Active Learning for Open-set Annotation

Kun-Peng Ning[1*]   Xun Zhao[2†]   Yu Li[3*]   Sheng-Jun Huang[1†]

[1]Nanjing University of Aeronautics and Astronautics

[2]Applied Research Center, Tencent PCG   [3]International Digital Economy Academy (IDEA)

{ningkp,huangsj}@nuaa.edu.cn   emmaxunzhao@tencent.com   liyu@idea.edu.cn

Abstract

Existing active learning studies typically work in the closed-set setting by assuming that all data examples to be labeled are drawn from known classes. However, in real annotation tasks, the unlabeled data usually contains a large amount of examples from unknown classes, resulting in the failure of most active learning methods. To tackle this open-set annotation (OSA) problem, we propose a new active learning framework called LfOSA, which boosts the classification performance with an effective sampling strategy to precisely detect examples from known classes for annotation. The LfOSA framework introduces an auxiliary network to model the per-example max activation value (MAV) distribution with a Gaussian Mixture Model, which can dynamically select the examples with highest probability from known classes in the unlabeled set. Moreover, by reducing the temperature T of the loss function, the detection model will be further optimized by exploiting both known and unknown supervision. The experimental results show that the proposed method can significantly improve the selection quality of known classes, and achieve higher classification accuracy with lower annotation cost than state-of-the-art active learning methods. To the best of our knowledge, this is the first work of active learning for open-set annotation.

1. Introduction

The remarkable success of deep learning is largely attributed to the collection of large datasets with human annotated labels [12, 15]. Nevertheless, it is extremely expensive and time-consuming to label large scale data with high-quality annotations [26, 29]. It is thus a significant challenge to learn with limited labeled data.

Active learning (AL) is a primary approach to tackle this problem. It iteratively selects the most useful examples from the unlabeled data to query their labels from the oracle,

---

*This work is done when Kun-Peng Ning and Yu Li worked at Applied Research Center, Tencent PCG.
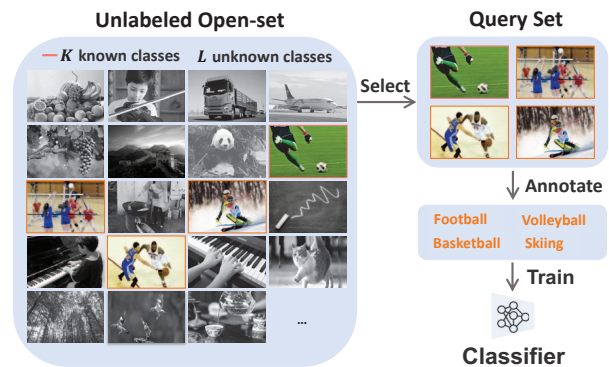
†Corresponding authors.

Figure 1. The illustration of open-set annotation (OSA) problem. The unlabeled open-set contains K known classes (color images with border) and L unknown classes (gray-scale images without border). The goal is to find and annotate the examples from known classes for training the classifier.

achieving competitive performance while reducing annotation costs [10, 26, 28]. Existing AL methods typically work in a closed-set setting where the labeled and unlabeled data are both drawn from the same class distribution.

However, in some real world scenarios, the unlabeled data are usually uncontrolled and large amounts of data examples are from unknown classes. Figure 1 displays an example of training a new sports image classification model for an image-sharing social platform, in which the database contains trillions of images from unconstrained categories uploaded by users. A large proportion of images in the unlabeled open-set are actually from irrelevant classes (*e.g.*, cats, pianos, *etc.*). As these irrelevant images are not necessary for training the desired classifier, labelling these images would lead to a waste of annotation budget. On the other hand, existing closed-set AL system cannot precisely distinguish these irrelevant images from unknown classes but tends to choose them for annotation as they contain more uncertainty or information. Hence, in the real-world open-set scenario, an effective and practical AL system is highly desired, which can precisely identify the examples of un-

wanted classes while querying the most useful cases from the wanted classes to train the classifier.

In this paper, we formulate this problem as an open-set annotation (OSA) task. As shown in Figure 1, the unlabeled set contains K known classes and L unknown classes, where $L > K$. The goal is to precisely filter out examples from unknown classes, while actively select a query set that contains examples from known classes as pure as possible. To overcome this challenge, we propose a new active learning framework called LfOSA (**L**earning **f**rom **O**pen-**S**et **A**nnotation), which includes two networks for detection and classification respectively. Specifically, the detector models the per-example max activation value (MAV) distribution with a Gaussian Mixture Model [21] to dynamically divide the unlabeled open-set into known and unknown set, then examples from the known set with larger certainty will be selected to construct a query set for annotation. After labeling, the classification model will be updated with the new examples from known classes. Meanwhile, as the query set will inevitably include a few invalid examples from unknown classes, these invalid examples will be utilized as negative training examples to update the detector, such that the detector can maintain a higher recall to identify known-class examples from the unlabeled open-set. Moreover, by reducing the temperature T of the cross-entropy (CE) loss, the distinguishability of the detector is further enhanced.

Experiments are conducted on multiple datasets with different mismatch ratios of known and unknown classes. The experimental results demonstrate that the proposed approach can significantly improve the selection quality of known classes, and achieve higher classification accuracy with lower annotation cost than state-of-the-art active learning methods.

The rest of this paper is organized as follows. We review the related work in Section 2 and introduce the proposed method in Section 3. Section 4 reports the experiments, followed by the conclusion in Section 5.

2. Related Work

Active learning. Active learning as a large-scale annotation tool has achieved a great success for learning with limited labeled data [8, 20]. It reduces the labeling cost by selecting the most useful examples to query their labels. Most researches focus on designing effective sampling strategies to make sure that the selected examples can improve the model performance most [3]. During the past decades, many criteria have been proposed for selecting examples [3,5,9,16,19,22,27,32]. Among of these approaches, some of them prefer to select the most informative examples to reduce the model uncertainty [16,27,32], while some others prefer to select the most representative examples to match the data distribution [5,22]. Moreover, some studies try to combine informativeness and representativeness to achieve better performance [9,11]. These standard active learning

methods are usually based on the closed-set assumption that the unlabeled data are drawn from known classes, which leads to failure on the open-set annotation (OSA) task.

Open-set recognition. Open-set recognition (OSR) attempts to address the classification setting where inference can face examples from unseen classes during training [13,23,24]. Its goal is to learn an open-set classifier with a mechanism to reject such unknown examples [6]. A representative approach called OpenMax has achieved remarkable results on the OSR problem, which employs deep neural networks to OSR by combining Extreme Value Theory with neural networks [1]. Another follow-up work proposed G-OpenMax by adopting GAN [2,7] for generating examples which are highly similar to training examples yet do not belong any of the training classes [4]. Moreover, [17] has been some efforts to use active learning in the open-set recognition. However, these OSR methods usually fail on the OSA problem for the following two essential differences between both. First, the training process of OSR has abundant labeled data and is based on the closed-set assumption, while the OSA problem has limited labeled data and its unlabeled data are open set. Second, the OSR focuses on rejecting unknown examples in testing phase after training, while the OSA aims to find more known examples from the unlabeled open-set for target model optimizing in training phase.

3. The Proposed Approach

In this section, we first formalize the open-set annotation (OSA) problem, and then introduce the proposed LfOSA approach in detail.

3.1. The OSA Problem Setting

In the OSA problem, we consider a large-scale annotation scenario with a limited labeled set D_L and an enormous unlabeled open-set D_U, where $D_L = \{(x_i^L, y_i^L)\}_{i=1}^{n^L}$ and $D_U = \{x_j^U\}_{j=1}^{n^U}$. Let $D_U = X_{kno} \cup X_{unk}$ and $X_{kno} \cap X_{unk} = \emptyset$, where X_{kno} and X_{unk} denote the examples from known and unknown classes respectively. Each labeled example x_i^L belongs to one of K known classes $Y = \{y_k\}_{k=1}^K$, while an unlabeled example x_j^U may belong to an unknown class not belonging to Y. Let X^{query} denotes the query set during each iteration, which consists of unknown query set X_{unk}^{query} and known query set X_{kno}^{query}, *i.e.*, $X^{query} = X_{kno}^{query} \cup X_{unk}^{query}$. The goal is to selectively construct the query set that contains known examples as many as possible.

Active learning (AL) iteratively selects the most useful examples from the unlabeled dataset to query their labels from the oracle [26]. After annotating the newly selected data, the model can be updated to achieve better performance. Specifically, in the i-th iteration, we train a classifier f_{θ_C} with parameters θ_C on labeled set D_L. Then, a batch of b examples X^{query} are selected with a specific criterion based
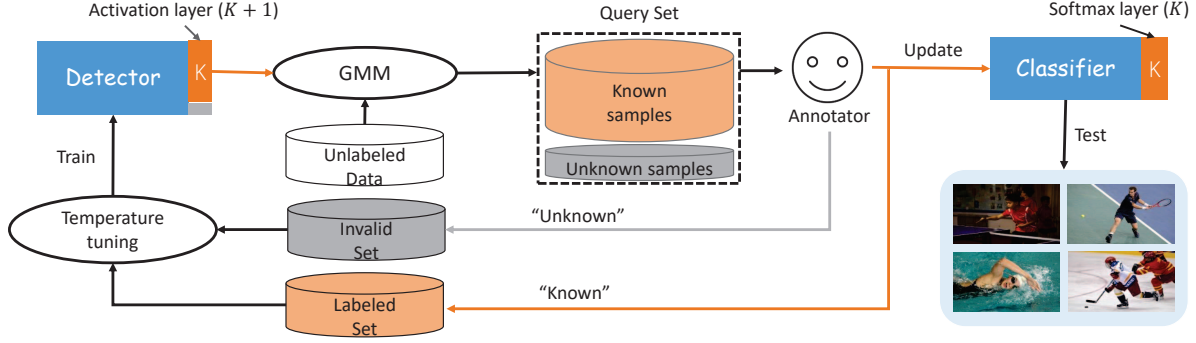
Figure 2. The framework of LfOSA. It includes two networks for detection and classification. The detector attempts to construct a query set for annotation by GMM modeling. After labeling, two networks will be updated for next iteration.

on the current trained model. After querying their labels, k_i known examples X_{kno}^{query} are annotated and the labeled set is updated to $D_L = D_L \cup X_{kno}^{query}$, while l_i examples X_{unk}^{query} with unknown classes are added to the invalid set D_I, where $b = k_i + l_i$. Thus, the recall and precision of known classes in the i-th selection can be defined as follow,

$$recall_i = \frac{\sum_{j=0}^{i} k^i}{n_{kno}}, \tag{1}$$

$$precision_i = \frac{k_i}{k_i + l_i}, \tag{2}$$

where n_{kno} denotes the number of examples from known classes in the unlabeled set. $recall_i$ calculates how many known examples are queried after i queries, and $precision_i$ denotes the proportion of the target examples in the i-th query. Obviously, if we maintain a high precision and recall to accurately select known examples, the trained target classifier will be more effective.

As discussed in the Introduction, most of the traditional AL methods are less effective in OSA problem, because their selection strategies tend to select open-set (unknown) examples with larger uncertainty. These examples from unknown classes are useless for training the target model, and thus traditional AL methods will probably fail with serious waste of the annotation budget. Fortunately, we should be aware that although these examples are useless for the target model, they could be exploited to improve the detector model for filtering out unknown classes from the open-set data. Moreover, we find that the activation (penultimate) layer of network has strong ability to distinguish unknown classes based on the observation that the maximum activation value (MAV) of open set examples are often far away from the average MAV of closed set examples. By decoupling detection and classification, we propose to exploit examples of both known and unknown classes to train a detector with strong distinguishability meanwhile train a more effective classifier for the target task.

3.2. Algorithm Detail

The framework of LfOSA is demonstrated in Figure 2, which mainly composed of three components: detector training, active sampling and classifier training. Specifically, we first train a network for detecting unknown examples by exploiting both known and unknown supervision while using a low-temperature mechanism. Then, by modeling per-example max activation value (MAV) distribution with a Gaussian Mixture Model (GMM), the most certain known examples can be actively selected for annotation. Finally, the classification model will be updated with the new examples from known classes. In the following part of this section, we will introduce these three components in detail.

Detector training. In addition to classifying K known classes, the detector has been extended with an additional $(K + 1)$-th output to predict unknown class. For a given example x from labeled or invalid set, we encode its label c with one-hot p, *i.e.*, the value of p_c is set to 1 and the others to 0. Then, we train the detector with the following cross-entropy loss:

$$\mathcal{L}_D(x, c) = -\sum_{c=1}^{K+1} p_c * log(q_c^T), \tag{3}$$

where

$$q_c^T = \frac{exp(a_c/T)}{\sum_j exp(a_j/T)},$$

where a_c is the c-th activation value of the last fully-connected layer, T is a temperature, which is set with a lower value ($T = 0.5$) to produce a sharper probability distribution q_c^T over classes. Obviously, by minimizing the loss function, examples of known classes will have larger activation values on the first K dimensions and smaller activation values on the $(K + 1)$-th dimension, while examples of unknown classes have the opposite phenomenon. Moreover, we find that the distinguishability of the activation layer can be further enhanced by reducing the temperature T of the

loss function. A brief analysis is as follows:

$$\frac{\partial \mathcal{L}_D}{\partial a_c} = \frac{1}{T}(q_c^T - p_c) = \frac{1}{T}(\frac{exp(a_c/T)}{\sum_j exp(a_j/T)} - p_c). \quad (4)$$

When we reduce the temperature $(T \downarrow)$ of the loss function $\mathcal{L}_R$, the probability distribution q_c^T will be more sharper, thus we have:

$$T \downarrow \Rightarrow \frac{1}{T} \uparrow, \frac{exp(a_c/T)}{\sum_j exp(a_j/T)} - p_c \uparrow \Rightarrow \frac{\partial \mathcal{L}_D}{\partial a_c} \uparrow.$$

As $\frac{\partial \mathcal{L}_R}{\partial a_c}$ becomes larger, the examples of known and unknown classes will be more distinguishable for the activation value a_c.

Active sampling. As mentioned earlier, the goal of OSA task is to precisely select as many known-class examples as possible from the unlabeled open-set. After training the detector as shown above, we find that the activation (penultimate) layer of network has the ability to distinguish unknown examples, that is, the maximum activation value (MAV) of unknown-class examples are often significantly different from the average MAV of known-class examples. Formally, for each unlabeled example x_i with predicted class c, its maximum activation value mav_i^c can be defined as follow:

$$mav_i^c = \max_c a_c^i. \quad (5)$$

All unlabeled examples will be classified into $K + 1$ classes according to the prediction of the current detector. We can select the examples predicted as the first K known classes for the next process while filtering out the examples predicted as "unknown". Then, for each known class c, we fit a two-component GMM (one for known classes and the other for unknown classes) to mav^c using the Expectation-Maximization algorithm, where mav^c is a set of activation values with prediction class c.

$$\mathcal{W}^c = GMM(mav^c, \theta_D), \quad (6)$$

where $\mathcal{W}^c$ is the probabilities of class c. For each unlabeled example x_i from class c, its known probability $w_i \in \mathcal{W}^c$ is the posterior probability $p(g|mav_i)$, where g is the Gaussian component with larger mean (larger activation value). Then we merge and sort the probabilities of all categories,

$$\mathcal{W} = sort(\mathcal{W}^1 \cup \mathcal{W}^2 \cup ... \cup \mathcal{W}^K). \quad (7)$$

Next, we select the first b examples with highest probability as the query set to ask for annotation. In other words, we can obtain the query set X^{query} by setting a threshold τ on w_i, where τ is equal to the b-th largest known probability:

$$X^{query} = \{(x_i, w_i)|w_i \geq \tau, \forall(x_i, w_i) \in (D_U, \mathcal{W})\}. \quad (8)$$

After querying their labels, the labeled and unknown sets will be updated by adding X_{kno}^{query} and X_{unk}^{query}, respectively.

Algorithm 1 The LfOSA algorithm

1: **Input:**
2:     Current detector f_{θ_D} and classifier f_{θ_C}
3:     Current labeled set D_L and invalid set D_I
4:     Query batch size b and temperature T
5: **Process:**
6:     *# Recognizer training*
7:     Update θ_D by minimizing $\mathcal{L}_D$ in Eq. 3 from D_L and D_I
8:     *# Examples sampling*
9:     Inference mav_i^c from detector θ_D for each unlabeled example x_i as Eq. 5
10:     **while** $c = 1, 2, ..., K$ **do**
11:         *# Collect the MAV set for each prediction class c*
12:         $mav^c = \{mav_i^c|f_{\theta_D}(x_i) = c, \forall x_i \in D_U\}$
13:         *# Obtain known probability by GMM*
14:         $\mathcal{W}^c = GMM(mav^c, \theta_D)$
15:     **end**
16:     *# Merge and sort the probability sets of all classes*
17:     $\mathcal{W} = sort(\mathcal{W}^1 \cup \mathcal{W}^2 \cup ... \cup \mathcal{W}^K)$
18:     *# Obtain the query set*
19:     $X^{query} = \{(x_i, w_i)|w_i \geq \tau, \forall(x_i, w_i) \in (D_U, \mathcal{W})\}$

20:     *# Ask for annotation from Oracle*
21:     Query their labels and obtain X_{kno}^{query} and X_{unk}^{query}
22:     *# Update labeled and invalid sets*
23:     $D_L = D_L \cup X_{kno}^{query}, D_I = D_I \cup X_{unk}^{query}$
24:     *# Classifier training*
25:     Update θ_C by minimizing $\mathcal{L}_C$ in Eq. 9 from D_L
26: **Output:** θ_D, θ_C, D_L and D_I for next iteration.

Classifier training. Based on the current labeled data D_L, we train the K-class classifier by minimizing the standard cross-entropy loss:

$$\mathcal{L}_C(x_i, y_i) = -\sum_{i=1}^{n^L} y_i * log(f(x_i; \theta_C)), \quad (9)$$

where $(x_i, y_i) \in D_L$, and n^L is the size of current D_L.

The process of the approach is summarized in Algorithm 1. Firstly, a small set of labeled data D_L, query batch size b and temperature T are given. Then the detector θ_D and classifier θ_C are randomly initialized, and the invalid set D_I is initialized as an empty set. At each iteration, we train the detector by minimizing Eq. 3 to inference mav_i^c for all unlabeled examples. Next, for each class, we collect the MAV set by the predictions of the detector and model per-example MAV to obtain known probabilities. After that, by merging and sorting these probabilities, the first b examples with highest probability are selected as the query set to ask for annotation. As a result, the classifier θ_C, labeled and invalid sets can be updated and output for the next iteration.
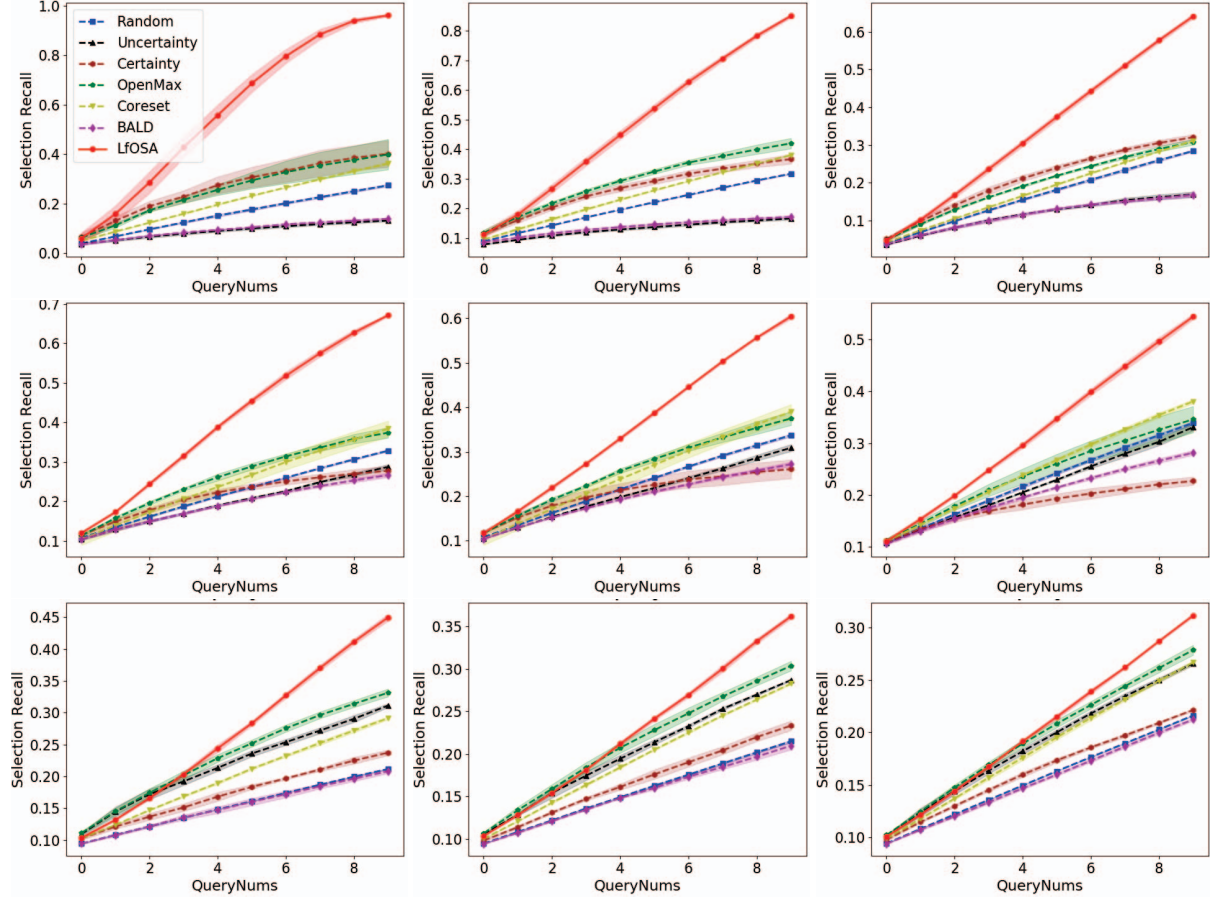
Figure 3. Selection recall comparison on CIFAR10 (first row), CIFAR100 (second row) and Tiny-Imagenet (third row) with 20% (first column), 30% (second column) and 40% (third column) mismatch ratio.

4. Experiments

To validate the effectiveness of the proposed approach, we perform experiments on CIFAR10, CIFAR100 [14] and Tiny-Imagenet [31] datasets, which contains 10, 100, 200 categories respectively. To construct open-set datasets, we set mismatch ratio as 20%, 30% and 40% for all our experiments, where the mismatch ratio denotes the proportion of the number of known classes in the total number of classes. For example, when the mismatch ratio is set as 20%, on CIFAR10, CIFAR100 and Tiny-Imagenet, the first 2, 20, 40 classes are known classes for classifier training, and the last 8, 80, 160 classes are seen as unknown classes respectively.

Baselines. To validate the effectiveness of the proposed LfOSA approach, we compare the following methods in the experiments. *i)* **Random**: it randomly selects examples from unlabeled pool for labeling. *ii)* **Uncertainty** [16, 18]: it selects the examples with largest uncertainty of predictions for annotation. *iii)* **Certainty** [16, 18]: it selects the examples with largest certainty of predictions for annotation. *iv)* **Coreset** [25]: it selects the representative examples by diver-

sity. *v)* **BALD** [30]: it uses dropout as an approximation to Bayesian inference for active sampling. *vi)* **OpenMax** [1]: a representative open-set recognition approach. *vii)* **LfOSA (ours)**: the proposed approach.

Active learning setting. For all AL methods, we randomly sampling 1%, 8% and 8% examples as initialization labeled set on CIFAR10, CIFAR100 and Tiny-Imagenet datasets, that is, each category contains only 50, 40 and 40 examples respectively. It is worth to note that the labeled sets only contain known classes. In each AL cycle, we train a ResNet18 model for 100 epochs, SGD is adopted as the optimizer with momentum 0.9, weight decay 5e-4, initialization learning rate 0.01, and batch size of 128, while a batch of 1500 examples is selected to query their labels for the next AL round.

Performance measurement. We compare the proposed LfOSA approach with other compared methods in the performance of selection recall (as Eq. 1), precision (as Eq. 2) and classification accuracy. Moreover, we perform the experiments for 4 runs and record the average results over 4 seeds ($seed = 1, 2, 3, 4$).
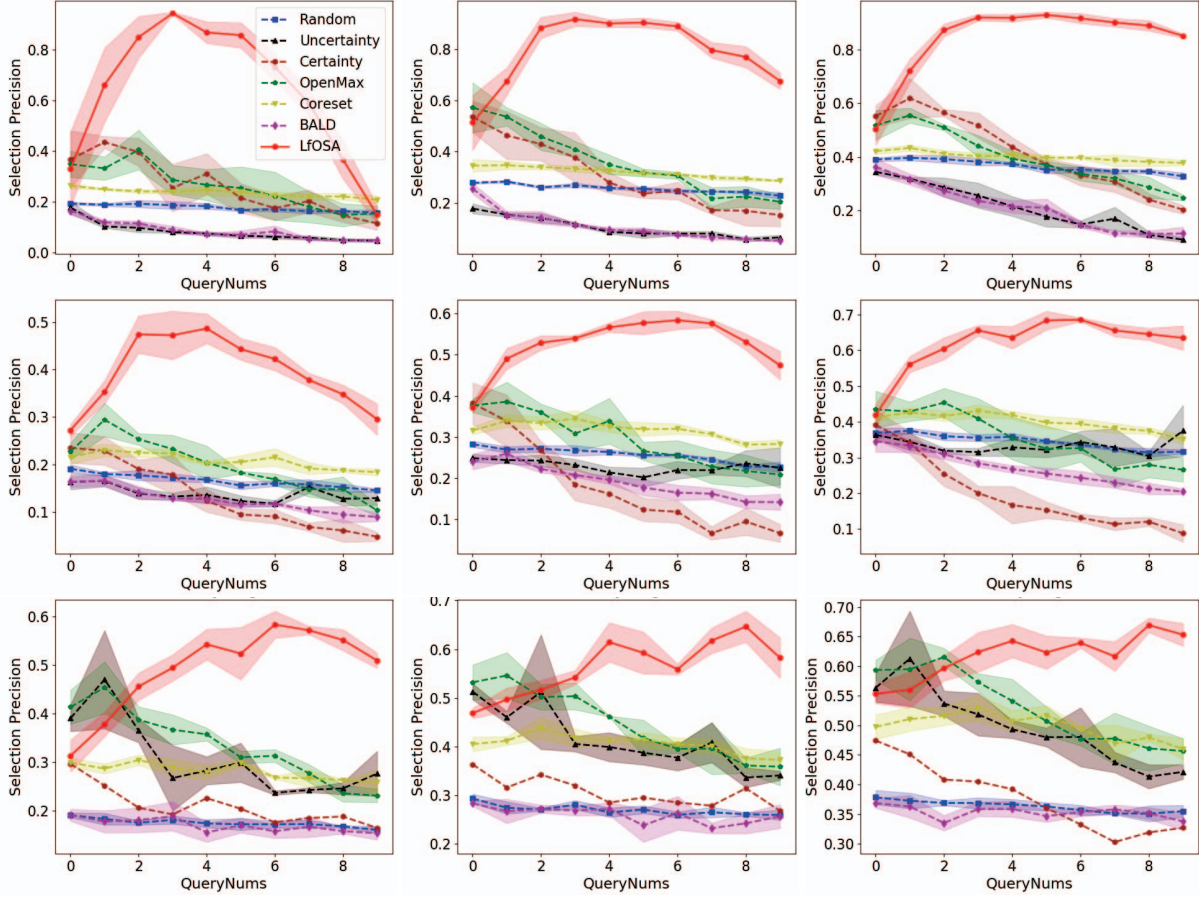
Figure 4. Selection precision comparison on CIFAR10 (first row), CIFAR100 (second row) and Tiny-Imagenet (third row) with 20% (first column), 30% (second column) and 40% (third column) mismatch ratio.

4.1. Performance Comparison

We evaluate the performance of the proposed LfOSA and compared methods by plotting curves with the number of queries increasing. The average results of recall, precision, accuracy are demonstrated in Figure 3, 4 and 5 respectively. The first, second and third rows represent the results on CIFAR10, CIFAR100 and TinyImagenet respectively. The first, second and third columns represent the results with 20%, 30% and 40% mismatch ratio.

It can be observed that no matter which dataset or mismatch ratio is used, the proposed LfOSA approach always outperforms other methods in all cases. LfOSA can achieve higher selection recall and precision during the AL process, while achieving better classification performance. *i*) For the performance of recall, the proposed LfOSA approach consistently outperforms other compared methods by a significant margin. Especially on CIFAR10 and CIFAR100, when the mismatch ratio is set to 20%, 30% and 40%, the average margins between the LfOSA and Random methods are 68.8%, 53.4% and 35.7% in the former and 34.3%, 26.7% and 20.5% in the later. *ii*) For the performance of precision, the pro-

posed LfOSA approach always maintain a higher selection precision than other baselines with a clear gap. It worth to note that adding invalid examples can significantly improve the detection ability (the precision of the first three queries is improving). Besides, as the number of known examples decreases, the precision is forced to decrease (the precision of the 10-th query is only 20% on "CIFAR10 with 20% mismatch ratio" because its recall has reached 96.7%). *iii*) For the performance of classification, LfOSA consistently exhibits the best performance in all cases. Especially on CIFAR100, compared to other AL methods, LfOSA achieves about 20%, 15% and 12% performance improvement under the 20%, 30%, 40% mismatch ratios respectively. Moreover, with the increase of unknown ratio, the superiority of LfOSA over the other methods becomes more significant. These results indicate that the proposed LfOSA method can effectively solve the open-set annotation (OSA) problem.

Compared methods analysis. It is interesting to observe that two popular AL methods, Uncertainty and BALD, perform worse even than the random method in most cases. One possible reason is that these informativeness-based AL
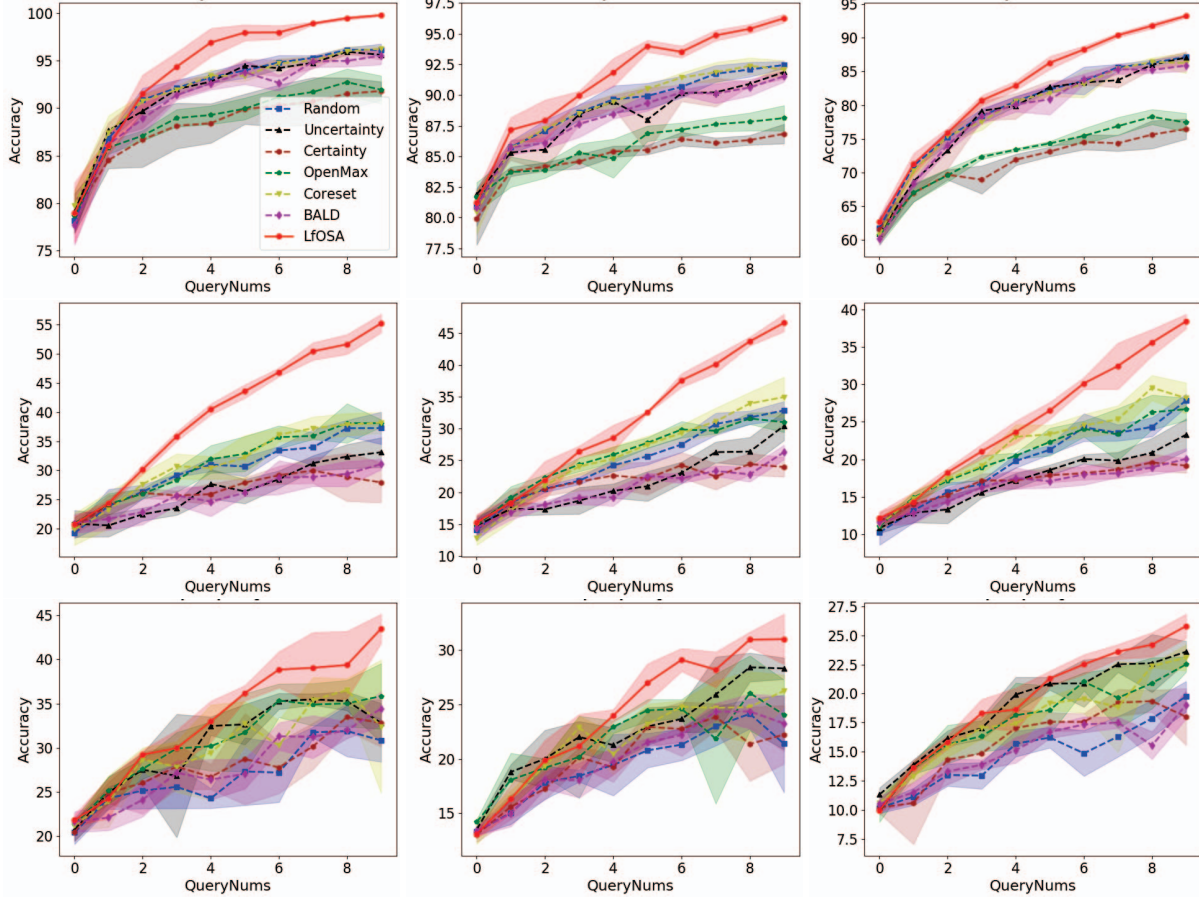
Figure 5. Classification performance comparison on CIFAR10 (first row), CIFAR100 (second row) and Tiny-Imagenet (third row) with 20% (first column), 30% (second column) and 40% (third column) mismatch ratio.
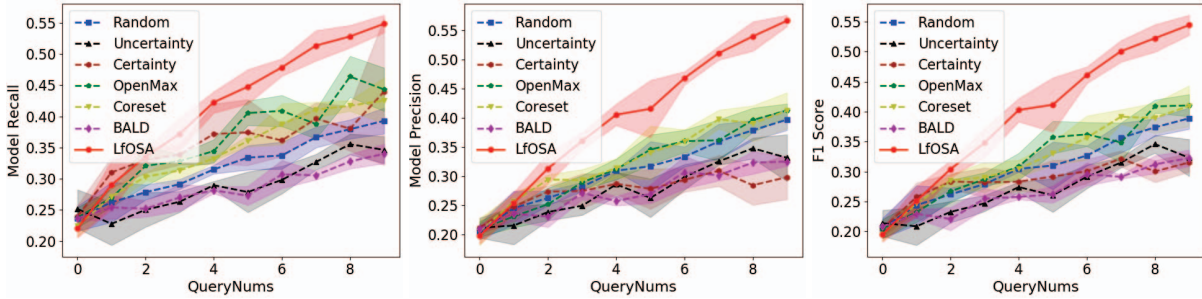

Figure 6. Classfication recall (first column), precision (second column), F1 (third column) performance comparison on CIFAR100 with 20% mismatch ratio.

methods tend to select unknown classes, because these unknown examples are more likely to be the most informative ones. On the other hand, the Certainty method also fails in the OSA problem, which means it may not be accurate to measure the certainty of examples by using the model's prediction entropy. The diversity-based Coreset method and the open-set recognition method OpenMax show limited effectiveness in the OSA tasks. The former has no recognition

ability for unknown classes, and the latter lacks sufficient supervision information.

4.2. Results Using More Metrics

The comparisons of classification recall, precision, and F1. To further validate the effectiveness of the proposed LfOSA approach, we compare with other methods in terms of classification recall, precision, and F1 on CIFAR100 with
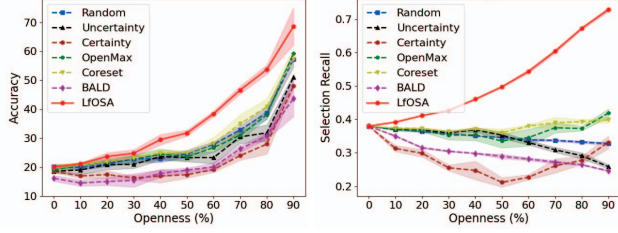
Figure 7. Accuracy / Recall (after 10 queries) *v.s.* openness on CIFAR100.

Figure 8. Accuracy / Recall (after 10 queries) *v.s.* different b on CIFAR100 with 20% mismatch ratio.

20% mismatch ratio. The experimental results are demonstrated in Figure 6. It can be observed that the proposed LfOSA approach always significantly outperforms other methods in all cases. LfOSA can achieve higher classification recall, precision, and F1 score.

The comparison of time complexity. The time complexity of LfOSA is $o(kn)$, where n denotes the number of unlabeled data, and k denotes the number of known classes. We have measured the time costs of different methods for one query on CIFAR100. As shown in Table 1, LfOSA is one of the most efficient methods.

Table 1. The comparison of time complexity.

Random	Certainty	Openmax	Coreset	BALD	LfOSA
$\sim 0s$	$23s$	$29s$	$165s$	$182s$	$26s$

4.3. Ablation Study

To further analyze the proposed LfOSA approach, we conduct following ablation study on CIFAR100 with 20% mismatch ratio.

The effect of openness. We study the effect of openness for all compared methods, where the openness denotes how many open-set examples in the unlabeled pool. By increasing the openness from 0% (closed-set) to 90% on CIFAR100, we record the results of final accuracy and recall after 10 queries. The experimental results are plotted in Fig 7. When openness is 0% or very small, LfOSA performs worse than most closed-set methods. With larger openness, the advantage of LfOSA in terms of both accuracy and recall becomes more noticeable.

The effect of batch size. The results on the effect of batch size b is in Fig 8. Our method consistently achieve the best performance when $b = 500, 1000, 1500, 2000, 2500$.

The effect of each component for LfOSA. The experimental results on the effect of each component for LfOSA are demonstrated in Figure 9. Where **w/o temperature** and **high temperature** denote the temperature T is set to 1 and 2 respectively. **w/o Detector** denotes the detector is not used, which means it employs for both detection and classification tasks. Similarly, **w/o Classifier** denote the classifier is not used. **Dichotomies** refers to using a binary classification to replace the detector. **w/o invalid set** denotes the detector
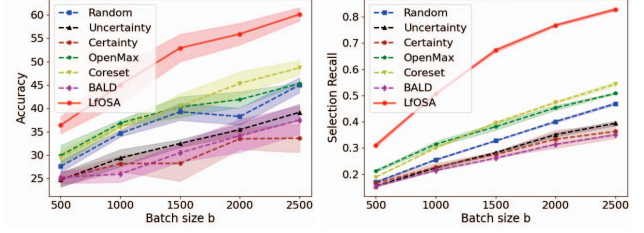
training without using the invalid set. Removing or replacing a component of LfOSA will damage its performance.
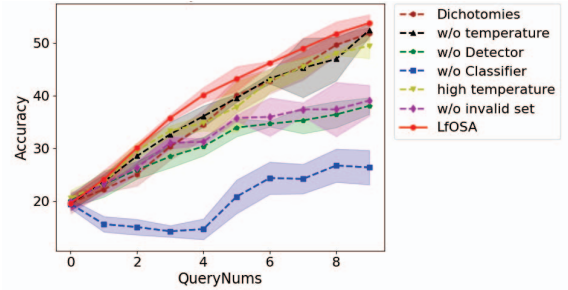

Figure 9. The effect of each component for LfOSA on CIFAR100 with 20% mismatch ratio.

5. Conclusion

In this paper, we formulate a new open-set annotation (OSA) problem for real-world large-scale annotation tasks. It introduces a practical challenge on how to maintain a high recall in identifying the examples of known classes for target model training from a massive unlabeled open-set. To overcome this challenge, we propose an active learning framework called LfOSA to precisely select examples of known classes by decoupling detection and classification. By minimizing low-temperature cross-entropy loss, it exploits both known and unknown supervision to train a detector, whose activation values will be fed into a mixture Gaussian model to estimate the per-example max activation value (MAV) distribution. Based on MAV distribution, we can distinguish examples of known classes against unknown classes in unlabeled data to build a query set for annotation. The classifier is then updated with labeled data. Experimental results on various tasks show the superiority of the LfOSA approach. In the future, we will extend the OSA problem to other computer vision tasks, *e.g.*, object detection.

6. Acknowledgments

References

[1] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1563–1572, 2016. 2, 5

[2] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018. 2

[3] Yifan Fu, Xingquan Zhu, and Bin Li. A survey on instance selection for active learning. *Knowledge and information systems*, 35(2):249–283, 2013. 2

[4] ZongYuan Ge, Sergey Demyanov, Zetao Chen, and Rahil Garnavi. Generative openmax for multi-class open set classification. *arXiv preprint arXiv:1707.07418*, 2017. 2

[5] Stuart Geman, Elie Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural computation*, 4(1):1–58, 1992. 2

[6] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 2

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 2

[8] Xian-Sheng Hua and Guo-Jun Qi. Online multi-label active annotation: towards large-scale content-based video search. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 141–150, 2008. 2

[9] Sheng-Jun Huang, Rong Jin, and Zhi-Hua Zhou. Active learning by querying informative and representative examples. In *Advances in neural information processing systems*, pages 892–900, 2010. 2

[10] Sheng-Jun Huang, Jia-Wei Zhao, and Zhao-Yang Liu. Cost-effective training of deep cnns with active model adaptation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1580–1588, 2018. 1

[11] Sheng-Jun Huang and Zhi-Hua Zhou. Active query driven by uncertainty and diversity for incremental multi-label learning. In *2013 IEEE 13th International Conference on Data Mining*, pages 1079–1084, 2013. 2

[12] Sheng-Jun Huang, Chen-Chen Zong, Kun-Peng Ning, and Hai-Bo Ye. Asynchronous active learning with distributed label querying. 1

[13] Lalit P Jain, Walter J Scheirer, and Terrance E Boult. Multi-class open set recognition using probability of inclusion. In *European Conference on Computer Vision*, pages 393–409, 2014. 2

[14] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 5

[15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 1

[16] David D Lewis and William A Gale. A sequential algorithm for training text classifiers. In *SIGIR'94*, pages 3–12, 1994. 2, 5

[17] Zhao-Yang Liu and Sheng-Jun Huang. Active sampling for open-set classification without initial annotation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4416–4423, 2019. 2

[18] Wenjie Luo, Alex Schwing, and Raquel Urtasun. Latent structured active learning. *Advances in Neural Information Processing Systems*, 26:728–736, 2013. 5

[19] Kun-Peng Ning, Lue Tao, Songcan Chen, and Sheng-Jun Huang. Improving model robustness by adaptively correcting perturbation levels with active queries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9161–9169, 2021. 2

[20] Buyue Qian, Xiang Wang, Jun Wang, Hongfei Li, Nan Cao, Weifeng Zhi, and Ian Davidson. Fast pairwise query selection for large-scale active learning to rank. In *2013 IEEE 13th International Conference on Data Mining*, pages 607–616, 2013. 2

[21] Douglas A Reynolds. Gaussian mixture models. *Encyclopedia of biometrics*, 741:659–663, 2009. 2

[22] N Roy and A McCallum. Toward optimal active learning through sampling estimation of error reduction. int. conf. on machine learning, 2001. 2

[23] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2012. 2

[24] Walter J Scheirer, Lalit P Jain, and Terrance E Boult. Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2317–2324, 2014. 2

[25] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017. 5

[26] Burr Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009. 1, 2

[27] H Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 287–294, 1992. 2

[28] Li-Li Sun and Xi-Zhao Wang. A survey on active learning strategy. In *2010 International Conference on Machine Learning and Cybernetics*, volume 1, pages 161–166, 2010. 1

[29] Ying-Peng Tang, Xiu-Shen Wei, Borui Zhao, and Sheng-Jun Huang. Qbox: Partial transfer learning with active querying for object detection. *IEEE Transactions on Neural Networks and Learning Systems*, 2021. 1

[30] Toan Tran, Thanh-Toan Do, Ian Reid, and Gustavo Carneiro. Bayesian generative active deep learning. In *International Conference on Machine Learning*, pages 6295–6304, 2019. 5

[31] Leon Yao and John Miller. Tiny imagenet classification with convolutional neural networks. *CS 231N*, 2(5):8, 2015. 5

[32] Xinge You, Ruxin Wang, and Dacheng Tao. Diverse expected gradient active learning for relative attributes. *IEEE transactions on image processing*, 23(7):3203–3217, 2014. 2