{i=1}^{n_u^{kno}} \cup \{(x_i^u)\}_{i=1}^{n_u^{unk}}$ containing examples from both known and unknown classes. Each known class example belongs to one of K known classes in the label space $\mathcal{Y} = \{1, 2, \ldots, K\}$. The unknown class examples are uniformly grouped into one category, denoted

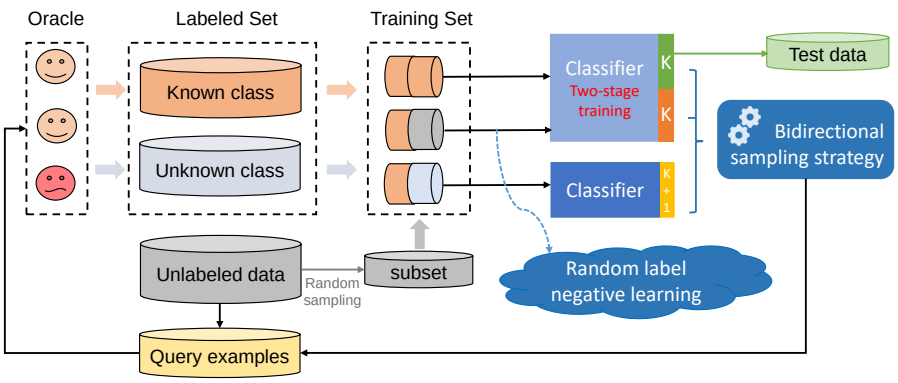


Fig. 2: The framework of BUAL. A two-stage K class classifier is maintained, where the first stage is trained in a normal manner saved as $f_p(\cdot)$ and the second stage is trained using the proposed random label negative learning method denoted as $f_n(\cdot)$. An auxiliary $K+1$ class classifier $f_{aux}(\cdot)$ is trained in parallel. By collecting the predicted uncertainty from $f_p(\cdot)$ and $f_n(\cdot)$ on each candidate example along with the global and local balancing factors, the proposed bidirectional sampling strategy can accurately estimate the potential utility of each example and perform effective sample sampling under complex open-set scenarios.

as $y = \emptyset$. In each training round, active learning (AL) queries a batch of b examples according to a given query strategy $\mathcal{A}$, denoted as $X_{query} = X_{query}^{kno} \cup X_{query}^{unk}$. Once the labeled feedback is obtained, we can calculate a ratio $r = \frac{|X_{query}^{kno}|}{|X_{query}|}$ to represent the precision of known classes.

Overview. The proposed Bidirectional Uncertainty-based Active Learning (BUAL) framework is depicted in Figure 2. In each iteration, there are in general three steps:

- Model Training: the algorithm first trains a target classifier in a normal learning manner, dubbed positive classifier $f_p(\cdot)$. Then, the algorithm fine-tunes the classifier with a new classifier head by the proposed *Random Label Negative Learning* method, dubbed negative classifier $f_n(\cdot)$. Similar to [20], a $K+1$ auxiliary classifier $f_{aux}(\cdot)$ is trained in parallel.
- Example Selection: the *Bidirectional Uncertainty Sampling Strategy* estimates uncertainty bidirectionally using both the positive and negative classifier heads. This estimation is combined with dynamic balance factors generated by the auxiliary classifier and query feedback to select the most informative known class examples.
- Oracle Labeling: the annotators assign class labels to the selected examples. Based on the feedback results, update the corresponding data pools accordingly.

3.2 Random Label Negative Learning

In OSA scenarios, conventional uncertainty-based AL methods tend to be ineffective for example selection, mainly due to the unconfident predictions generated for the unknown class examples. Existing OSA methods prioritize sample purity without fully exploring sample informativeness, resulting in queries with too many already mastered examples that are not useful for model training. Thus, a key question arises: how to distinguish between unknown and "informative" known class examples.

To cope with this problem, we propose a general method by pushing unknown class examples toward the high-confidence regions, while pushing known class examples to the low-confidence regions. If this separation is achieved, we can leverage existing uncertainty-based AL methods directly to handle various complex open-set scenarios. Fortunately, we achieve this to some extent by employing negative learning (NL), where we assign random labels to unlabeled examples and fine-tune the model with a new classifier head accordingly, dubbed Random Label Negative Learning (RLNL).

Specifically, in the model training stage, we first train a K-class classifier $f_p(\cdot)$ in a normal training manner (*e.g.*, cross-entropy loss) based on D_l^{kno}. Note that $f_p(\cdot)$ can be the target model we eventually need to output. After training on the labeled data, $f_p(\cdot)$ can own a good discriminative ability for known class examples at the representation level. Then, we replace the last layer of $f_p(\cdot)$ and fine-tune a new classifier head $f_n(\cdot)$ with Equation 1. In this phase, unlabeled examples in D_u are involved in training to help achieve the goal mentioned above: pushing unknown and known class examples toward the high and low confidence regions, respectively. All labeled examples in D_l^{kno} are still involved to prevent them from being incorrectly shifted in the distribution. Eventually, we assign random labels to the unlabeled examples in D_u, while using the complementary labels instead for the labeled known class examples, *i.e.*, for an example $\boldsymbol{x}$:

$$P(\bar{y} = s) = \frac{1}{|S|}, s \in \begin{cases} S = \mathcal{Y} \setminus y^l, & \text{if } \boldsymbol{x} \in D_l^{kno}, \\ S = \mathcal{Y}, & \text{if } \boldsymbol{x} \in D_u. \end{cases} \tag{2}$$

where $\bar{y}$ is uniformly sampled at each training iteration. Here, considering that the number of examples in D_u is usually much larger than that in D_l^{kno}, a small subset D_{sub} is randomly selected from the D_u to save training cost.

Why RLNL works? The first question one may hold is why it shouldn't be optimal for the unlabeled examples to be the uniform distribution as $\bar{y}$ is sampled at each iteration. To explain this, we take two different kinds of updating for a single example in the binary classification scenario of "Dog vs. Cat" as illustrations: negative learning using all labels at each iteration and negative learning using only one random label at each iteration. One possible case is shown in Figure 3. If each iteration uses all labels for negative learning, *i.e.*, such an example is neither a "Cat" nor a "Dog", it is apparent that an update gradient of 0 is only achieved when the example is at the decision boundary. In contrast, if each iteration gives only one random label for negative learning,

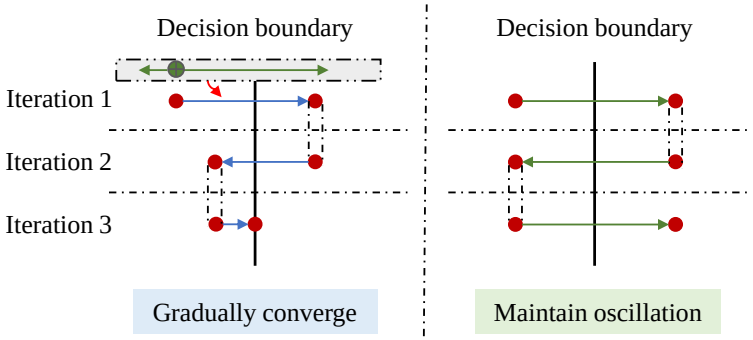


Fig. 3: Use all labels per iteration for negative learning (left) vs. use one random label per iteration for negative learning (right).

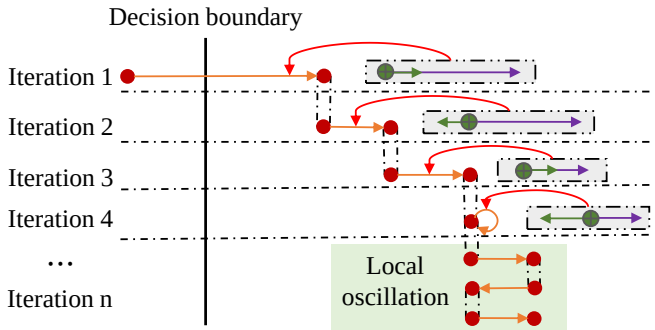
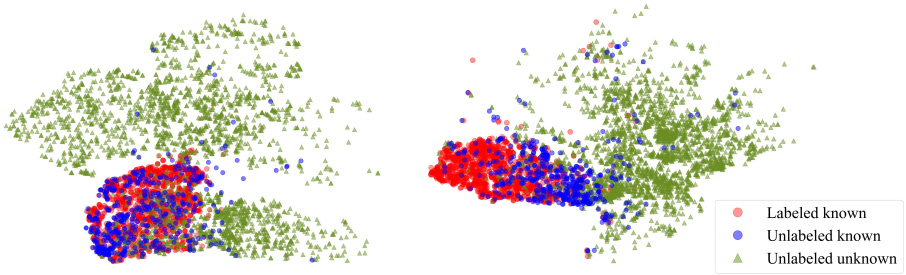


Fig. 4: The possible RLNL update scenario for unlabeled unknown class data in batch deep learning manner. The green "$\longrightarrow$" is the batch update gradient produced by the example itself, and the purple "$\longrightarrow$" is the update gradient produced by labeled data. Initially, the decision boundary is close to the left-hand category.

i.e., such an example is not a "Cat" or a "Dog", then for an instance that is not a "Cat", pushing it as much as possible towards the "Dog" side will result in a gradient update of 0, and vice versa. Obviously, there is no fixed optimal scenario for this type of data.

In RLNL, we utilize this property by exploiting the prior knowledge contained in the labeled data. For unlabeled known class data, they might have overlapping regions with labeled data in the feature space, benefiting from the feature representations learned in the previous stage and the simultaneous introduction of labeled data for negative learning. The mapping of such examples in the feature space will remain or be close to the labeled ones in subsequent model updates, as they will receive invisible constraints provided by the prior knowledge from the labeled data, owing to similar features.

As shown in Figure 4, we present the possible update scenario for unlabeled unknown class examples in the commonly adopted batch deep learning manner.



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

Fig. 5: The t-SNE feature visualization of labeled data, unlabeled known class data, and unlabeled unknown class data on CIFAR-10 with an openness ratio of 0.5 before and after performing RLNL. For a more intuitive visualization, we only show a single known class. More visualization results are shown in the supplementary file.

Here, the green arrow indicates the batch update gradient produced by itself, while the purple arrow indicates the batch update gradient produced by labeled data. Different from Figure 3, where data oscillates on both sides of the decision boundary, the unlabeled unknown class examples will oscillate at uncharted away from the decision boundary to counteract the update gradient due to the labeled ones. In contrast, unlabeled known class examples will move much less than unknown class ones in magnitude in the feature space within the constraints of the prior knowledge provided by the labeled data. This is further confirmed in Figure 5, which illustrates the t-SNE feature visualization of labeled data, unlabeled known class data, and unlabeled unknown class data before and after performing RLNL. This ultimately leads to the result in Figure 1b and proves that RLNL does work.

3.3 Bidirectional Uncertainty Sampling Strategy

During the fine-tuning process, we observed that the predictions of $f_n(\cdot)$ for examples in D_u tend to oscillate between epochs. To ensure stable sampling, on one hand, we test all unlabeled examples t times at m round intervals to obtain the predicted probabilities $\boldsymbol{p_t^-} = (p_1^t, ..., p_K^t)$, which will further be averaged as $\boldsymbol{\mathcal{P}^-} = \frac{1}{t}\sum_{i=1}^{t} \boldsymbol{p_i^-} = (\frac{1}{t}\sum_{i=1}^{t} p_1^i, ..., \frac{1}{t}\sum_{i=1}^{t} p_K^i)$. On the other hand, we reuse the predictions of $f_p(\cdot)$ to produce predicted probabilities $\boldsymbol{p^+} = (p_1, ..., p_K)$ for all unlabeled examples, as $f_p(\cdot)$ are accurate for measuring known class ones.

Compared to the positive head, the negative head is slightly biased for the measurement of sample uncertainty due to the unstable training. Therefore, if an example is likelier to belong to known classes, we prefer to utilize the sample uncertainty obtained from $f_p(\cdot)$. On the contrary, once an example has a higher risk of belonging to the unknown classes, the uncertainty obtained from $f_p(\cdot)$ is unreliable, and thus the uncertainty by $f_n(\cdot)$ should be given a higher weight.

To achieve this, we introduce two balancing factors based on global and local observations. In open-set scenarios, some unknown class examples might be

mistakenly selected for annotation. Although these examples cannot be directly used for training the target model, they are valuable for measuring sample purity. Therefore, similar to [20], we train a $K+1$ classifier $f_{aux}(\cdot)$ in a normal training manner based on the examples from both D_l^{kno} and D_l^{unk}. Then, we can obtain the predicted probability $p_{K+1}^{aux}(\boldsymbol{x}) = f_{aux}(\emptyset \mid \boldsymbol{x})$ for each example. This can act as a local balancing factor, since the larger the value of $p_{K+1}^{aux}(\boldsymbol{x})$ the more likely $\boldsymbol{x}$ is to belong to the unknown class.

Additionally, once the selected examples are sent to the oracle for annotation, we can calculate the ratio r, which provides a rough estimate of the current openness of D_u and can serve as a global balancing factor. With the two balancing factors, we propose a bidirectional uncertainty sampling strategy defined as follows:

$$\boldsymbol{x}^* = arg\max_{\boldsymbol{x}} p_{K+1}^{aux}(\boldsymbol{x}) unc_n + r\left[1 - p_{K+1}^{aux}(\boldsymbol{x})\right] unc_p, \tag{3}$$

where unc_p and unc_n denote the uncertainty of $\boldsymbol{x}$ to the positive classifier head and the negative classifier head, respectively. Noteworthy, this sampling strategy remains applicable even in closed-set settings. If there are no unknown class examples, the ratio r will always be equal to 1, and $p_{K+1}^{aux}(\boldsymbol{x})$ will be 0, effectively making the strategy equivalent to a normal uncertainty sampling strategy.

With Equation 3, extending the existing closed-set uncertainty-based active learning methods to open-set scenarios is possible. In this paper, we focus on three classical uncertainty sampling strategies: least confident sampling, margin-based sampling, and entropy-based sampling. The corresponding modified versions for open-set scenarios are as follows:

- **Bidirectional Least Confident Sampling**

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} p_{K+1}^{aux}(\boldsymbol{x}) \left[1 - \boldsymbol{\mathcal{P}}_{y^-}^-(\boldsymbol{x})\right] + r\left[1 - p_{K+1}^{aux}(\boldsymbol{x})\right]\left[1 - \boldsymbol{p}_{y^+}^+(\boldsymbol{x})\right], \tag{4}$$

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

- **Bidirectional Margin-Based Sampling**

$$\begin{aligned} \boldsymbol{x}^* = &\arg\max_{\boldsymbol{x}} p_{K+1}^{aux}(\boldsymbol{x}) \left[\boldsymbol{\mathcal{P}}_{y_1^-}^-(\boldsymbol{x}) - \boldsymbol{\mathcal{P}}_{y_2^-}^-(\boldsymbol{x})\right] \\ &+ r\left[1 - p_{K+1}^{aux}(\boldsymbol{x})\right]\left[\boldsymbol{p}_{y_1^+}^+(\boldsymbol{x}) - \boldsymbol{p}_{y_2^+}^+(\boldsymbol{x})\right], \end{aligned} \tag{5}$$

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

- **Bidirectional Entropy-Based Sampling**

$$\begin{aligned} \boldsymbol{x}^* = &\arg\max_{\boldsymbol{x}} p_{K+1}^{aux}(\boldsymbol{x}) \left[-\boldsymbol{\mathcal{P}}_{y^-}^-(\boldsymbol{x}) \log \boldsymbol{\mathcal{P}}_{y^-}^-(\boldsymbol{x})\right] \\ &+ r\left[1 - p_{K+1}^{aux}(\boldsymbol{x})\right]\left[-\boldsymbol{p}_{y^+}^+(\boldsymbol{x}) \log \boldsymbol{p}_{y^+}^+(\boldsymbol{x})\right], \end{aligned} \tag{6}$$

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

The main procedures of BUAL are summarized in Algorithm 1.

Algorithm 1 BUAL Training Procedure

Input: D_l^{kno}, D_l^{unk}, D_u, r, $\mathcal{A}$, b, m, t, subset size s, training epoch E, optimizer $\mathcal{O}$.
Output: D_l^{kno}, D_l^{unk}, D_u, r, model parameters θ_p.
Process:

```
# Model training
for j = 1; j ≤ E do
    L_p = Σ_{(x,y)∈D_l^kno} ℓ(f_p(x), y);
    L_aux = Σ_{(x,y)∈D_l^kno ∪ D_l^unk} ℓ(f_aux(x), y);
    Update θ_p, θ_aux = O(L_p, θ_p), O(L_aux, θ_aux);
end for
Use f_aux(x) to remove confidently unknown examples in D_u and randomly select
s examples as D_sub;
for j = 1; j ≤ E do
    Generate label ȳ for each example.
    L_n = Σ_{(x,y)∈D_l^kno ∪ D_sub} ℓ_NL(f_n(x), ȳ);
    Update θ_n = O(L_n, θ_n);
    if j mod m = 0 then
        Obtain p_j^- for each sample in D_u by f_n(x);
    end if
end for
# Example selection
Calculate P^- for each sample across rounds;
Obtain p^+ for each sample in D_u by f_p(x);
Obtain p_{K+1}^aux for each sample in D_u by f_aux(x);
Obtain X_query using BU sampling strategy A;
# Oracle labeling
Ask for annotation and update D_l^kno, D_l^unk, D_u and r.
```

4 Experiments

Datasets. We conduct experiments on three benchmark datasets: CIFAR-10, CIFAR-100 [12], and Tiny-Imagenet [28]. Tiny-Imagenet is a subset of the Imagenet [13], consisting of 200 classes with 500 training images per class. The openness of each dataset is defined as the ratio of unknown classes to the total number of classes, and we set its value to 0.2, 0.4, 0.6, and 0.8 for all datasets.

Comparing methods. We select nine AL strategies for comparison, which can be further categorized into six groups: (1) Random: Randomly select examples from the unlabeled data pool for labeling. (2) Traditional uncertainty-based strategies: Least confident sampling (LC), Margin-based sampling (Margin), and Entropy-based sampling (Entropy). (3) Diversity-based strategy: Coreset. (4) Hybrid-based strategy: BADGE. (5) OSA methods: CCAL and LfOSA. (6) OSR method: DIAS [18]. Correspondingly, our methods are Bidirectional Least confident sampling (B-LC), Bidirectional Margin-based sampling (B-Margin), and Bidirectional Entropy-based sampling (B-Entropy).

Training details. On CIFAR-10, CIFAR-100, and Tiny-Imagenet, we randomly sample 1%, 8%, and 8% known class examples as the initial labeled data,

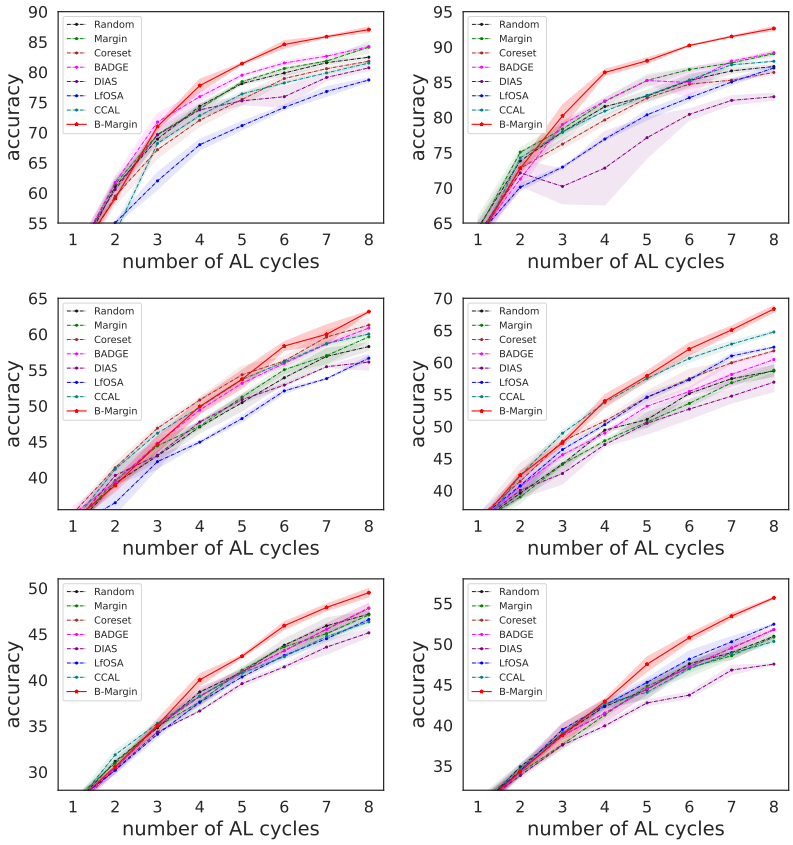


Fig. 6: Accuracy comparison on CIFAR-10 (first row), CIFAR-100 (second row), and Tiny-Imagenet (third row). The ratio of unknown class examples to the total number of examples is fixed at 0.4 (first column) and 0.6 (second column) for each dataset.

respectively. All the models involved in the experiments are ResNet18 [6], trained for 100 epochs, using SGD as the optimizer, where the learning rate is 0.01, momentum is 0.9, weight decay is 1e-4, and batch size is 128. We perform the experiments for 3 runs and report the average results. For CIFAR-10 and CIFAR-100, 5000 examples are randomly selected from the unlabeled pool as D_{sub}, and 1500 examples are queried in each query round. Due to the doubling of data volume, we randomly select 10000 examples as D_{sub} and query 3000 examples in each query round for Tiny-Imagenet.

4.1 Performance Comparison

Figure 6 presents the variation curves for the classification accuracy of the proposed methods and the comparison methods. Here, to better observe the variation curves, we only show the B-Margin for the proposed methods and the

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

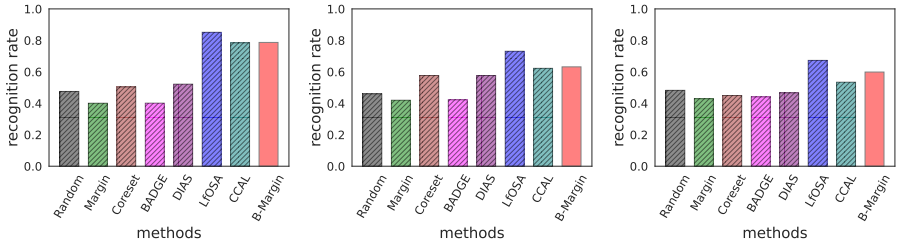
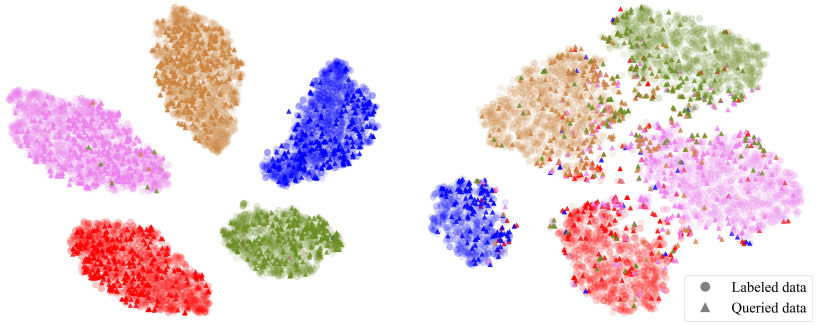


Fig. 7: The average recognition rate on CIFAR-10 (first column), CIFAR-100 (second column), and Tiny-Imagenet (third column).

Margin for the compared traditional uncertainty methods. Table 1 reports all methods' final round average accuracy.

We can observe that our proposed methods consistently achieve the highest classification accuracy, demonstrating the effectiveness and superiority of our BUAL framework over other methods. Some notable observations are as follows: 1) The OSA methods, CCAL and LfOSA, gradually lose effectiveness as the proportion of known classes in the dataset increases. In contrast, traditional uncertainty-based methods perform better as the openness ratio decreases. This is consistent with our previous analysis. 2) The performance of Coreset and BADGE also deteriorates when the openness ratio is high. These methods tend to select examples with diverse characteristics. However, unknown class examples often differ significantly in characteristics from the labeled ones, making them more likely to be sampled and thus undermining the effectiveness of this type of method. 3) The OSR method DIAS does not perform well in all situations. The



(a) LfOSA **(b)** BUAL (ours)

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.

limited labeled data prevents the model from learning a robust representation for identifying open-set examples effectively. Moreover, DIAS cannot identify highly informative examples, resulting in its queries often being simple and unhelpful examples. 4) All our methods remain stable and do not suffer significantly from changes in the openness ratio. This can demonstrate two points. The dynamic balance factors in the proposed framework can adaptively assign appropriate weights to positive and negative uncertainties regardless of the openness of the dataset, and negative uncertainties are indeed effective for querying highly informative known class examples.

Figure 7 shows the average recognition rate of known class examples across queries and openness ratios. We can observe that traditional AL methods all perform poorly due to their inability to recognize known class examples from those with high uncertainty and/or representative. As an OSR method, DIAS shows only marginal improvement compared to the traditional AL methods, unable to fully exploit its exceptional performance with limited training data. The recognition ability of CCAL decreases slightly with the increase of dataset categories compared to our method. It is noted that LfOSA achieved the highest query precision on all datasets. However, combining the results in Table 1, we can find that the performance of models trained by LfOSA is not satisfactory compared to other methods. To explain why, we visualize the queried examples' feature representation and labeled examples' feature representation for our method and LfOSA in Figure 8. We can observe that the features of examples queried by LfOSA are highly overlapping with the labeled ones, which are often examples that the model has already mastered and cannot provide an effective bonus for model training. On the contrary, the sample features queried by our method have little overlap and are more distributed in low-density regions, consistent with the goal of AL. These results validate the effectiveness of our approach in querying more informative examples of known classes and maintaining a high recognition rate.

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.

Method / Dataset	$\boldsymbol{unc_p}$	$\boldsymbol{unc_n}$	$\boldsymbol{w/o\ w}$	$\boldsymbol{w/o\ f_{aux}}$	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**

4.2 Ablation Study

The ablation study is conducted on CIFAR-10 with an openness ratio of 0.6 to validate the effectiveness of each component in our proposed query strategy (Equation 3). The final round accuracy is shown in Table 2. Here, $\boldsymbol{unc_n}$ and $\boldsymbol{unc_p}$ indicate that only unc_n and unc_p is adopted for active sampling, respectively. $\boldsymbol{w/o\ w}$ denotes the removal of all balancing factors, *i.e.*, both f_{aux} and r. $\boldsymbol{w/o\ f_{aux}}$ means only removing the local balancing factor f_{aux}.

We can observe that removing any of the components leads to performance degradation. Although $\boldsymbol{unc_n}$ and $\boldsymbol{unc_p}$ are significantly less effective than B-LC, $\boldsymbol{unc_n}$ has a substantial improvement compared to $\boldsymbol{unc_p}$ due to its ability to distinguish informative examples of known classes from examples of unknown classes. The direct combination of unc_n and unc_p, *i.e.*, $\boldsymbol{w/o\ w}$, works better than $\boldsymbol{unc_n}$, as models trained by RLNL may produce oscillate output and thus the uncertainty obtained in $\boldsymbol{unc_n}$ is not necessarily accurate. By adding the dynamic global balancing factor r, $\boldsymbol{w/o\ f_{aux}}$ achieves a better performance. However, it still falls short in comparison to B-LC, which validates the effectiveness of local balancing factor f_{aux}. These results further corroborate the soundness of our strategy design.

5 Conclusion

In this paper, we successfully expand the existing uncertainty-based active learning methods to complex and ever-changing open-set scenarios by proposing a Bidirectional uncertainty-based Active Learning (BUAL) framework. On one hand, to achieve the goal of distinguishing known and unknown class examples with high uncertainty, we propose a simple but effective Random Label Negative Learning (RLNL) method for pushing unknown and known class examples toward the high and low confidence regions respectively. On the other hand, to better measure sample uncertainty, we propose a Bidirectional Uncertainty (BU) sampling strategy by dynamically fusing the sample uncertainty obtained from positive learning and negative learning. The dynamic balancing factors in it can ensure that the strategy is effective under various openness ratios. Extensive experimental results show that the model trained with BUAL can achieve state-of-the-art performance under various open-set scenarios.

Acknowledgements

This work was supported by the Natural Science Foundation of Jiangsu Province of China (BK20222012, BK20211517), the National Key R&D Program of China (2020AAA0107000), and NSFC (62222605).

References

1. Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., Agarwal, A.: Deep batch active learning by diverse, uncertain gradient lower bounds. arXiv preprint arXiv:1906.03671 (2019)
2. Balcan, M.F., Broder, A., Zhang, T.: Margin based active learning. In: International Conference on Computational Learning Theory. pp. 35–50. Springer (2007)
3. Du, P., Zhao, S., Chen, H., Chai, S., Chen, H., Li, C.: Contrastive coding for active learning under class distribution mismatch. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8927–8936 (2021)
4. Feng, L., Kaneko, T., Han, B., Niu, G., An, B., Sugiyama, M.: Learning with multiple complementary labels. In: International Conference on Machine Learning. pp. 3072–3081. PMLR (2020)
5. Fu, Y., Zhu, X., Li, B.: A survey on instance selection for active learning. Knowledge and information systems **35**(2), 249–283 (2013)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
7. Holub, A., Perona, P., Burl, M.C.: Entropy-based active learning for object recognition. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. pp. 1–8. IEEE (2008)
8. Huang, S.J., Jin, R., Zhou, Z.H.: Active learning by querying informative and representative examples. Advances in neural information processing systems **23** (2010)
9. Huang, S.J., Zong, C.C., Ning, K.P., Ye, H.B.: Asynchronous active learning with distributed label querying. In: IJCAI. pp. 2570–2576 (2021)
10. Ishida, T., Niu, G., Hu, W., Sugiyama, M.: Learning from complementary labels. Advances in neural information processing systems **30** (2017)
11. Kim, Y., Yim, J., Yun, J., Kim, J.: Nlnl: Negative learning for noisy labels. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 101–110 (2019)
12. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)
14. Lee, J.H., Astrid, M., Zaheer, M.Z., Lee, S.I.: Deep visual anomaly detection with negative learning. In: International Workshop on Frontiers of Computer Vision. pp. 218–232. Springer (2021)
15. Li, M., Sethi, I.K.: Confidence-based active learning. IEEE transactions on pattern analysis and machine intelligence **28**(8), 1251–1261 (2006)
16. Luo, X., Chen, W., Tan, Y., Li, C., He, Y., Jia, X.: Exploiting negative learning for implicit pseudo label rectification in source-free domain adaptive semantic segmentation. arXiv preprint arXiv:2106.12123 (2021)

17. Mahmood, R., Fidler, S., Law, M.T.: Low budget active learning via wasserstein distance: An integer programming approach. arXiv preprint arXiv:2106.02968 (2021)
18. Moon, W., Park, J., Seong, H.S., Cho, C.H., Heo, J.P.: Difficulty-aware simulator for open set recognition. arXiv preprint arXiv:2207.10024 (2022)
19. Ning, K.P., Tao, L., Chen, S., Huang, S.J.: Improving model robustness by adaptively correcting perturbation levels with active queries. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 9161–9169 (2021)
20. Ning, K.P., Zhao, X., Li, Y., Huang, S.J.: Active learning for open-set annotation. arXiv preprint arXiv:2201.06758 (2022)
21. Ren, P., Xiao, Y., Chang, X., Huang, P.Y., Li, Z., Gupta, B.B., Chen, X., Wang, X.: A survey of deep active learning. ACM computing surveys (CSUR) **54**(9), 1–40 (2021)
22. Roy, N., McCallum, A.: Toward optimal active learning through sampling estimation of error reduction. int. conf. on machine learning (2001)
23. Salehi, M., Mirzaei, H., Hendrycks, D., Li, Y., Rohban, M.H., Sabokrou, M.: A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges. arXiv preprint arXiv:2110.14051 (2021)
24. Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boult, T.E.: Toward open set recognition. IEEE transactions on pattern analysis and machine intelligence **35**(7), 1757–1772 (2012)
25. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. arXiv preprint arXiv:1708.00489 (2017)
26. Settles, B.: Active learning literature survey (2009)
27. Sinha, S., Ebrahimi, S., Darrell, T.: Variational adversarial active learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5972–5981 (2019)
28. Yao, L., Miller, J.: Tiny imagenet classification with convolutional neural networks. CS 231N **2**(5), 8 (2015)
29. Yoo, D., Kweon, I.S.: Learning loss for active learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 93–102 (2019)
30. You, X., Wang, R., Tao, D.: Diverse expected gradient active learning for relative attributes. IEEE transactions on image processing **23**(7), 3203–3217 (2014)
31. Yu, X., Liu, T., Gong, M., Tao, D.: Learning with biased complementary labels. In: Proceedings of the European conference on computer vision (ECCV). pp. 68–83 (2018)
32. Zong, C.C., Cao, Z.T., Guo, H.T., Du, Y., Xie, M.K., Li, S.Y., Huang, S.J.: Noise-robust bidirectional learning with dynamic sample reweighting. arXiv preprint arXiv:2209.01334 (2022)
33. Zong, C.C., Wang, Y.W., Xie, M.K., Huang, S.J.: Dirichlet-based prediction calibration for learning with noisy labels. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 17254–17262 (2024)