Revisiting Unknowns: Towards Effective and Efficient Open-Set Active Learning

Supplementary Material

A. Full Expression of Equation (4)

The KL divergence term $\mathcal{L}_{\mathrm{KL}}$ between the predicted Dirichlet distribution and the uniform prior can be derived as:

$$
\begin{aligned}
&\mathcal{L}_{\mathrm{KL}} = \mathrm{KL}\big(\mathrm{Dir}(\tilde{\boldsymbol{\alpha}}) \parallel \mathrm{Dir}(\mathbf{1})\big) \\
&= \int \mathrm{Dir}(\boldsymbol{p};\tilde{\boldsymbol{\alpha}}) \log \frac{\mathrm{Dir}(\boldsymbol{p};\tilde{\boldsymbol{\alpha}})}{\mathrm{Dir}(\boldsymbol{p};\mathbf{1})} d\boldsymbol{p} \\
&= \int \left(\frac{1}{\mathbb{B}\left(\tilde{\boldsymbol{\alpha}}\right)} \prod_{i=1}^{k+\hat{u}} p_i^{\tilde{\alpha}_i - 1} \right) \log \left(\frac{\mathbb{B}(\mathbf{1})}{\mathbb{B}\left(\tilde{\boldsymbol{\alpha}}\right)} \prod_{i=1}^{k+\hat{u}} p_i^{\tilde{\alpha}_i - 1} \right) d\boldsymbol{p} \\
&= \log \frac{\mathbb{B}(\mathbf{1})}{\mathbb{B}\left(\tilde{\boldsymbol{\alpha}}\right)} \int \frac{1}{\mathbb{B}\left(\tilde{\boldsymbol{\alpha}}\right)} \prod_{i=1}^{k+\hat{u}} p_i^{\tilde{\alpha}_i - 1} d\boldsymbol{p} \\
&+ \int \left(\log \prod_{i=1}^{k+\hat{u}} p_i^{\tilde{\alpha}_i - 1} \right) \left(\frac{1}{\mathbb{B}\left(\tilde{\alpha}\right)} \prod_{i=1}^{k+\hat{u}} p_i^{\tilde{\alpha}_i - 1} \right) d\boldsymbol{p} \\
&= \log \frac{\mathbb{B}(\mathbf{1})}{\mathbb{B}\left(\tilde{\boldsymbol{\alpha}}\right)} + \mathbb{E}_{\boldsymbol{p}\sim\mathrm{Dir}(\tilde{\boldsymbol{\alpha}})} \left[\log \textstyle\prod_{i=1}^{k+\hat{u}} p_i^{\tilde{\alpha}_i - 1} \right] \\
&= \log \frac{\mathbb{B}(\mathbf{1})}{\mathbb{B}\left(\tilde{\boldsymbol{\alpha}}\right)} + \sum_{j=1}^{C} \left(\tilde{\alpha}_j - 1\right) \mathbb{E}_{\boldsymbol{p}_j\sim\mathcal{B}\left(\tilde{\alpha}_j, \sum_{i\neq j}\tilde{\alpha}_i\right)} \left[\log p_j\right] \\
&= \log \left[\frac{\Gamma\left(\sum_{i=1}^{k+\hat{u}} \tilde{\alpha}_i\right)}{\Gamma(k+\hat{u}) \prod_{i=1}^{k+\hat{u}} \Gamma\left(\tilde{\alpha}_i\right)} \right] \\
&+ \sum_{j=1}^{k+\hat{u}} \left(\tilde{\alpha}_j - 1\right) \left[\psi\left(\tilde{\alpha}_j\right) - \psi\left(\sum_{i=1}^{k+\hat{u}} \tilde{\alpha}_i \right) \right],
\end{aligned}
\tag{9}
$$

where $\mathbb{B}(\cdot)$ denotes the multivariate Beta function, $\mathcal{B}(\cdot,\cdot)$ is the standard Beta function, $\Gamma(\cdot)$ represents the Gamma function, and $\psi(\cdot)$ denotes the digamma function. The explicit definitions of $\mathbb{B}(\cdot)$ and $\mathcal{B}(\cdot,\cdot)$ are given by:

$$
\mathbb{B}(\tilde{\boldsymbol{\alpha}}) = \frac{\prod_{i=1}^{k+\hat{u}} \Gamma(\tilde{\alpha}_i)}{\Gamma\left(\sum_{i=1}^{k+\hat{u}} \tilde{\alpha}_i\right)}, \tag{10}
$$

and

$$
\mathcal{B}\left(\tilde{\alpha}_j, \sum_{i\neq j} \tilde{\alpha}_i\right) = \frac{\Gamma\left(\tilde{\alpha}_j\right) \Gamma\left(\sum_{i\neq j} \tilde{\alpha}_i\right)}{\Gamma\left(\tilde{\alpha}_j + \sum_{i\neq j} \tilde{\alpha}_i\right)}. \tag{11}
$$

B. The Pseudocode of E²OAL

We outline the full workflow of E^{2}OAL in Algorithm B1, which covers each stage of the active learning process.

C. Additional Results for Figure 1

We present additional results over a wider range of mismatch ratios and network architectures beyond those in Figure 1. As illustrated in Figure C1, we evaluate three sampling strategies to assess whether labeled `unknowns` can enhance known-class learning: (1) random sampling; (2) certainty-based sampling via maximum softmax probability (MSP [11]); and (3) uncertainty-based sampling using the least-confidence criterion [18]. Our observations indicate that, regardless of sampling strategy, mismatch ratio, or network architecture, leveraging fine-grained labels of labeled `unknowns` within the auxiliary classifier consistently yields the best performance—especially as network capacity increases. Treating all labeled `unknowns` as a single class typically improves over ignoring them but results in less stable and less substantial gains, occasionally even degrading performance. These findings highlight the significant benefits of effectively exploiting labeled `unknowns` during training, motivating the in-depth study and method proposed in this work.

D. Additional Results for Figure 3

Figure D1 illustrates the evolution of test accuracy across rounds under intermediate mismatch ratios on three benchmark datasets. E^{2}OAL consistently surpasses all baselines, exhibiting a clearly superior accuracy trajectory throughout the process. Notably, the performance gap widens as the dataset complexity increases, highlighting the robustness and scalability of the proposed framework.

E. Additional Results for Figure 4

Figure E1 reports the mean test accuracy and average query precision across rounds under moderate mismatch ratios on three benchmark datasets. E^{2}OAL achieves the highest query precision on CIFAR-100 and Tiny-ImageNet, while consistently outperforming all competing methods in accuracy. On CIFAR-10, the query precision is slightly lower than EOAL [27], which is expected since our purity control is explicitly regulated by a target precision of 0.6. This observation underscores two key insights: (1) Simply maximizing query purity is not optimal for open-set active learning—although EOAL attains high precision, its queried samples, albeit from known classes, tend to be less informative, leading to suboptimal accuracy; (2) Although E^{2}OAL does not achieve the absolute highest query precision, it adheres more closely to the target value than EAOA [41], which is also guided by the target precision, demonstrating

Algorithm B1 The E^{2}OAL algorithm

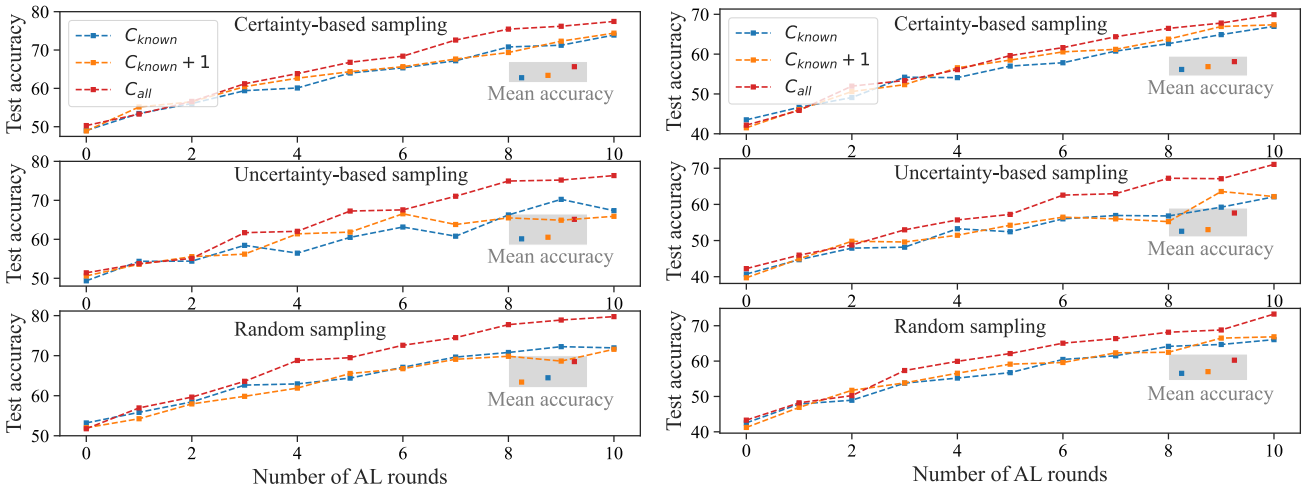
Input: Labeled pool $\mathcal{D}_L = \mathcal{D}_L^{kno} \cup \mathcal{D}_L^{unk}$, unlabeled pool $\mathcal{D}_U$, known class count k, upper limit $\hat{u}_{max}$, target classifier f_θ, query budget $|\mathcal{B}|$, and target query precision p^*.

Process: (The t-th active learning round)

if $t = 1$ **then**
  # Model training
  Train classifier f_θ using $\mathcal{L}_{\text{CE}}$ for primary head (k-way) and $\mathcal{L}_{\text{EDL}}$ for auxiliary head (k-way)
  # Purity-based candidate selection
  Obtain auxiliary head's outputs for all $x \in \mathcal{D}_L \cup \mathcal{D}_U$
  for each $x \in \mathcal{D}_L \cup \mathcal{D}_U$ **do**
    Compute purity margin: $S_{\text{purity}}(x) = \max_{c \in \mathcal{C}_k} o_c$
  end for
else
  # Unknown class estimation
  Estimate $\hat{u}$ based on labeled pool $\mathcal{D}_L$, known class count k, and upper bound $\hat{u}_{\max}$ using Algorithm 1
  # Model training
  Train classifier f_θ using $\mathcal{L}_{\text{CE}}$ for primary head (k-way) and $\mathcal{L}_{\text{EDL}}$ for auxiliary head ($(k+\hat{u})$-way)
  # Purity-based candidate selection
  Obtain auxiliary head's outputs for all $x \in \mathcal{D}_L \cup \mathcal{D}_U$.
  for each $x \in \mathcal{D}_L \cup \mathcal{D}_U$ **do**
    Compute purity margin: $S_{\text{purity}}(x) = \max_{c \in \mathcal{C}_k} o_c - \max_{c \in \mathcal{C}_{\hat{u}}} o_c$
  end for
end if
Fit a 3-component Gaussian Mixture Model (GMM) [24] on logit margins $\{S_{\text{purity}}(x)\}$ to model different purity regimes: high (known), low (unknown), and intermediate (ambiguous)
For each $x \in \mathcal{D}_U$, compute its likelihood under the high-purity component
Sort $\mathcal{D}_U$ in descending order of the computed likelihoods
Initialize the candidate pool $\mathcal{C}_{pool}$ with the top $|\mathcal{B}|$ samples from the sorted $\mathcal{D}_U$
 # Precision-based candidate refinement
Compute the calibrated target query precision $\hat{p}_t^* = \begin{cases} \max\left(\min(\hat{p}_{t-1}^* + (p^* - \bar{p}_{t-1}^*), 0), 1\right) & \text{if } t > 1 \\ p^* & \text{if } t = 1 \end{cases}$
while the mean likelihood of the lowest $|\mathcal{B}|$ samples in $\mathcal{C}_{\text{pool}}$ is greater than $\hat{p}_t^*$ **do**
  Add the next highest-likelihood sample from $\mathcal{D}_U$ into $\mathcal{C}_{\text{pool}}$
end while
 # Information-based final query selection
for each $x \in \mathcal{C}_{\text{pool}}$ **do**
  Obtain predicted probability vector $\mathbf{p}$ of x from the primary head
  Let $\mathbf{u}$ be a uniform distribution over all classes
  Let $\mathbf{p}^{\max}$ be a one-hot vector with 1 at $\arg\max(\mathbf{p})$
  Compute information score: $S_{\text{info}}(x) = \text{JS}\big(\mathbf{p} \parallel \mathbf{u}\big) \cdot \text{JS}\big(\mathbf{p} \parallel \mathbf{p}^{\max}\big)$
end for
Select top $|\mathcal{B}|$ samples from $\mathcal{C}_{pool}$ with the highest $S_{\text{info}}(x)$ scores as the query set $\mathcal{B}$
Compute observed query precision: $\bar{p}_t^* = \frac{|\mathcal{B}^{kno}|}{|\mathcal{B}|}$
Update data pools: $\mathcal{D}_U \leftarrow \mathcal{D}_U \setminus \mathcal{B}$, $\mathcal{D}_L \leftarrow \mathcal{D}_L \cup \mathcal{B}$
return $\mathcal{D}_L$, $\mathcal{D}_U$, $\bar{p}_t^*$, and f_θ for the next round

superior control and stability in purity regulation.

Figure E2 further illustrates how query precision evolves over rounds. It can be observed that our method maintains a high query precision even in the early training stages—a desirable property, as highlighted in [21], where high-purity samples tend to provide greater utility when the model is still undertrained. This advantage stems from our proposed purity metric and flexible sampling strategy, which effectively constructs high-purity candidate sets. In contrast, the previous state-of-the-art method, EAOA, shows subop-



(a) ResNet-18, 20 known / 80 unknown classes

(b) ResNet-18, 30 known / 70 unknown classes

(c) ResNet-34, 30 known / 70 unknown classes

(d) ResNet-50, 30 known / 70 unknown classes

Figure C1. Per-round and mean test accuracy on CIFAR-100 under varying class splits and network architectures. C_{known} excludes labeled `unknowns`, $C_{known+1}$ collapses them into a single class, and C_{all} leverages their ground-truth labels.

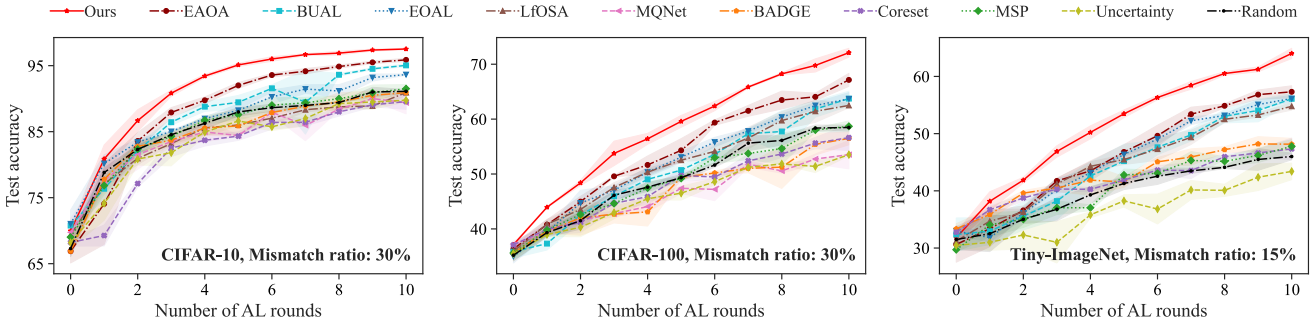


Figure D1. Test accuracy across rounds under mismatch ratio 30% on CIFAR-10/100 and 15% on Tiny-ImageNet.

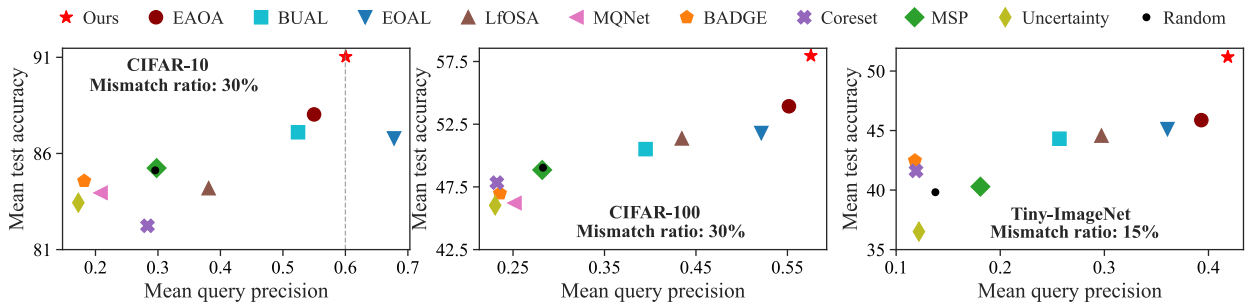


Figure E1. Mean query precision and test accuracy under mismatch ratio 30% on CIFAR-10/100 and 15% on Tiny-ImageNet.

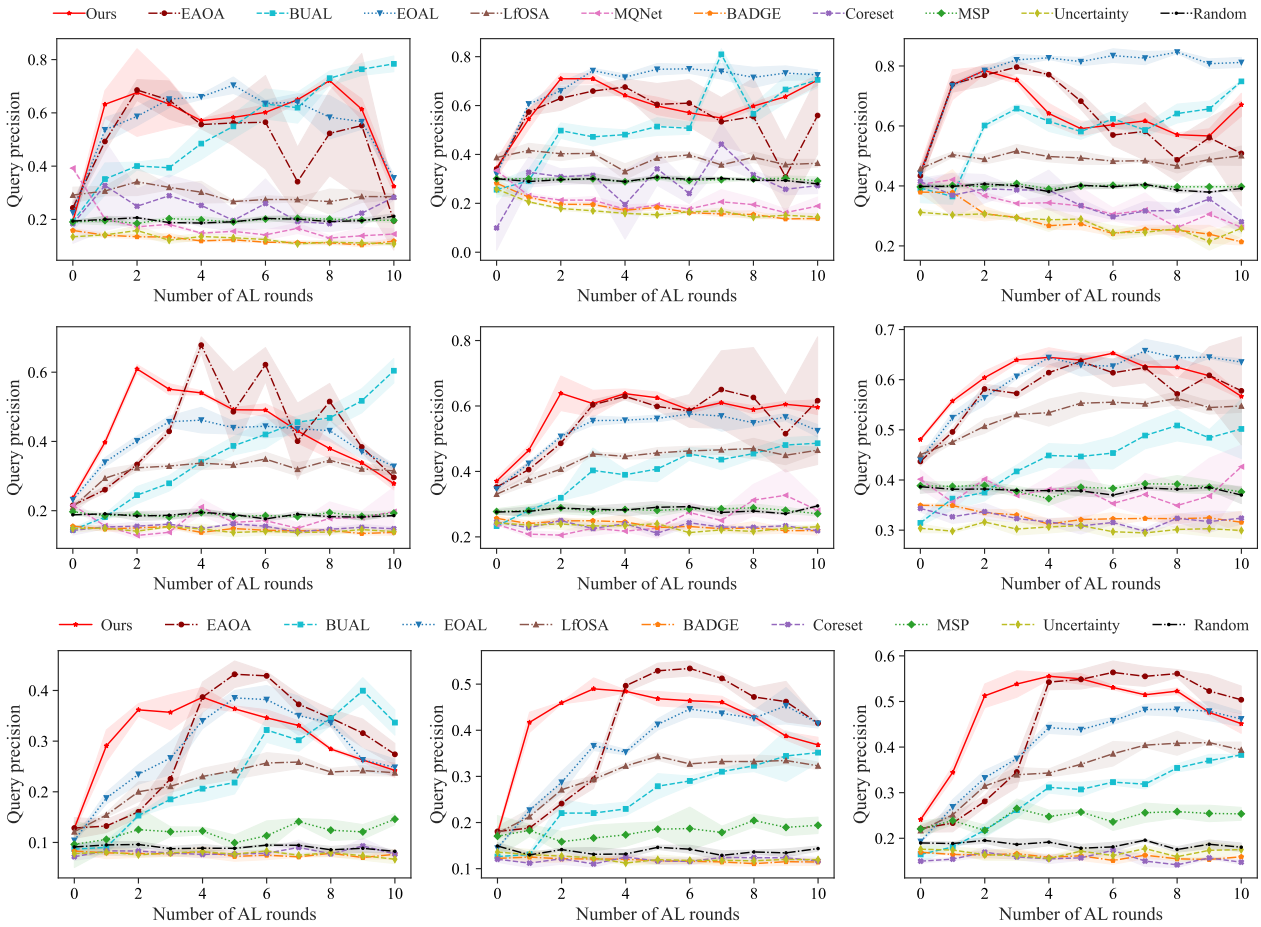


Figure E2. Cross-round query precision under varying mismatch ratios on three benchmarks. From top to bottom: results on CIFAR-10, CIFAR-100, and Tiny-ImageNet; from left to right: mismatch ratios of 20%, 30%, and 40%, respectively.

timal query precision in the early rounds and exhibits significant fluctuations, especially on CIFAR-100 with a 20% mismatch ratio. This instability arises from its inherent limitation in the adaptive sampling strategy, where the adjustment of the parameter k is limited to a fixed step size per round, hindering rapid convergence to an optimal value. Our method, by comparison, introduces no additional hyperparameters and achieves stable, high-precision queries from the outset, demonstrating clear superiority.

F. Additional Results for Table 1

Table F1 summarizes the final-round test accuracy of all methods under different mismatch ratios across three

Dataset	**CIFAR-10**			**CIFAR-100**			**Tiny-ImageNet**		
Mismatch ratio	20%	30%	40%	20%	30%	40%	10%	15%	20%
Random	94.48	91.11	87.18	64.45	58.48	55.48	47.93	46.00	43.32
Uncertainty [18]	95.70	89.77	83.61	62.25	53.52	50.83	45.83	43.40	35.43
Coreset [28]	94.20	89.56	86.38	63.53	56.62	55.00	50.60	47.33	45.35
BADGE [2]	94.95	90.91	87.12	64.00	56.49	50.20	49.70	48.16	46.23
MSP [11]	94.15	91.51	87.21	65.33	58.69	56.68	51.43	47.78	46.57
LfOSA [20]	94.15	90.91	87.43	70.32	62.49	58.49	58.37	54.78	51.33
MQNet [21]	95.12	89.39	87.42	63.70	53.52	55.44	-	-	-
EOAL [27]	96.23	93.64	91.63	73.73	63.69	59.55	61.40	56.13	52.65
BUAL [42]	96.48	95.04	92.52	73.43	63.73	59.89	63.80	56.09	50.52
EAOA [41]	97.23	95.88	93.09	74.60	67.14	63.49	62.33	57.31	53.33
Ours*	97.33	95.94	93.13	75.90	67.54	63.85	64.23	60.44	54.73
↑ over best baseline (%)	0.10	0.06	0.04	1.30	0.40	0.36	1.15	3.13	1.40
Ours	**98.77**	**97.52**	**95.69**	**82.20**	**72.10**	**67.98**	**68.53**	**64.02**	**57.10**
↑ over best baseline (%)	1.44	1.64	2.60	7.60	4.96	4.49	4.73	6.71	3.77

Table F1. Final-round test accuracy (%) of all methods under varying mismatch ratios on CIFAR-10, CIFAR-100, and Tiny-ImageNet. "Ours*" denotes a variant of our method where the target classifier is trained independently without leveraging labeled `unknowns`. The best result in each setting is highlighted in bold, while the second best is underlined. Due to the poor performance and high training cost of MQNet, we do not include it on Tiny-ImageNet, and thus mark it with "–".

benchmark datasets. We report results for both "Ours" (the proposed E^{2}OAL) and "Ours*" (a variant where the target classifier is trained independently without utilizing labeled `unknowns`, similar to prior baselines). Both E^{2}OAL and its variant consistently outperform existing methods by a notable margin. The strong performance of "Ours*" demonstrates the effectiveness of our adaptive sampling strategy in selecting more informative known-class samples, while the additional improvement achieved by E^{2}OAL further highlights the value of leveraging labeled `unknowns` to enhance known-class learning.

G. Ablation on CLIP Representations

Feature source	CIFAR-10	CIFAR-100	Tiny-ImageNet
CLIP	**97.52**	72.10	**64.02**
MoCo	97.44	**72.31**	63.87

Table G1. Final-round test accuracy (%) under fixed mismatch ratios (30% for CIFAR-10/100 and 15% for Tiny-ImageNet). "CLIP" and "MoCo" respectively refer to adaptive class estimation performed using features extracted from a CLIP model and from a self-supervised MoCo pretrained model.

Table G1 presents the final-round results when the default CLIP features are replaced with self-supervised MoCo pretrained representations for adaptive class estimation. E^{2}OAL exhibits stable performance across all datasets, with only marginal differences between the two feature sources. These results demonstrate that E^{2}OAL is robust to the choice of pretrained representation, and its effectiveness does not depend on a specific feature extractor. In practice, CLIP can be substituted with any pretrained model capable of providing high-quality, task-agnostic features.

H. Ablation on $\hat{u}$ Estimation

Figure H1 illustrates the estimated total number of classes, $k + \hat{u}$, across rounds on CIFAR-100 under mismatch ratios of 20%, 30%, and 40%, where k denotes the number of known classes (20, 30, and 40, respectively). For reference, the figure also includes the final number of queried samples per class. Our adaptive class estimation module consistently yields stable and reliable estimates across rounds and mismatch settings. Since the true class prior of unknown samples is inaccessible, the estimated total number of classes may not exactly match the ground truth. This deviation partly arises from the inherent ambiguity of class granularity—for instance, CIFAR-100 can be organized into either 20 coarse-grained or 100 fine-grained categories. Nevertheless, our method consistently provides estimates within the correct order of magnitude, with accuracy improving as the mismatch ratio increases (i.e., as the open-set problem becomes less challenging). In particular, under the 40% setting, the estimated class count closely fluctuates around the

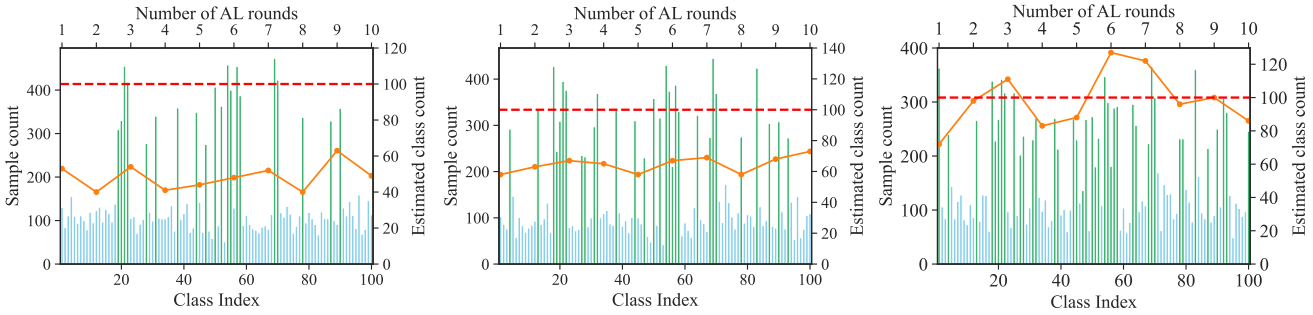


Figure H1. Ablation results for unknown class estimation on CIFAR-100 under mismatch ratios of 20%, 30%, and 40% (from left to right). Bar charts (bottom x-axis and left y-axis) show the total number of samples labeled per class in the final round. Green bars represent known classes, and blue bars represent unknown classes. Line plots (top x-axis and right y-axis) illustrate the evolution of the estimated total number of classes ($k + \hat{u}$), where the ground-truth total is 100.

true value, highlighting the effectiveness and robustness of the proposed class estimation strategy.