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

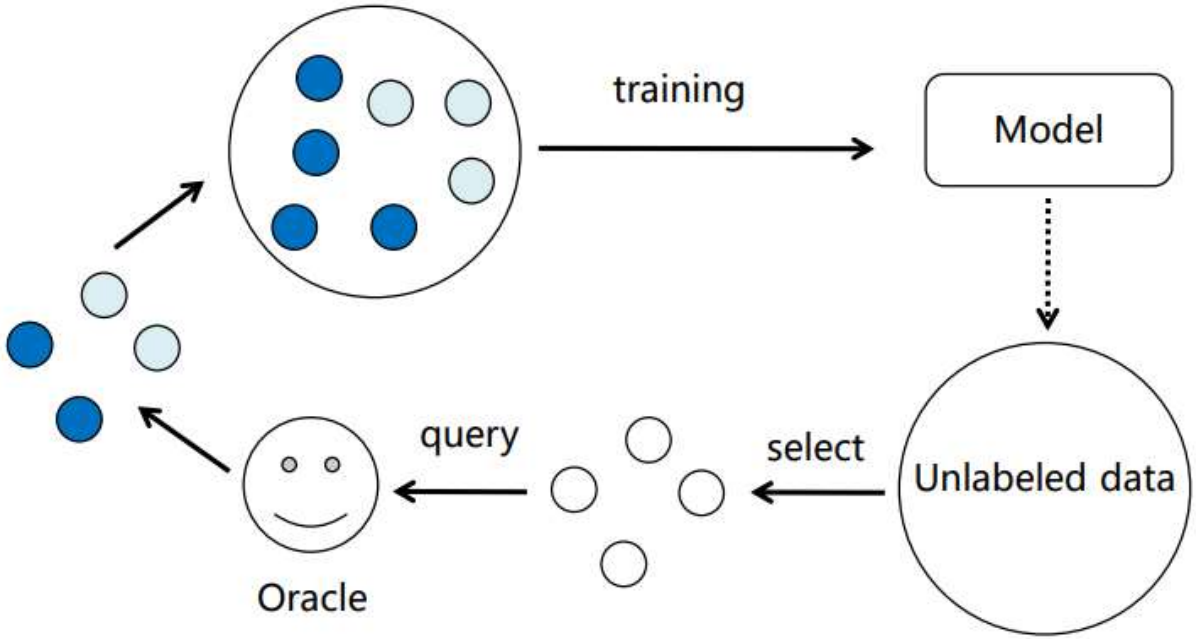
Inconsistency-Based Data-Centric Active Open-Set Annotation

Ruiyu Mao, Ouyang Xu, Yunhui Guo

The University of Texas at Dallas
{ruiyu.mao, oxu, yunhui.guo}@utdallas.edu

AAAI 2024

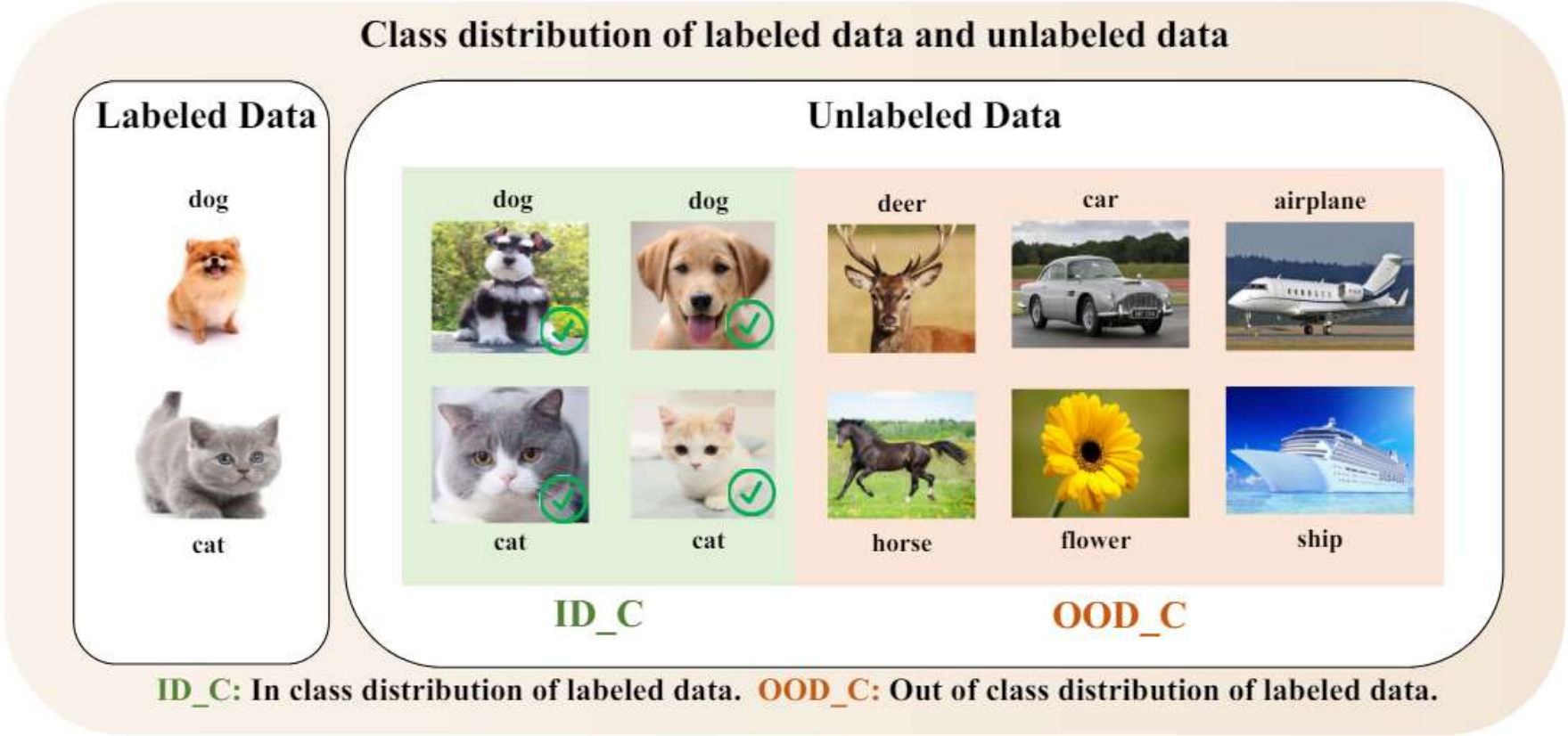
Background: Active Learning



Goal: query less for more.

- Uncertainty-based sampling
  - Least-confidence
  - Margin
  - Entropy
  - MC-dropout
- Diversity-based sampling
  - CoreSet
- Hybird sampling
  - BADGE

Background: Active Open-Set Annotation



OOD examples exhibit high uncertainty and diversity because they share neither class-distinctive features nor other inductive biases with ID examples

Motivation: LfOSA

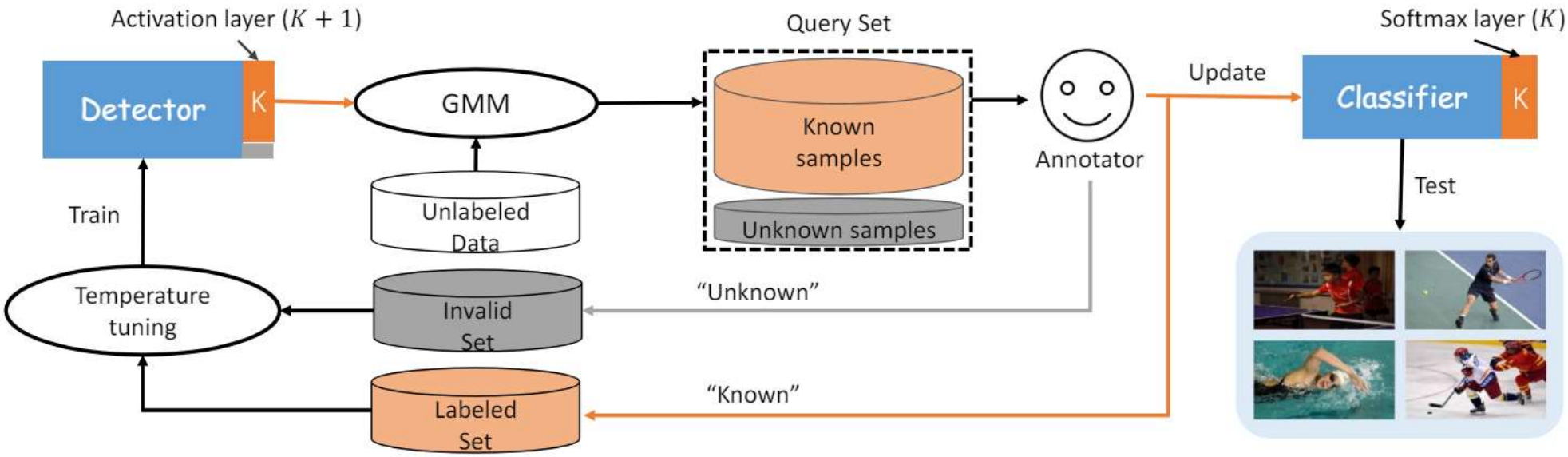


Figure 2. The framework of LfOSA. It includes two networks for detection and classification. The detector attempts to construct a query set for annotation by GMM modeling. After labeling, two networks will be updated for next iteration.

Two limitations:

- ✓ Training the additional detector network is costly;
- ✓ It is difficult to identify informative examples from the known classes.

Method: NEAT

- We suggest a data-centric perspective that naturally separates examples by label clusterability, eliminating the need for an additional detector network.

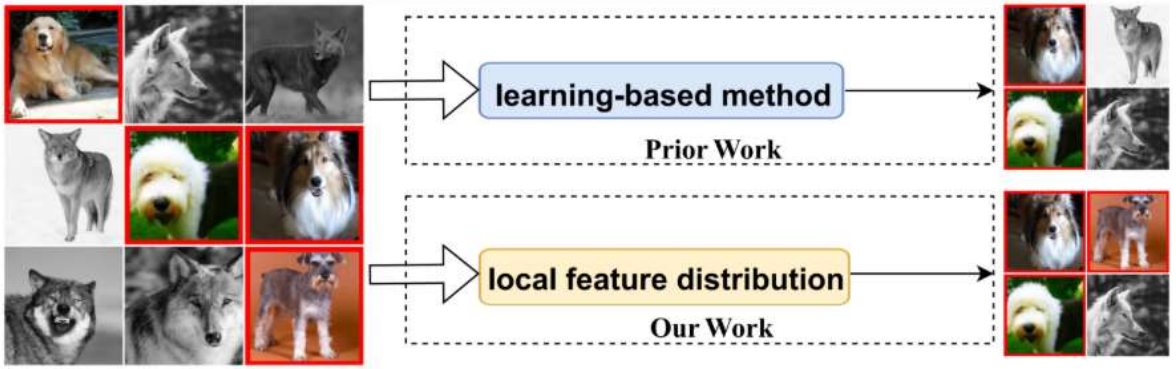


Figure 1: Dataset consists of color images as known dog class and gray-scale images as unknown wolf class. Prior work using learning-based approach may identify some unknown classes as known classes. Our work focusing on local feature distribution can find known classes more accurately.

- Selecting informative examples from known classes by estimating the inconsistency between the model prediction and the local feature distribution.

Method: NEAT

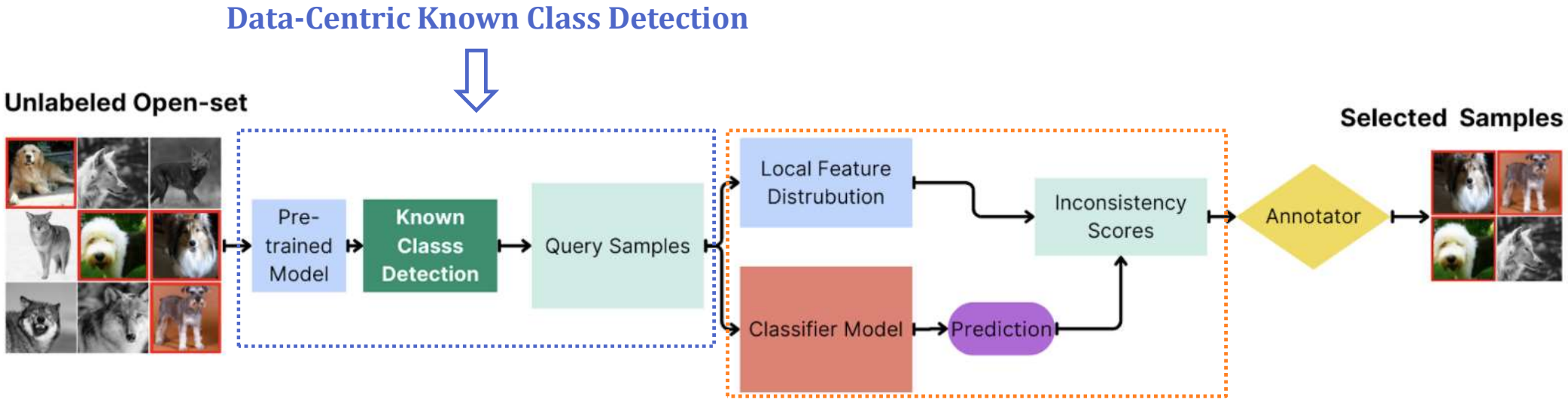


Figure 9: The architecture of NEAT includes a classifier model and a local feature distribution extractor, designed to identify both known and unknown samples.

Data-Centric Known Class Detection

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

- **Label clusterability.**
  - Samples with similar features should belong to the same class.

- **Feature extraction.**
  - CLIP provides high-quality features for calculating feature similarity.

- **Known class detection.**
  a. Each $N_k(x) \in L$ represents the k-th closest samples in L to the unlabeled example x in U.
  b. Compute the count of neighbors with known and unknown classes for each x.
  c. Classify an unlabeled sample x as belonging to a known class if all of its neighboring samples are from known classes.

Inconsistency-Based Active Learning

a. Given the K-nearest neighbors $\{N_i(x)\}_{i=1}^{K} \subset L$ of the example x.

b. Construct a vector $V_x \in \mathbb{R}^C$ with

$$V_x[c] = \sum_k 1(Y_k(x) = c).$$

c. Normalize V_x via the softmax function to be a probabilistic vector $\widetilde{V}_x$.

d. The inconsistency is computed using cross-entropy as

$$I(\mathbf{x}) = -\sum_{c=1}^{C} P_{\mathbf{x}}[c] \log \tilde{V}_{\mathbf{x}}[c].$$

e. Rank all the identified known classes examples using $I(x)$.

f. Select the top B examples for labeling.

Method: NEAT

Algorithm 1: NEAT: Inconsistency-Based Data-Centric Active Open-Set annotation.

Require: A deep neural network f_θ, initial labeled set L, a set of known class Y_{Known} , unlabeled labeled set U, number of query rounds T, number of examples in each query batch B, a pre-trained model M, number of neighbors K.

Use the pre-trained model M for extracting features on L and U.
for $t \leftarrow 1$ to T **do**
    Train the model f_θ on L by minimizing $\sum_{(x,y)\in L} \ell_{\text{CE}}(f_\theta(x), y)$
    $S \leftarrow \{\}$.
    For each sample $x \in U$:
        Compute the output of the softmax function as P_x.
        Find the K-nearest neighbors $\{N_i(\mathbf{x})\}_{i=1}^{K}$ of $\mathbf{x}$ in L based on the extracted features.
        If all the labels from $\{N_i(\mathbf{x})\}_{i=1}^{K}$ belong to known classes Y_{Known}, then $S \leftarrow S \bigcup \{\mathbf{x}\}$.
    Compute the score $I(x)$ using Eq. 5 for each sample $\mathbf{x} \in S$.
    Rank the samples based on $I(\mathbf{x})$ and denote the B samples which have the largest scores as U_B.
    Query the labels of each sample in U_B.
    $U \leftarrow U \setminus U_B$, $L \leftarrow L \bigcup U_B$.
end for

Experiments

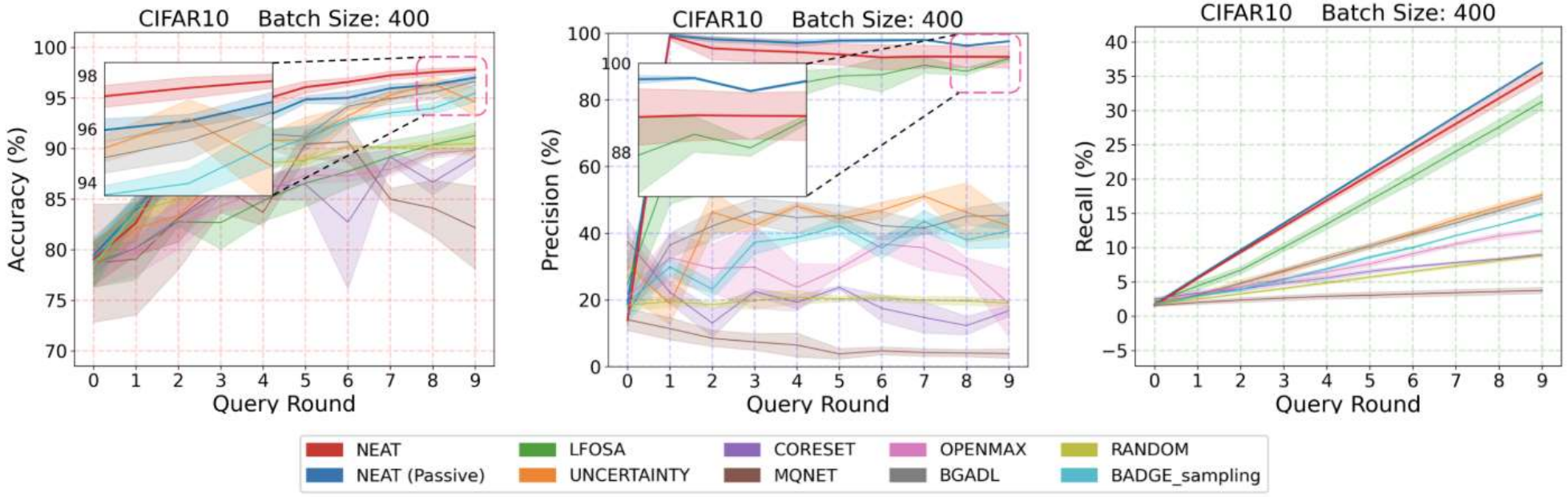


Figure 2: NEAT achieves higher precision, recall and accuracy compared with existing active learning methods for active open-set annotation. We evaluated NEAT and the baseline active learning methods on **CIFAR10**, **CIFAR100** and **Tiny-ImageNet** based on accuracy, precision and recall.

Experiments

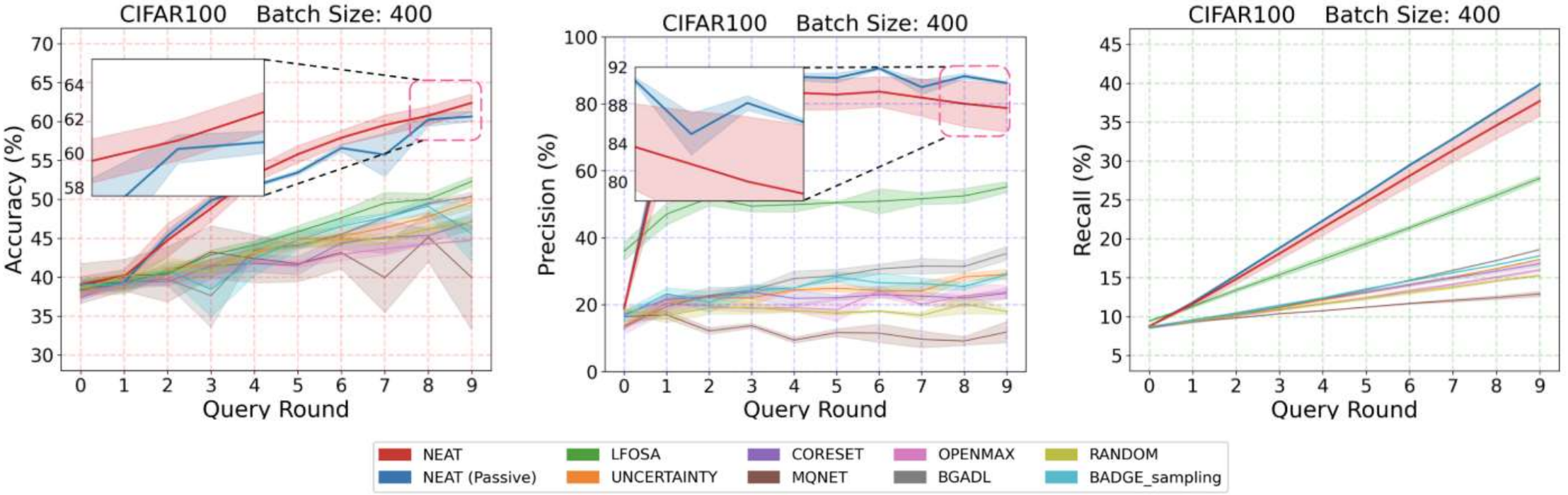


Figure 2: NEAT achieves higher precision, recall and accuracy compared with existing active learning methods for active open-set annotation. We evaluated NEAT and the baseline active learning methods on **CIFAR10**, **CIFAR100** and **Tiny-ImageNet** based on accuracy, precision and recall.

Experiments

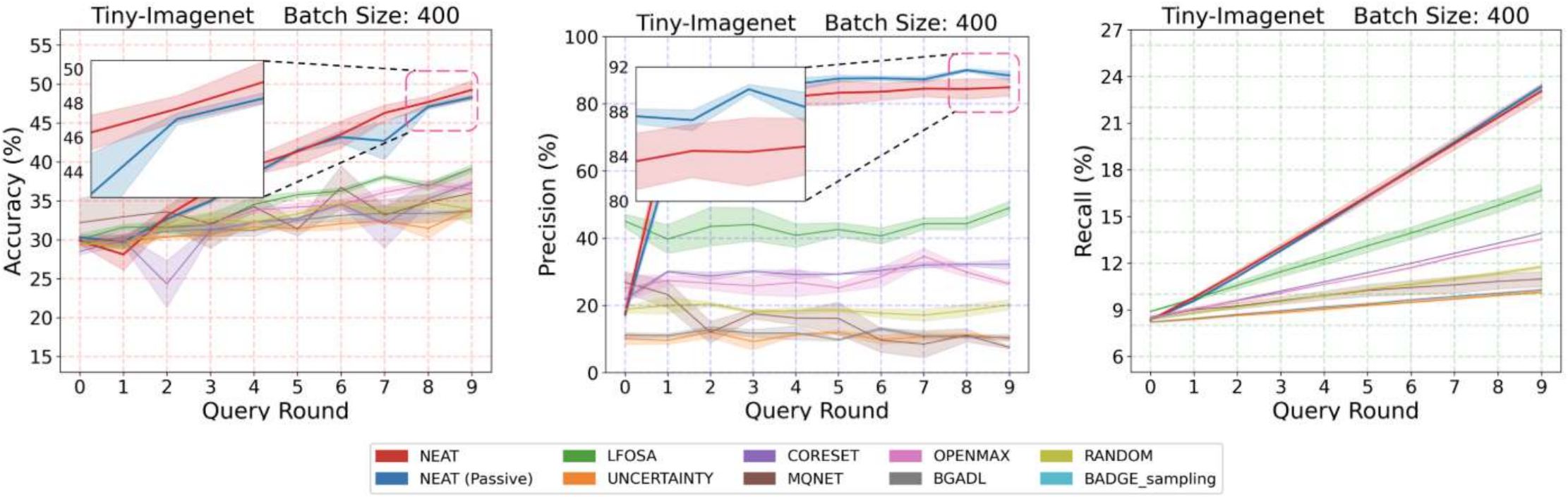


Figure 2: NEAT achieves higher precision, recall and accuracy compared with existing active learning methods for active open-set annotation. We evaluated NEAT and the baseline active learning methods on **CIFAR10**, **CIFAR100** and **Tiny-ImageNet** based on accuracy, precision and recall.

Experiments

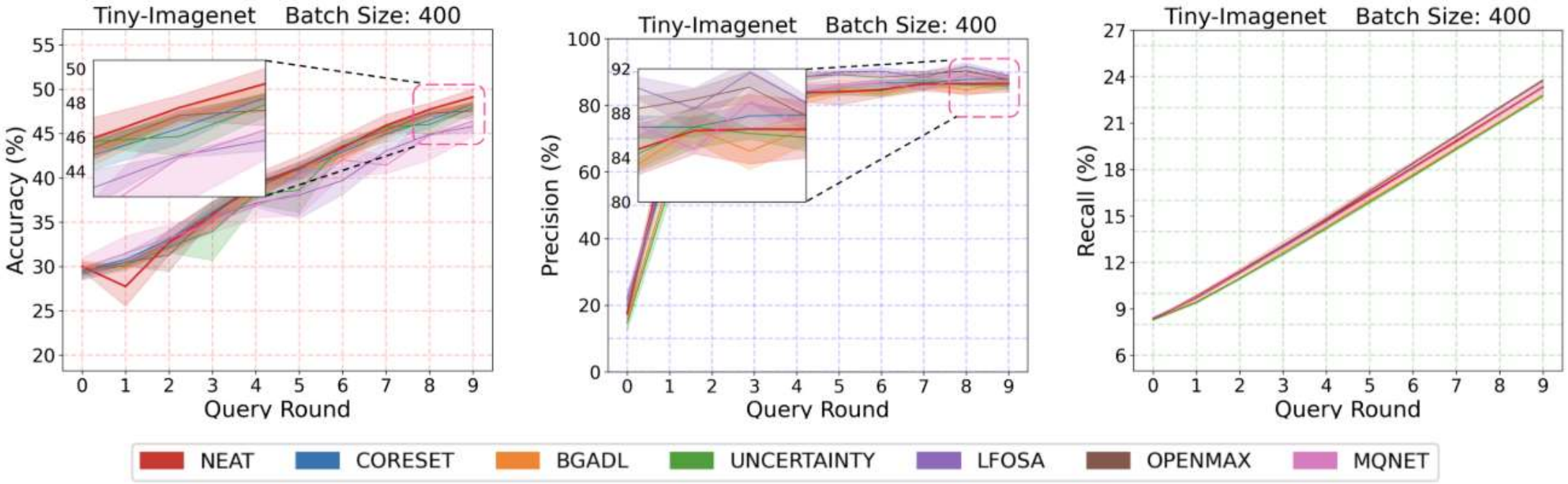


Figure 3: NEAT is effective compared with other active learning methods for deep neural networks.

Experiments

Dataset	Model	Accuracy (avg ± std)
Tiny-Imagenet	ResNet-50	50.53 ± 1.15
	ResNet-34	50.17 ± 0.20
	ResNet-18	48.98 ± 0.42
	CLIP	49.01 ± 1.60
CIFAR100	ResNet-50	62.00 ± 1.26
	ResNet-34	62.73 ± 0.45
	ResNet-18	60.75 ± 0.58
	CLIP	63.68 ± 0.67
CIFAR10	ResNet-50	97.53 ± 0.24
	ResNet-34	97.48 ± 0.19
	ResNet-18	97.48 ± 0.26
	CLIP	98.15 ± 0.14

Table 1: NEAT is robust to the choices of pre-trained models.

Experiments

Dataset	# of Neighbors	Accuracy (avg ± std)
Tiny-Imagenet	5	47.23 ± 1.21
	10	49.01 ± 1.60
	15	49.02 ± 0.61
	20	50.73 ± 0.29
CIFAR100	5	62.63 ± 0.66
	10	63.68 ± 0.67
	15	62.77 ± 0.84
	20	60.85 ± 1.70
CIFAR10	5	97.95 ± 0.27
	10	98.15 ± 0.14
	15	97.75 ± 0.14
	20	97.85 ± 0.21

Table 2: The impact of number of neighbors on the final performance of NEAT.

Experiments

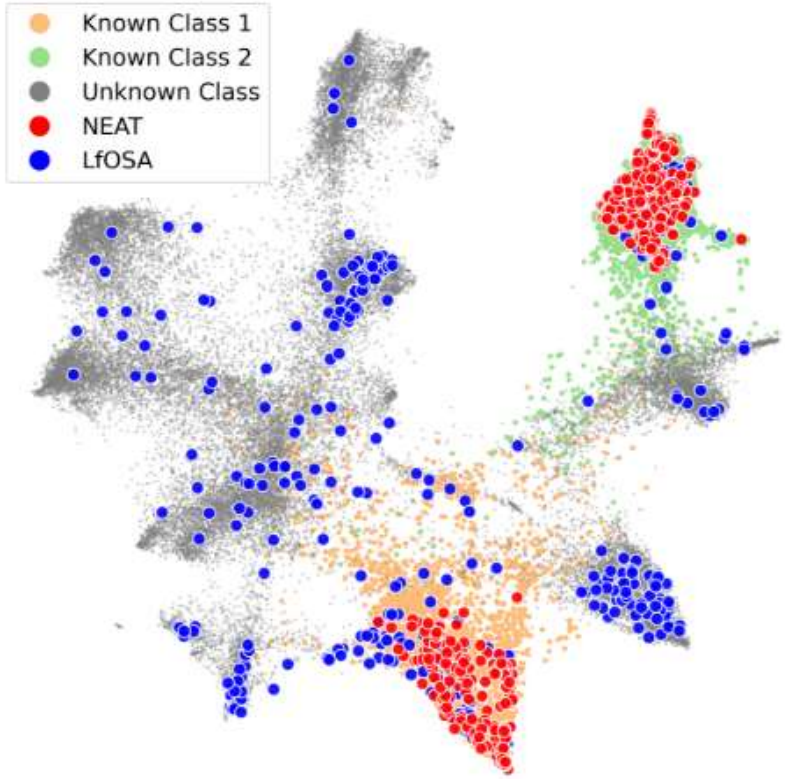


Figure 4: NEAT can accurately identify known classes from the unlabeled pool.

南京航空航天大學

Nanjing University of Aeronautics and Astronautics

Not All Out-of-Distribution Data Are Harmful to Open-Set Active Learning

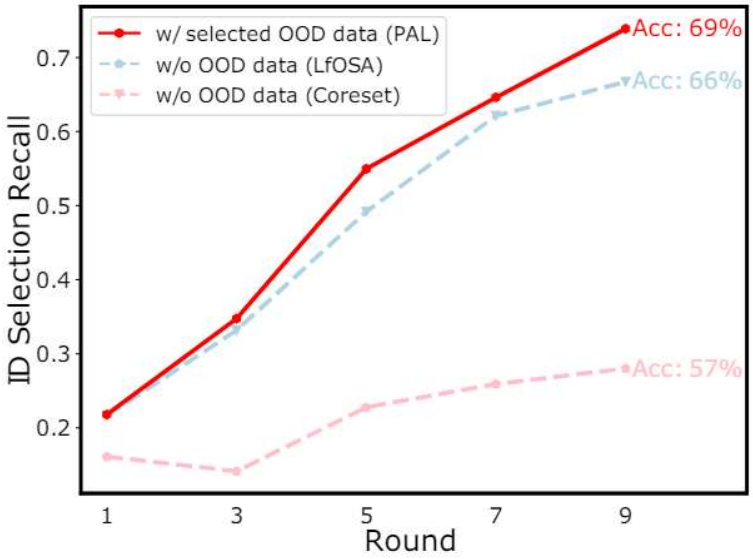
Yang Yang[1], **Yuxuan Zhang**[1], **Xin Song**[2], **Yi Xu**[3*]
[1]Nanjing University of Science and Technology
[2]Baidu Talent Intelligence Center, Baidu Inc
[3]Dalian University of Technology
{yyang, xuan_zhang}@njust.edu.cn, songxin06@baidu.com, yxu@dlut.edu.cn

NeurIPS 2024

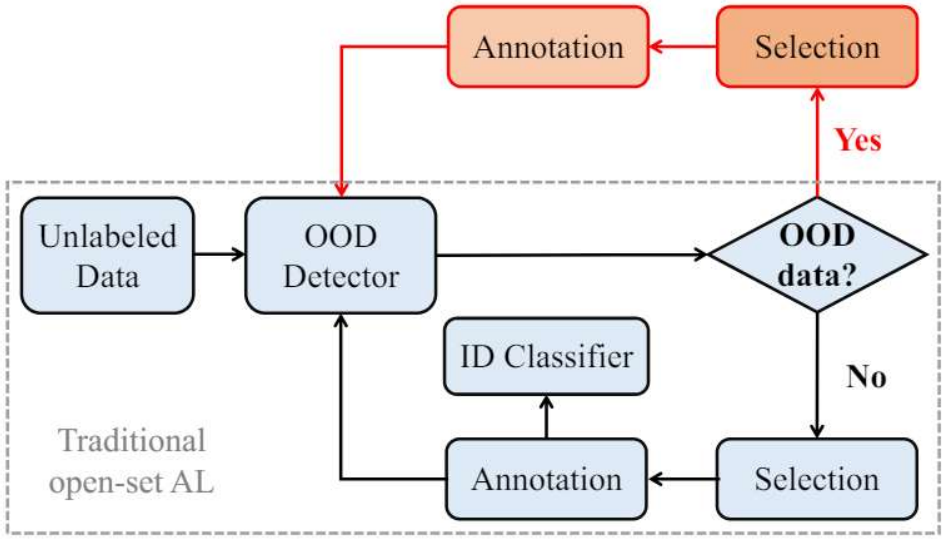
Motivation

Limitations:

- ✓ Existing methods mainly focus on learning the ID classifier, consequently restraining the resolution of the OOD detector.
- ✓ Valuable OOD instances can promptly strengthen the boundary between ID and OOD instances, and further improve the ID purity for classifier retraining.



(*a*) The power of OOD instances.



(*b*) The proposed PAL process

PAL: Progressive Active Learning

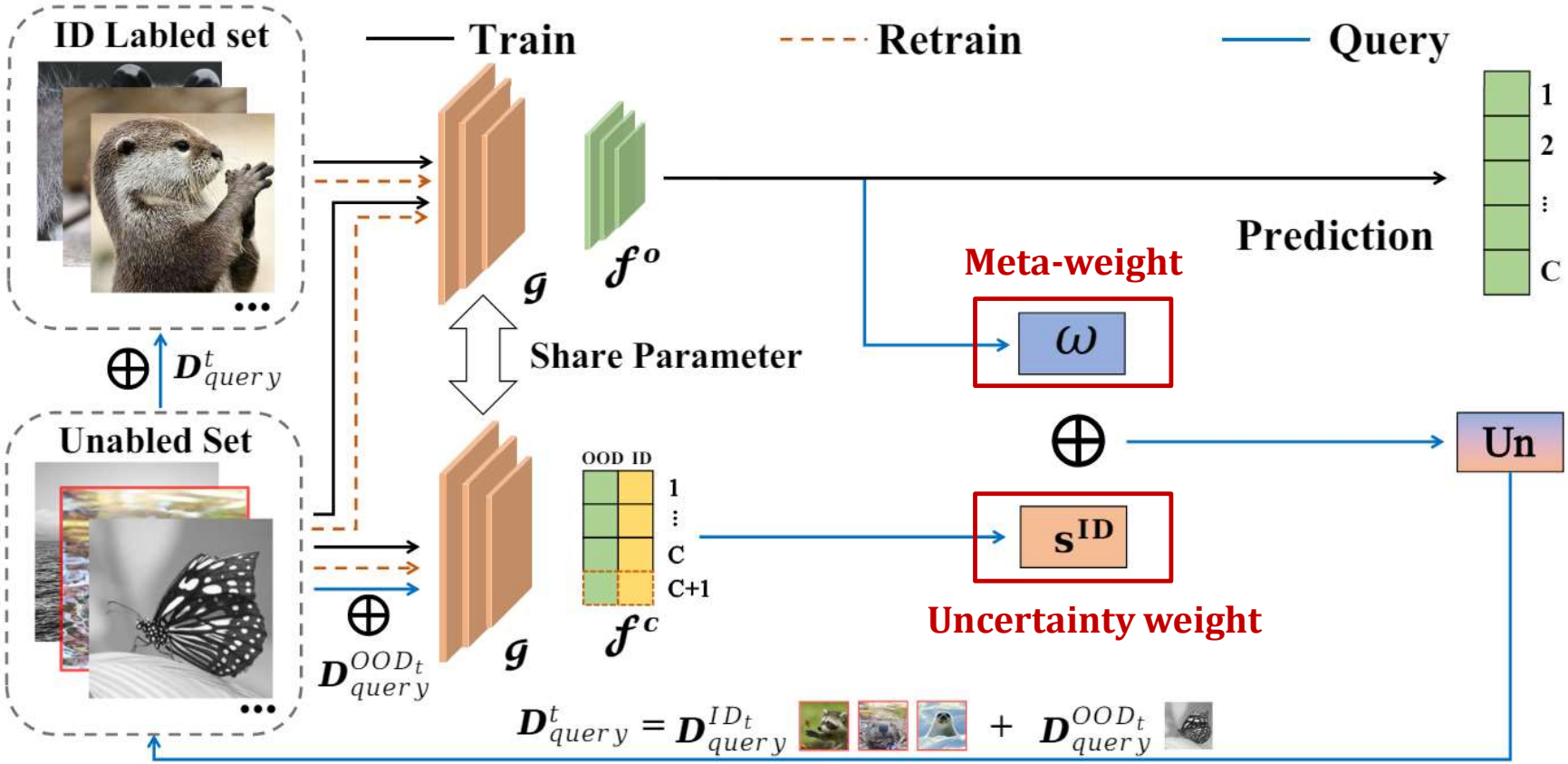
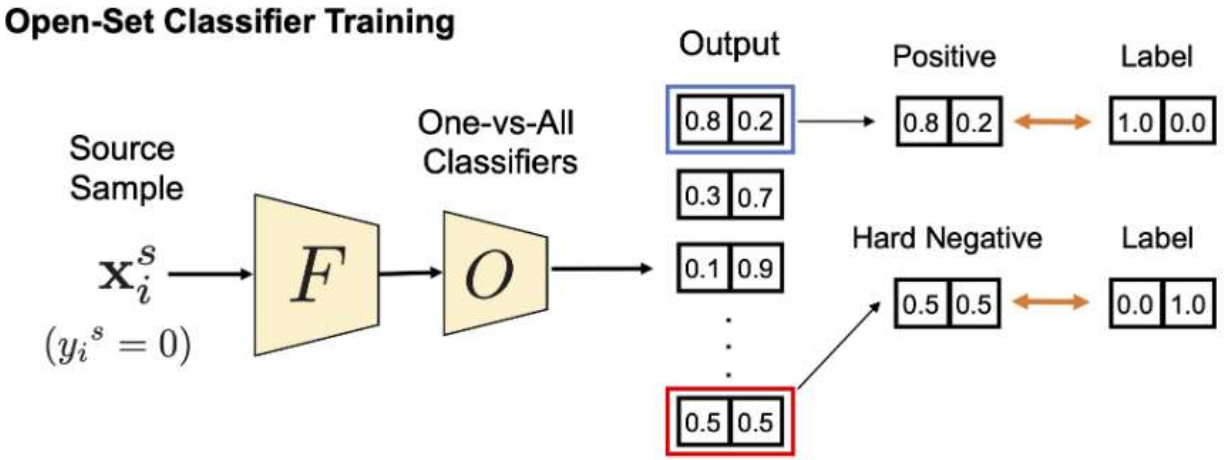


Figure 2: **Framework of the proposed Progressive Active Learning (PAL).** We initialized the ID classifier and OOD detector with the ID labeled data. During the query rounds, we explicitly query both pseudo-ID and pseudo-OOD instances with the designed sampling criterion to enhance the ID classifier and OOD detector.

Informativeness: Uncertainty weight



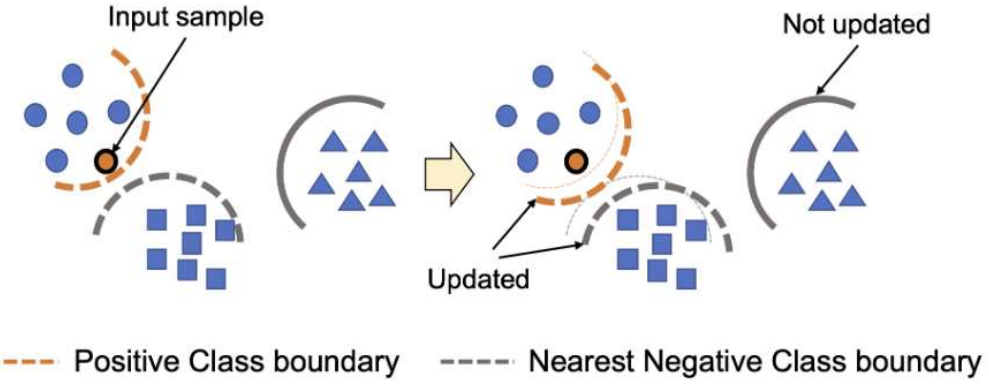


Figure 3: **Overview of the open-set classifier training.**

$$\ell_{OVA} = -\sum_{i=1}^{N_l}(\log \mathbf{p}^{y_{\mathbf{x}_i^l}}(p=1|\mathbf{x}_i^l) + \min_{c \neq y_{\mathbf{x}_i^l}} \log \mathbf{p}^c(p=0|\mathbf{x}_i^l))$$

Uncertainty weight

$$s^{ID} = 1 + \mathbf{p}^{\hat{c}}(p=1|\mathbf{x}^u)\log \mathbf{p}^{\hat{c}}(p=1|\mathbf{x}^u)$$

$$\hat{c} = \arg\max_c \mathbf{p}^c(p=1|\mathbf{x}^u)$$

Representativeness: Meta-weight

- Automatically learn the weight of each unlabeled instance via tracking the effect of the supervised learning model to prevent OOD instances, which can be formulated as the following **bi-level optimization problem**:

$$\begin{aligned} &\min_{\omega} \sum_{i=1}^{N_l} \ell_{sup}(\mathbf{x}_i^l, \mathbf{y}_i^l, \hat{\Theta}(\omega)), \\ &s.t. \quad \hat{\Theta}(\omega) = \arg\min_{\Theta} \sum_{i=1}^{N_l} \ell_{sup}(\mathbf{x}_i^l, \mathbf{y}_i^l, \Theta) + \sum_{j=1}^{N_u} \omega_j \ell_{un}(\mathbf{x}_j^u, \Theta), \end{aligned} \tag{2}$$

where $\Theta = (\Theta_1, \Theta_2)$, $\ell_{sup}(\mathbf{x}_i^l, \mathbf{y}_i^l, \Theta) = -\sum_{c=1}^{C} \mathbf{y}_i^l \log f_{\Theta_1}^{o,c}(g_{\Theta_2}(\mathbf{x}_i^l))$ is the supervised loss, $\ell_{un}(\mathbf{x}_j^u, \Theta) = -\sum_{c=1}^{C} f_{\Theta_1}^{o,c}(g_{\Theta_2}(\mathbf{x}_j^u)) \log f_{\Theta_1}^{o,c}(g_{\Theta_2}(\mathbf{x}_j^u))$ is the unsupervised loss, and f^o denotes the ID classifier.

Progressive Sampling

- OSA aims to precisely select as many ID instances as possible from the unlabeled pool. With the s_{ID} and ω, we select the first b instances with the highest probability as the query set:

$$un(\mathbf{x}^u) = \omega + \mu s^{ID}$$

- The selected b instances including N_{query}^{ID} pseudo-ID instances and N_{query}^{OOD} pseudo-OOD instances.
  - At the first query, OOD instances with the lowest $un(x^u)$ are selected, $N_{query}^{OOD} \gg N_{query}^{ID}$.
  - After that, the OVA classifiers are extended to $C+1$ classifiers, and the $un(x^u)$ is given by

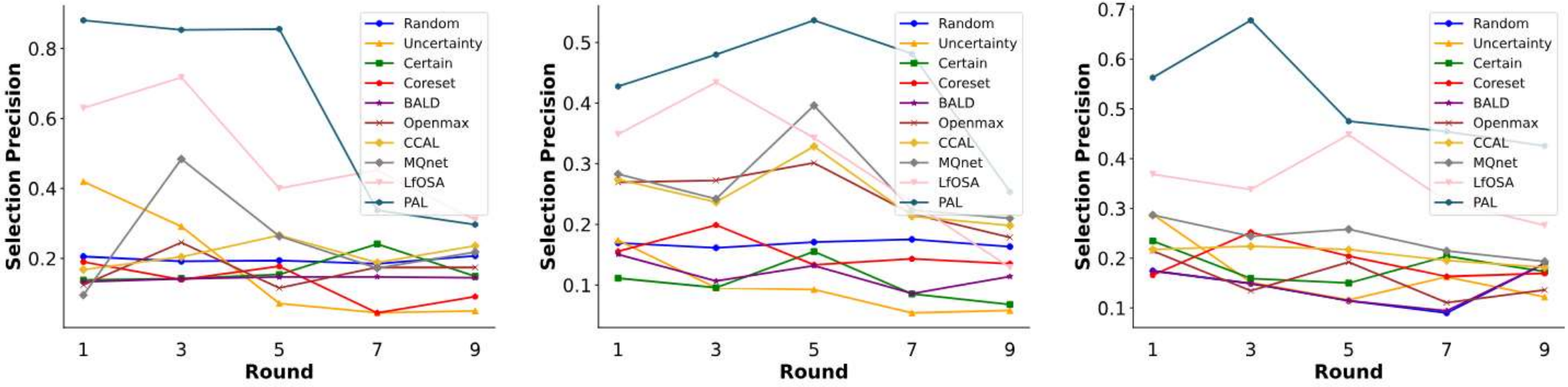
$$un(\mathbf{x}^u) = \begin{cases} \omega + \mu(1 - s^{ID}), & \text{if } \hat{c} = C+1 \\ \omega + \mu s^{ID}, & \text{otherwise} \end{cases}$$

Experiments

Table 1: Comparison of testing accuracy (%) on CIFAR-10, CIFAR-100, and Tiny-Imagenet datasets with an ID proportion of 20%. The best results are highlighted in bold, and the second-best results are underlined.

Datasets	CIFAR-10					CIFAR-100					Tiny-Imagenet				
Rounds	1	3	5	7	9	1	3	5	7	9	1	3	5	7	9
Label Only	78.6					42.5					21.4				
Random	88.1	93.1	95.9	96.9	97.4	44.6	49.2	52.2	54.5	56.7	22.9	26.8	32.7	37.8	39.2
Uncertainty	88.2	93.0	96.1	97.2	97.5	44.4	49.4	49.7	54.8	55.3	21.6	28.5	35.6	37.6	41.1
Certainty	88.1	90.8	91.6	92.4	93.0	45.0	50.7	53.1	54.4	54.9	21.7	27.9	35.4	39.8	44.7
Coreset	86.5	94.7	95.8	96.7	96.8	43.6	49.5	51.9	55.0	57.0	23.1	26.3	33.0	39.9	43.3
BALD	84.6	93.1	94.3	96.4	96.7	42.9	48.1	52.3	54.4	56.2	22.9	26.8	32.9	38.1	39.2
OpenMax	81.3	85.8	86.6	87.0	90.7	45.0	47.3	50.1	53.9	56.0	22.0	26.1	32.2	36.9	41.9
CCAL	88.9	94.3	96.2	97.4	97.7	45.1	50.9	53.4	57.2	60.4	23.2	28.5	35.6	40.6	44.9
MQnet	88.8	94.9	96.8	97.4	97.8	45.3	51.1	57.9	59.1	61.3	23.8	28.6	35.7	42.0	45.8
LfOSA	84.2	95.4	97.1	97.5	98.3	45.6	52.2	59.0	62.5	66.1	22.6	28.8	36.4	43.7	47.9
PAL	**91.1**	**95.6**	**97.6**	**98.5**	**98.7**	**45.6**	**53.0**	**60.0**	**65.6**	**69.4**	**24.3**	**33.4**	**43.9**	**47.6**	**52.1**

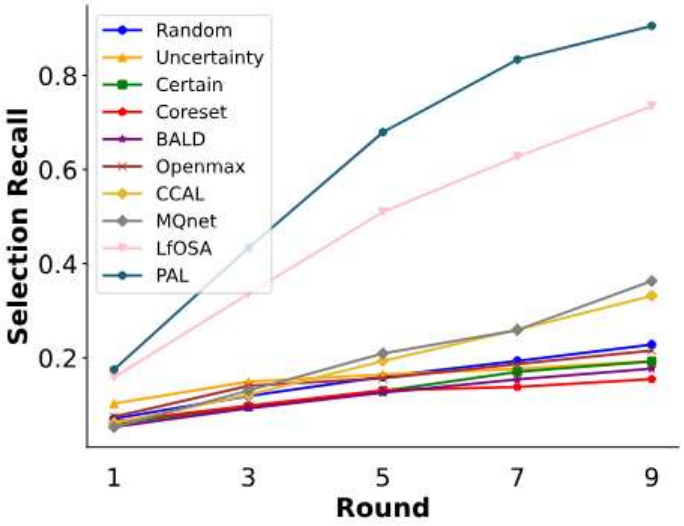
Experiments



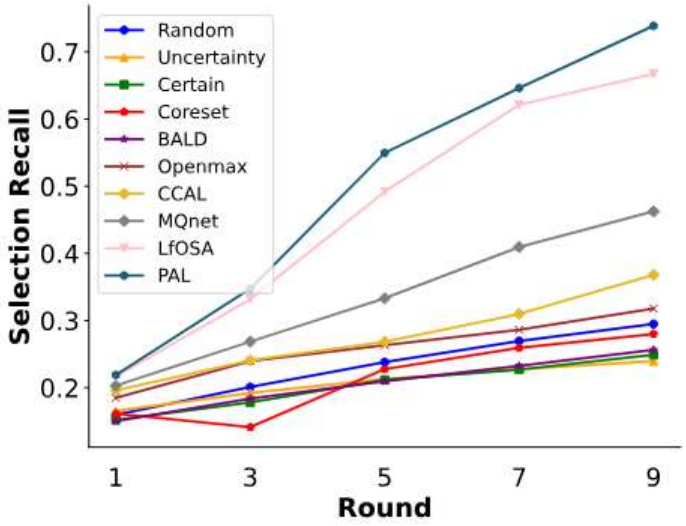
(a) CIFAR-10 Precision (b) CIFAR-100 Precision (c) Tiny-Imagenet Precision

Figure 3: The precision and recall comparisons of OOD detection on CIFAR-10, CIFAR-100, and Tiny-Imagenet datasets with ID proportion of 20%.

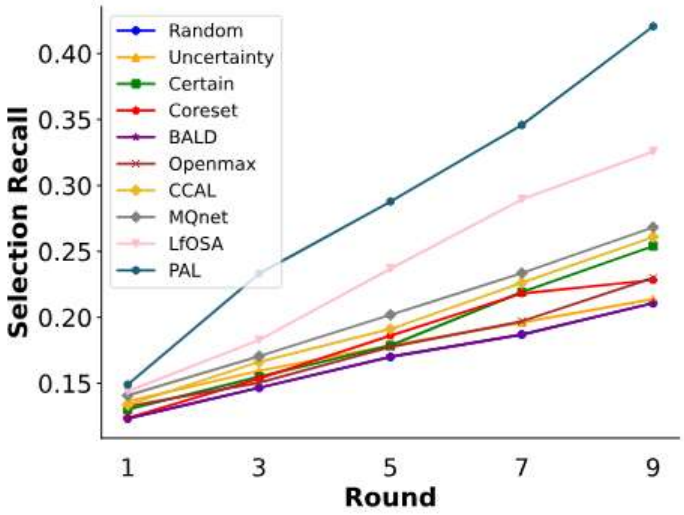
Experiments



(d) CIFAR-10 Recall



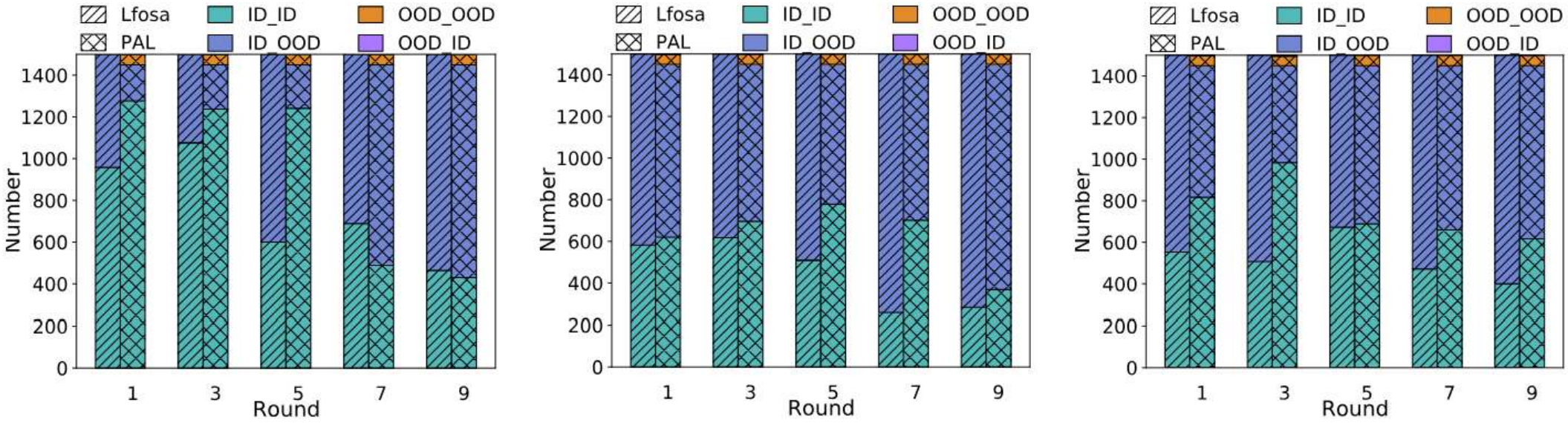
(e) CIFAR-100 Recall



(f) Tiny-Imagenet Recall

Figure 3: The precision and recall comparisons of OOD detection on CIFAR-10, CIFAR-100, and Tiny-Imagenet datasets with ID proportion of 20%.

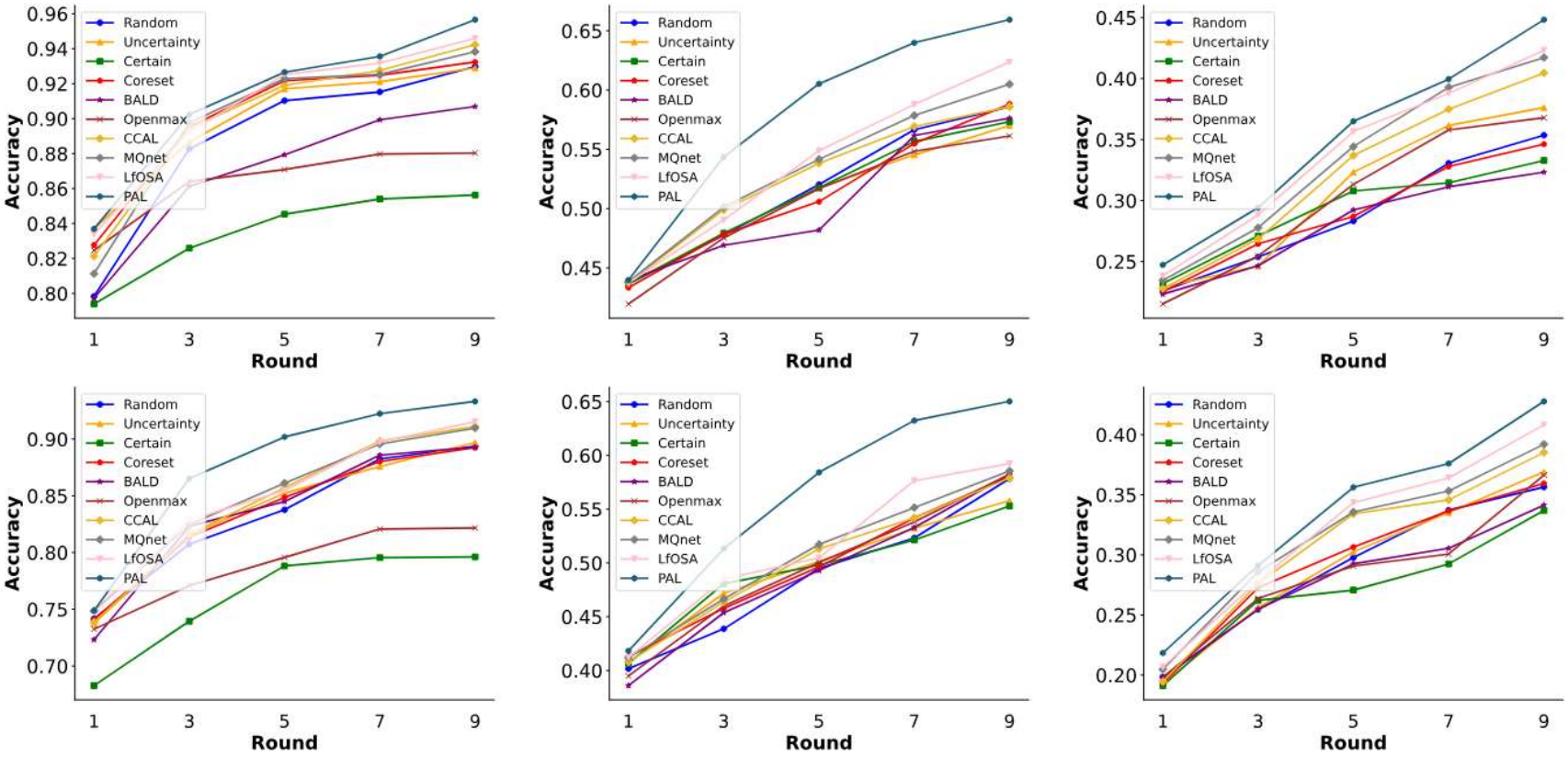
Experiments



(a) CIFAR-10 (b) CIFAR-100 (c) Tiny-Imagenet

Figure 4: The query purity comparison between LfOSA and PAL. Query purity denotes the ratio of ID_ID instances to the total number of the annotation budget. ID_ID denotes the true positive in the pseudo-ID instances, while ID_OOD represents the false positive in the pseudo-ID instances. Similarly, OOD_ID denotes the false negative in the pseudo-OOD instances, while OOD_OOD represents the true negative in the pseudo-OOD instances. Note that LfOSA only queries the pseudo-ID instances.

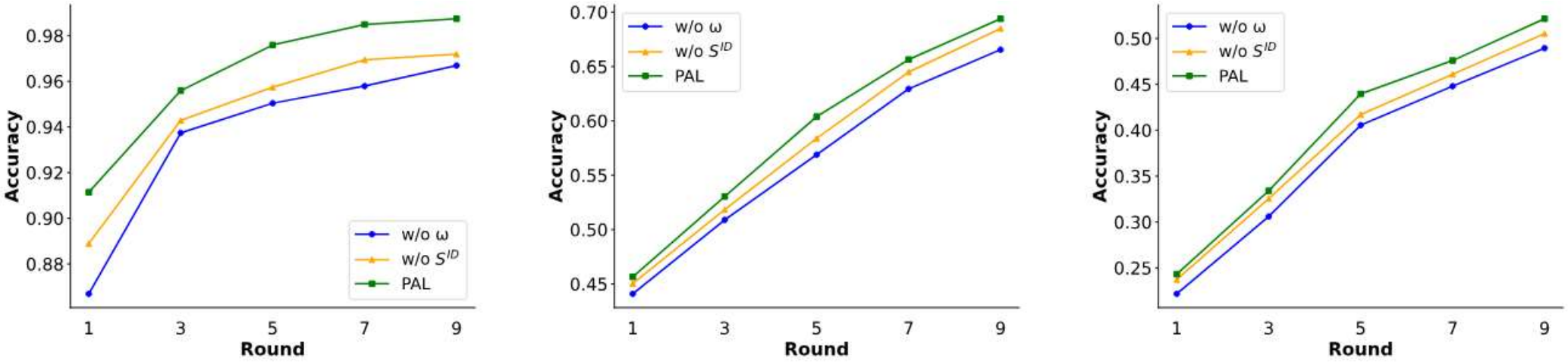
Experiments



(a) CIFAR-10 Accuracy (b) CIFAR-100 Accuracy (c) Tiny-Imagenet Accuracy

Figure 5: Comparison of testing accuracy(%) on CIFAR-10 (first column), CIFAR-100 (second column), and Tiny-Imagenet (third column) datasets, with an ID proportion of 30% (first row) and 40% (second row).

Experiments



(a) CIFAR-10 (b) CIFAR-100 (c) Tiny-Imagenet

Figure 8: Ablation study conducted on CIFAR-10, CIFAR-100, and Tiny-Imagenet datasets with an ID proportion of 20%.

Experiments

Table 2: Comparison of accuracy, precision, and recall across different values of μ in CIFAR-100, with an ID proportion of 20%.

μ	0.2	0.4	0.6	0.8	1
Accuracy (%)	66.2	67.2	68.5	**69.4**	67.6
Precision (%)	**32.4**	27.3	25.6	25.4	17.6
Recall (%)	67.0	69.2	71.0	**73.8**	69.7

Experiments

Table 3: Comparison of testing accuracy (%) for PAL_Core, PAL_MSP, PAL_energy, and PAL on CIFAR-10 and CIFAR-100 with an ID proportion of 20%.

Datasets	**CIFAR-10**					**CIFAR-100**				
Rounds	1	3	5	7	9	1	3	5	7	9
PAL_Core	87.1	93.3	96.7	97.4	98.3	44.8	49.7	59.6	63.7	65.9
PAL_MSP	86.0	95.1	95.2	97.5	98.0	45.1	48.4	55.8	60.1	64.7
PAL_energy	85.6	93.7	94.7	97.0	97.9	45.32	52.5	58.6	64.5	66.2
PAL	**91.1**	**95.6**	**97.6**	**98.5**	**98.7**	**45.6**	**53.0**	**60.0**	**65.6**	**69.4**

THANKS