

JSOSAL:Joint Sampling for Open-Set Active Learning

Yongxiang Zhang

Zhengzhou University, Zhengzhou, China

MOJIEZUOT@163.COM

Bo Zhang

Zhengzhou University, Zhengzhou, China

ZHANGBO2050@ZZU.EDU.CN

Zhiqiang Dai

Zhengzhou University, Zhengzhou, China

158624591@QQ.COM

Yangjie Cao*

Zhengzhou University, Zhengzhou, China

CAOYJ@ZZU.EDU.CN

Zhengzhou University, Zhengzhou, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Traditional active learning methods typically operate under closed-set assumptions, where unlabeled data samples are selected for annotation from a pool consisting exclusively of known classes. However, real-world scenarios predominantly exhibit open-set conditions, characterized by the presence of substantial unknown-class instances within datasets. This fundamental discrepancy renders most conventional active learning approaches ineffective in practical applications. To address the annotation challenge in open-set environments, we propose JSOSAL (Joint Sampling for Open-Set Active Learning), an innovative approach that applies a Bayesian Gaussian Mixture Model (BGMM) to represent the probability distribution of the highest activation values, enabling effective discrimination between known and unknown classes. Our method subsequently selects high-entropy samples from the identified known-class subset for annotation. Rigorous testing on CIFAR-10 and CIFAR-100 shows that JSOSAL achieves superior performance compared to existing leading methods.

Keywords: Active Learning; Open Set; Bayesian Gaussian Mixture Model; Entropy Sampling; Joint Sampling

1. Introduction

In computer vision, the remarkable achievements of deep learning are fundamentally supported by large-scale labeled datasets (Huang et al., 2021; LeCun et al., 2015), but obtaining quality labels is costly and may introduce noise (Settles, 2009; Tang et al., 2021). Active learning addresses this by strategically selecting the most informative samples for labeling. Current approaches fall into two categories: diversity-based methods that select varied samples to improve model coverage, and uncertainty-based methods that prioritize ambiguous samples to refine decision boundaries. However, these traditional methods assume all unlabeled data belong to known classes (closed-set), while real-world scenarios often involve substantial unknown-class samples, causing performance degradation as annotations are wasted on irrelevant examples.

To solve this, we formalize the Open-Set Annotation (OSA) task, where unlabeled data contain K known and U unknown classes ($U > K$). Our approach improves upon LfOSA (Ning et al., 2022) by: (1) using a Bayesian GMM (Rakesh and Jain, 2021) to better distinguish known/unknown samples via maximum activation values, (2) applying entropy sampling (Safaei et al., 2024) to select high-confidence known-class samples for annotation, and (3) utilizing misclassified unknowns as negative examples to enhance detection. Comprehensive testing confirms that our framework outperforms leading existing techniques in the selection of meaningful known-class samples while lowering annotation expenditures.

2. Related Work

The primary objective of Active Learning (AL) is to minimize labeling expenses by strategically identifying the most valuable instances within unlabeled datasets, thereby improving model performance with minimal supervision. Most studies focus on designing sampling strategies—featuring three primary categories: uncertainty-oriented techniques, diversity-centered approaches, and hybrid solutions combining both—that are effective in closed-set settings, where all unlabeled data belong to known categories. However, these methods often perform poorly in open-set scenarios. Open-set recognition has gained increasing attention. For example, OpenMax (Bendale and Boult, 2016) introduces an additional output class and employs extreme value theory to estimate the probability that a sample is unknown. Other approaches use GANs to synthesize open-domain samples for modeling the open space. Despite their success in recognition, these methods fall short in open-set annotation, which differs in key ways: annotation begins with limited labeled data and must actively incorporate newly discovered unknowns, requiring dedicated designs to leverage information collected across AL iterations. These challenges highlight the need for tailored open-world active learning solutions. Joint Sampling has shown promise in domains where sampling and downstream tasks are jointly optimized. For example, JoJoNet (Zhao et al., 2022) improves multi-contrast MRI reconstruction via co-optimized sampling and reconstruction networks. In image compression, joint sampling reduces gradient variance, accelerating convergence (Balint et al., 2023). In video perception, JsrNet (Chen et al., 2019) enhances reconstruction by aligning sampling and decoding in distributed settings.

3. Methodology

3.1. Model framework

The framework of JSOSAL is illustrated in Figure 1, and it primarily consists of three fundamental modules: detection network training, intelligent sample acquisition, and classification model learning. The detailed procedure is as follows: First, A neural network is optimized to identify unknown instances by utilizing supervision from both labeled and unlabeled data, incorporating temperature scaling to improve class separation. Next, We utilize a Bayesian Gaussian Mixture Model (BGM) to characterize the probability distribution of Maximum Activation Values (MAV) across individual samples. Based on this model, samples with higher entropy values are selected from the known instances for annotation. Finally, The classifier undergoes iterative refinement by incorporating newly annotated samples from established categories. Subsequent sections elaborate on each constituent element of our proposed framework in detail.

The training process incorporates both known class discrimination (K categories) and unknown class identification through a supplementary ($K + 1$)-th output branch. For a given sample x from either the labeled set or the invalid set, we represent its label c in one-hot encoding by setting the c -th element P_C to 1 while assigning 0 to all other elements. The detection network is trained using the subsequent cross-entropy objective:

$$L_D(x, c) = - \sum_{c=1}^{K+1} w_c p_c \log \left(\frac{\exp(b_c/T)}{\sum_j \exp(b_j/T)} \right) \quad (1)$$

$$w_c = \frac{N}{n_c} \quad (2)$$

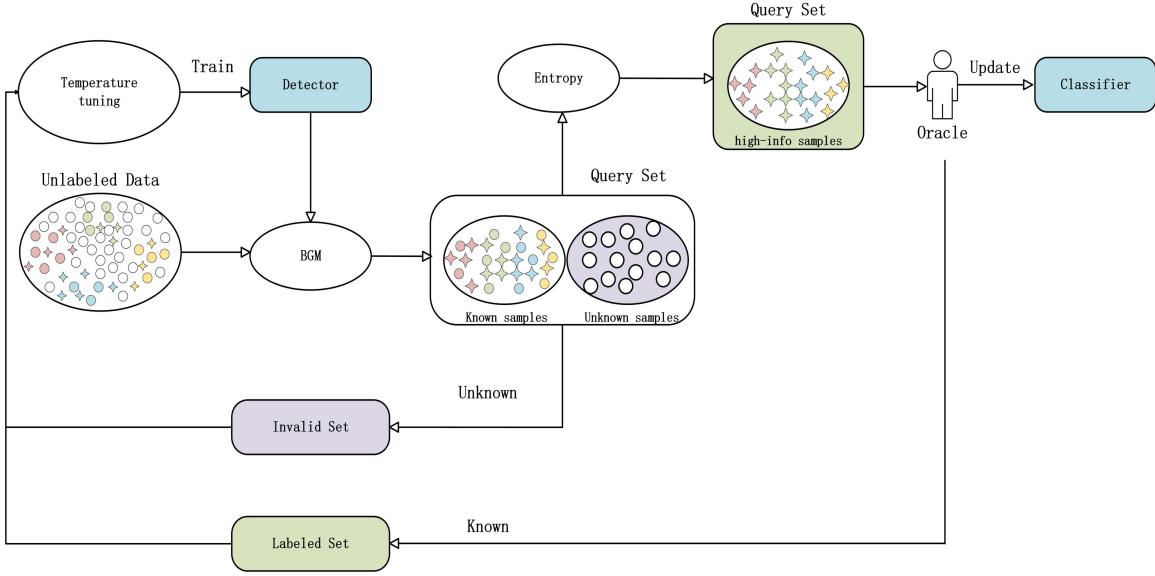


Figure 1: The framework of JSOSAL

Here, b_c signifies the c -th activation value produced by the ultimate fully-connected layer. The temperature parameter T set to a low value ($T = 0.5$) to produce a more distinct categorical probability distribution. W_C represents the weight for class c , N corresponds to the aggregate number of data instances in the collection, and n_c signifies the population size of the c -th category samples. By employing this inverse-proportional weighting, the detector can better learn the features of known classes. Clearly, through minimizing the loss function, samples from known classes will exhibit higher activation values in the first K dimensions and lower activation values in the $(K + 1)$ -th dimension, whereas samples from unknown classes will demonstrate the opposite behavior. Furthermore, reducing the temperature T can further enhance the discriminative capability of the activation layer. Effective open-set annotation (OSA) requires active sampling strategies that maximize acquisition of informative known-category samples from open-world unlabeled collections. Following the detector's training in the preceding stage, Lfosa observes that the penultimate layer of the network effectively distinguishes unknown samples—specifically, Unknown-class samples demonstrate markedly different Maximum Activation Values (MAV) compared to the average MAV of known classes. For any unlabeled instance x_i classified as category c , we define its maximum activation value mav_i^c as follows:

$$mav_i^c = \max_c b_c^i \quad (3)$$

All unlabeled samples are classified into $K + 1$ categories based on the predictions of the current detector. Our selection process retains samples classified among the initial K known categories for subsequent analysis, while excluding those identified as unknown-class instances. Regarding each predefined class c , our approach implements BGM to characterize the density distribution of peak activation values (mav_c) for each category.. The BGM provides better uncertainty estimation and prevents overfitting in cases of insufficient data or class imbalance, thereby improving the model's

robustness and classification performance.

$$W_c = BGM(mav^c, \theta_D) \quad (4)$$

Where W_c represents the probability distribution for class c. For every unlabeled instance x_i belonging to category c, its established probability weight $w_i \in W_c$ corresponds to the posterior probability $p(g|MAV_i)$, where g represents the Gaussian distribution component exhibiting higher mean values (the component associated with higher activation values). Subsequently, We consolidate and rank the probability values spanning all classes.

$$W = \text{sort} \left(\sum_{i=1}^K W_i \right) \quad (5)$$

The top-a highest-probability instances are then selected to comprise a closed-set collection, followed by computing the entropy for each sample.

$$H(w_i) = - \sum_j P_{ij} \log P_{ij} \quad (6)$$

where P_{ij} represents the probability of sample x_i being predicted as class j. We subsequently identify the b most uncertain instances, as measured by entropy, to constitute our annotation query set. Specifically, the query set X_{query} can be obtained by setting a threshold τ equal to the b-th largest entropy value $H(w_i)$.

$$X_{\text{query}} = \{(x_i, w_i) \mid \tau \leq H(w_i), \forall (x_i, w_i) \in (D_U, W)\} \quad (7)$$

We subsequently query their labels and update them into the known-class query set X_{kno}^{query} and unknown-class query set X_{unk}^{query} respectively. The classifier is trained on the currently labeled data D_L by minimizing the standard cross-entropy loss used in LFOSA for the K-class classification task:

$$L_C(x_i, y_i) = - \sum_{i=1}^{n^L} y_i * \log(f(x_i; \theta_C)) \quad (8)$$

where $(x_i, y_i) \in D_L$, and n^L denotes the number of samples in the current D_L .

Figure 1 shows the model workflow, which takes as inputs a small labeled dataset D_L , batch size b, and temperature T. The detector θ_D and classifier θ_C are randomly initialized alongside an empty invalid set D_I . Each iteration involves training the detector via Equation (1) to compute maximum activation values (mav_i^c) for unlabeled samples, modeling these MAV to estimate known-class probabilities, then selecting and annotating the most informative samples (top-a by probability, top-b by entropy) before updating all components for the next cycle.

4. Experiments

We empirically validated our method's performance through extensive testing on the established CIFAR-10 and CIFAR-100 (Krizhevsky et al., 2009) datasets. These datasets contain 10 and 100 classes respectively. To satisfy the open-set setting, we configured the experiments as follows: for CIFAR-10, we designated the first 2, 3, and 4 classes as known categories; for CIFAR-100, we set the first 20, 30, and 40 classes as known categories.

4.1. Comparative Algorithms

Random Sampling: Randomly selects b examples for annotation

Uncertainty Sampling: Identifies the b instances exhibiting maximal predictive uncertainty to include in the annotation pool.

Coreset Selection: Optimizes sample selection by choosing instances that maximize coverage of the feature space representation for annotation.

Entropy Sampling: Prioritizes the top-b most uncertain instances, as measured by predictive entropy, for manual annotation.

LFOSA (Ning et al., 2022) :Trains a dedicated detector network and employs Gaussian Mixture Models to model the distribution of Maximum Activation Values for dynamically selecting the top-b highest-probability examples.

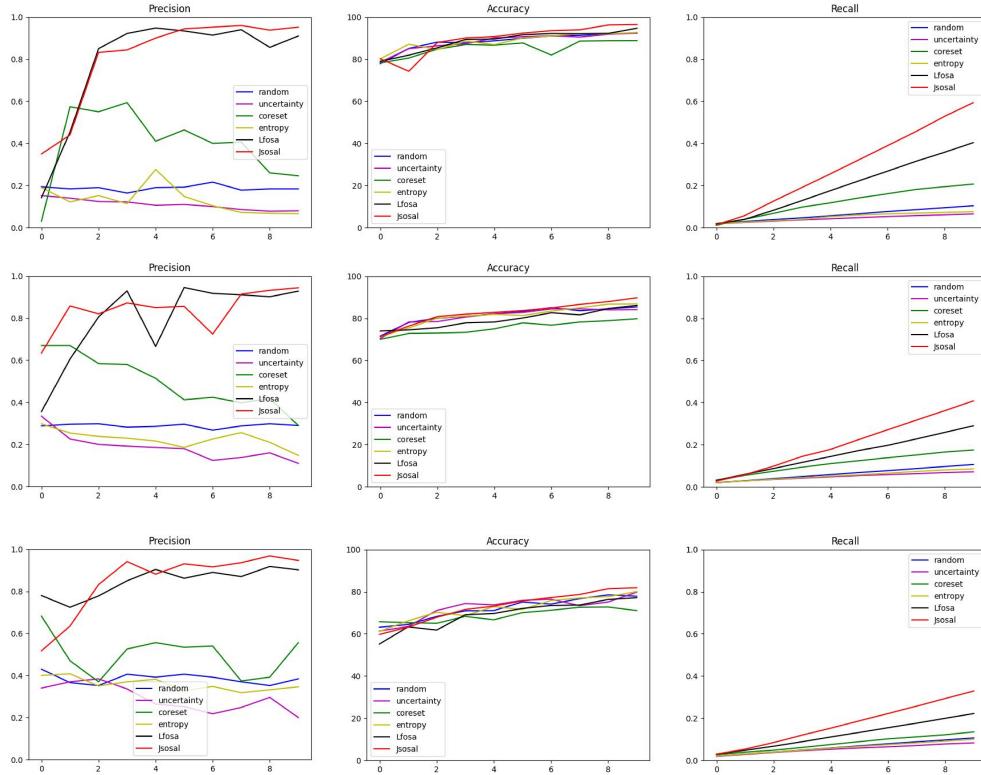


Figure 2: The performance of the model under different known class distributions in the CIFAR-10 dataset, where the first row represents 20%, the second row 30%, and the third row 40%.

4.2. Parameter Settings

Initialization Phase: During the setup phase, we construct our initial labeled dataset by randomly selecting 1% of CIFAR-10 and 8% of CIFAR-100 data, ensuring all labeled instances belong exclusively to known categories. During each active learning cycle, We optimize a ResNet-18 architecture for 100 training cycles per active learning round, employing stochastic gradient descent (SGD) with

0.9 momentum, weight decay 1e-3, and initial learning rate 0.01, with a batch size of 128. Each selection cycle identifies 1000 informative samples for label acquisition, which are subsequently incorporated into the next training iteration.

4.3. Performance Comparison

We assessed JSOSAL’s performance through progressive query-round analyses on CIFAR-10/100 datasets (Figure 2 and Figure 3), testing known-class proportions of 20%/30%/40%. The results show JSOSAL (red/brown curves) consistently surpasses baselines across all configurations, particularly in recall metrics. Our method’s superiority derives from its integrated Bayesian GMM for known-unknown separation and entropy-based sample selection, enabling precise discrimination. Performance improves as unknown-class ratios increase (40%→20% known classes), confirming JSOSAL’s effectiveness for OSA tasks.

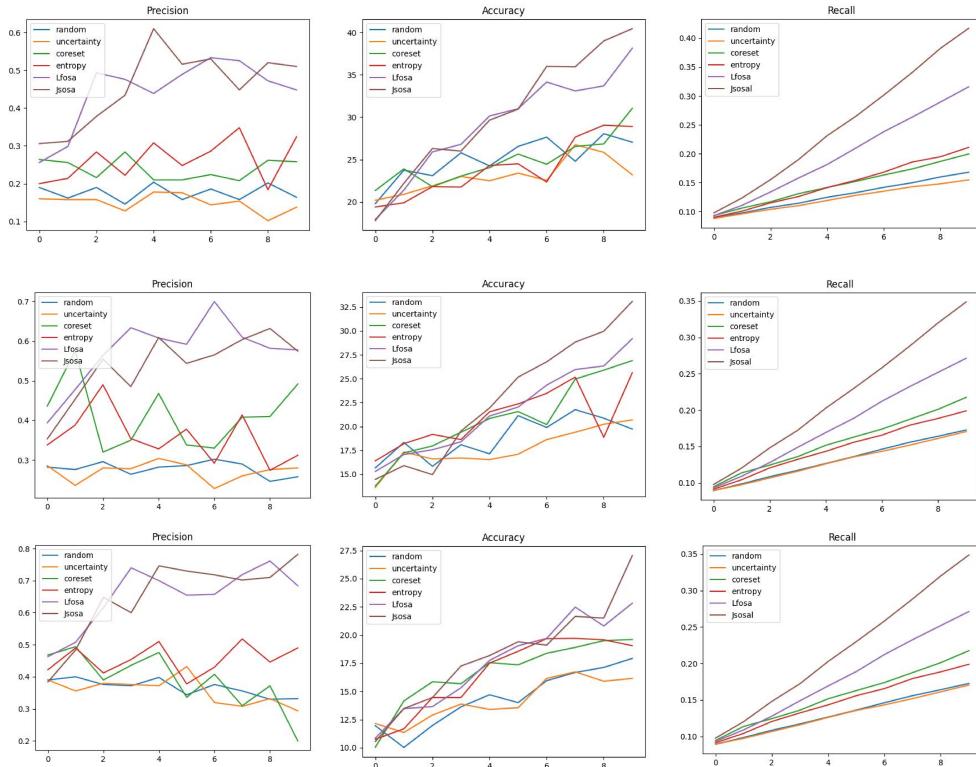


Figure 3: The performance of the model under different known class distributions in the CIFAR-100 dataset, where the first row represents 20%, the second row 30%, and the third row 40%.

4.4. Ablation Study

To comprehensively evaluate JSOSAL’s components, we performed ablation studies on both CIFAR-10 and CIFAR-100 using a 20% known-class proportion, with results visualized in Figure 4.

By comparing the first row of Figure 4 with the first row of Figure 2, and the second row of Figure 4 with the first row of Figure 3, Experimental results in Figures 2 and 3 demonstrate our method’s superior performance over all baseline approaches according to all three evaluation criteria, with particularly more pronounced advantages in recall. However, Figure 4 shows varying degrees of performance degradation across all three metrics, especially a more significant decline in recall. This demonstrates that without employing the high-entropy selection method to choose more informative samples for annotation, the model’s overall performance deteriorates, with this issue being particularly prominent in terms of recall.

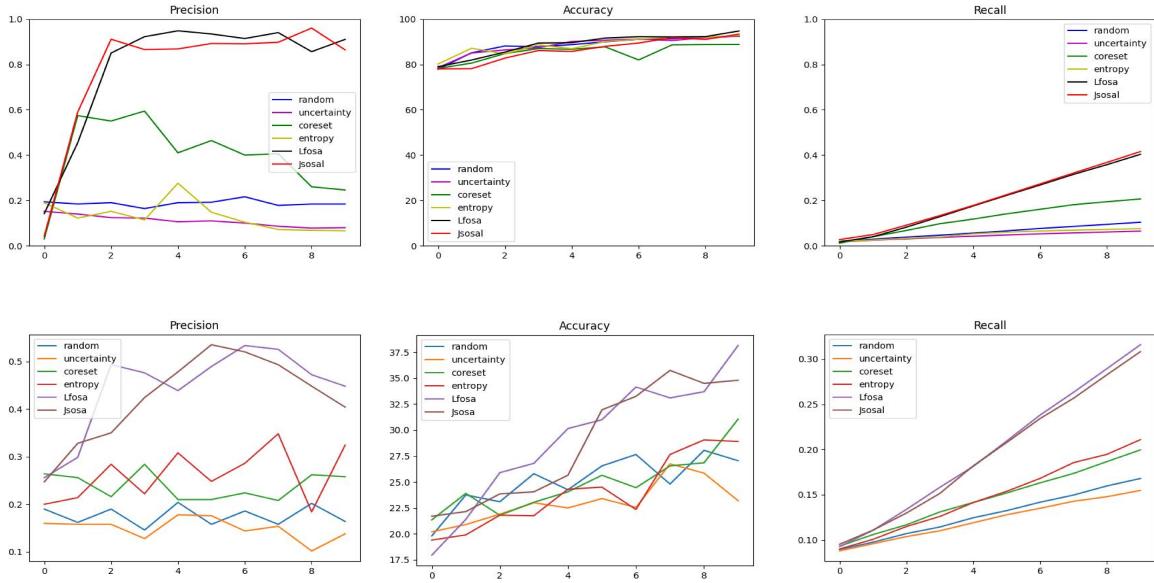


Figure 4: The performance of the model on the CIFAR-10 and CIFAR-100 datasets with 20% known classes, using BGM without entropy calculation.

5. Conclusion

We present JSOSAL, a novel active learning framework that effectively selects the most informative known-class samples from datasets containing substantial unknown-class instances to enhance model performance. JSOSAL utilizes dual detection and classification networks to accurately identify high-value known-class samples. The framework employs a low-temperature cross-entropy loss to train a detector using both known and unknown supervision, with the detector’s activation values processed by a Bayesian Gaussian Mixture Model to estimate each sample’s maximum activation value distribution. This distribution enables partitioning of unlabeled data into known and unknown categories. For the known-class samples, We compute entropy measures to identify maximally informative instances for manual labeling prior to classifier refinement. Empirical evaluations confirm JSOSAL’s effectiveness in open-set active learning, with clear potential for generalization to other vision tasks such as object detection.

References

- Martin Balint, Karol Myszkowski, Hans-Peter Seidel, and Gurprit Singh. Joint sampling and optimisation for inverse rendering. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–10, 2023.
- 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.
- Can Chen, Yutong Wu, Chao Zhou, and Dengyin Zhang. Jsrnet: A joint sampling–reconstruction framework for distributed compressive video sensing. *Sensors*, 20(1):206, 2019.
- Sheng-Jun Huang, Chen-Chen Zong, Kun-Peng Ning, and Haibo Ye. Asynchronous active learning with distributed label querying. In *IJCAI*, pages 2570–2576, 2021.
- Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- Kun-Peng Ning, Xun Zhao, Yu Li, and Sheng-Jun Huang. Active learning for open-set annotation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 41–49, 2022.
- Vineeth Rakesh and Swayambhoo Jain. Efficacy of bayesian neural networks in active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2601–2609, 2021.
- Bardia Safaei, VS Vibashan, Celso M de Melo, and Vishal M Patel. Entropic open-set active learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 4686–4694, 2024.
- Burr Settles. Active learning literature survey. 2009.
- 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*, 34(6):3058–3070, 2021.
- Lin Zhao, Xiao Chen, Eric Z Chen, Yikang Liu, Dinggang Shen, Terrence Chen, and Shanhui Sun. Jojonet: Joint-contrast and joint-sampling-and-reconstruction network for multi-contrast mri. *arXiv preprint arXiv:2210.12548*, 2022.