

An Active Learning Method based on the Label Diversity of Nearest Neighbors

1st Guoqiang Liang*

Shenzhen Research Institute of Northwestern Polytechnical University
Shenzhen, China
gqliang@nwpu.edu.cn

2nd Guanghui Han

Northwestern Polytechnical University
Xi'an, China
hangh@mail.nwpu.edu.cn

3rd Shuailong Zhang

Northwestern Polytechnical University
Xi'an, China
18530793257@163.com

4th Lingyan Ran

Northwestern Polytechnical University
Xi'an, China
Iran@nwpu.edu.cn

Abstract—Deep learning models rely significantly on large volumes of labeled data, the gathering of which is both costly and time-intensive. To tackle this challenge, active learning, which seeks to attain comparable results by selecting and labeling a small portion of data, has attracted more interests. Traditionally, most studies assessing sample importance have focused solely on its individual features. Nonetheless, the relationship between samples is also crucial. For instance, a sample is more likely located around the decision boundary if its several closest neighbors are from different classes. To incorporate this aspect, we propose to explore the diversity of the predicted labels of nearby samples. Specifically, we calculate the variance of the class distribution among the K nearest neighbors, treating it as an indicator of a sample's significance. Those samples with higher scores are selected for querying the oracle to obtain annotations. Moreover, we employ a nearest neighbor classifier, eventually enhancing generalization performance on unlabeled samples. To evaluate this approach, we conduct experiments using the CIFAR-10 and CIFAR-100 datasets, well-known for image classification based on active learning. The results demonstrate the effectiveness of the proposed approach.

Index Terms—Active learning, Sample importance, Decision boundary, Nearest neighbor classifier

I. INTRODUCTION

Supervised deep learning has attained notable success across almost all computer vision tasks. Despite its success, training a powerful model requires a large amount of labeled data. Even with well-known crowd-sourcing platforms, like the Amazon Mechanical Turk, it is costly and time-consuming to construct a large-scale dataset with accurate annotation. To address this problem, active learning, which aims to identify a smaller set of critical samples for annotation, has garnered great attention recently [1], [2].

Based on the sample selection strategy, current active learning works consist of two primary streams: diversity-based approaches and uncertainty-based approaches [2]. Diversity-based methods devise strategies to select a subset of samples

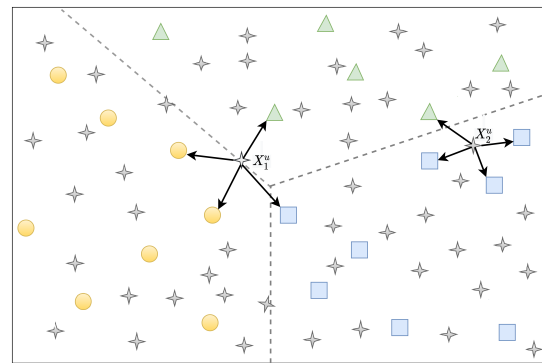


Fig. 1. An illustration of our motivation. Colored squares, triangles, and circles represent the labeled samples, whereas gray stars denote the unlabeled data. Compared to sample X_2^u , the diversity of labels in the samples surrounding X_1^u is higher, indicating that it is closer to the decision boundary.

that best represents the true distribution of datasets [3]–[5]. Nonetheless, they might choose samples that are not particularly beneficial for model training. On the other hand, uncertainty-based approaches target identifying some most ambiguous samples, which provide more useful information for future training cycles [6], [7]. They have been thoroughly examined and demonstrated outstanding results. However, few studies explored the correlation between relative positional relationship of samples in feature space and predicted category probability distribution.

This study seeks to identify samples that are proximate to the classification boundary, as they offer more insights into the boundary compared to other samples [8]. Although the boundary is generally unknown, a sample is more likely located near the boundary if its neighboring samples belong to different categories. Moreover, the greater the difference in categories of its neighbors, the closer the sample is to the boundary. An example for classification is depicted in Fig. 1, where the dotted lines represent the classification boundary. The colored circulars, triangles, squares denote

This work was supported in part by Guangdong Basic and Applied Basic Research Foundation (No. 2023A1515011298); the Natural Science Basic Research Program of Shaanxi Province (No. 2024JC-DXWT-07).

labeled samples from different categories respectively. The gray star indicate unlabeled samples. As shown in this figure, the variety of labels in the samples surrounding X_1^u is greater than that of X_2^u and it is situated nearer to the boundary.

Inspired by the above observation, we propose to explore the label diversity of a sample's neighbors for selecting important samples. Specifically, we first search K nearest neighbors for each unlabeled sample by assessing the distance of samples. Then, we calculate the variance of predicted labels of these K nearest samples, whose results are regarded as their importance score. A higher score suggests an increased probability that this sample is nearer the classification boundary. After that, we select some samples with higher scores, which are passed to the oracle for annotating. To address the distribution discrepancy between the unlabeled and labeled sample sets, we replace the softmax classifier with the nearest neighbor classifier inspired by [9]. To assess the efficacy of our approach, we performed extensive evaluations using the CIFAR-10 and CIFAR-100 benchmarks [10], both of which are popular for active learning-based classification. And the results show the superiority of this method.

II. RELATED WORKS

This part offers a brief summary of active learning. For a comprehensive review, refer to [1], [2].

A. Diversity-based Active Learning

This direction seeks to pick samples which are maximally different from the existing training data and the samples that have already been selected. In the early stage, clustering techniques like K-means and K-medoids were employed to identify diverse and representative samples [11], [12]. For example, Nguyen and Smeulders first [11] incorporated the the K -medoid clustering technique into active learning. Then, more effect criteria were proposed to enhance the diversity of selected samples. Based on a general maximum mutual information principle, [13] selected a set of samples that best represents the dataset distribution. Sener and Savarese [3] utilized the geometric attributes of data points within the feature space to build a core set. To alleviate sampling redundancy, Zhang et al. [14] introduced a heuristic clique rescaling technique to identify the most valuable instances. Lin et al. [15] emphasized the spatial and temporal diversity.

Although these methods tried to select representative and diverse data, some selected samples may not provide much information for model training.

B. Uncertainty-based Active Learning

Another direction is uncertainty-based methods, focusing on choosing samples with higher uncertainty by current model. Thus, how to measure the uncertainty is the core problem. The confidence of a model's prediction can serve as a reliable metric. For instance, Xie et al. [6] proposed an energy-based sampling strategy. Elenter et al. [16] adopted a Lagrangian duality approach to selects samples with the highest predicted dual variable. [7] leveraged the noise stability to estimate

data uncertainty. Gul et al. [17] proposed to label samples with highest gradient norm of the entropy values. Moreover, uncertainty can also be estimated through automated feature learning models, referred to as data-driven uncertainty learning. According to [2], there exist three types, i.e. loss learning [18], discriminate learning [19] and adversarial learning [20]. For instance, a module for predicting the loss of unlabeled data was employed in [18]. Zhang et al. [21] identified the most valuable instances by utilizing the state predicted from a discriminator.

C. Hybrid Methods

To integrate the strengths of uncertainty and diversity criteria, many methods have attempted to combine them, yielding promising outcomes. Li and Guo [22] developed an adaptive way to combine the uncertainty and information density measure. [23] considered both uncertainty and diversity through evaluating the inconsistency of predictions over a random set of data augmentations. Yan et al. [24] introduced a clustering-based sampling technique that concurrently evaluate the representativeness and informativeness of samples. Informativeness, diversity, and information density are fused in [25]. By building a sequential graph convolution network, Caramulau et al. [26] applied the existing sampling methods on the graph embedding along with their confidence score.

Our model belongs to the uncertainty-based method. Different from the above works, we measure the sample's uncertainty by calculating its distance to the decision boundary.

III. THE PROPOSED METHOD

A. Formulation of Pool-based Active Learning

In the pool-based active learning paradigm, there exists a large pool of unlabeled data $U = \{X_i^u\}_{i=1}^{N_u}$ and an initial labeled dataset $L_0 = \{(X_j^l, Y_j)\}_{j=1}^{N_l}$. Generally, N_l is much smaller than N_u , meaning that the labeled data is insufficient for developing an effective model. In the t -th training cycle, the old model $\mathcal{M}$ is updated using the labeled set L_{t-1} , which is obtained after previous cycle. Then, the new model is used to select b samples from U , where b is the budget size for each cycle. Next, these selected samples are annotated by oracles and merged with L_{t-1} , leading to a new labeled set L_t for next cycle. The above cycles will continue until the total labeled budget B is exhausted or a satisfactory model is obtained. In the end, the final model $\mathcal{M}$ is trained on all labeled samples from initialization.

B. Sample Selection based on Label Diversity

As illustrated in introduction, we want to choose samples located around the decision boundary. To accomplish this objective, we propose to choose samples where the predicted labels of their K nearest neighbors exhibit notable variations. In the following, we will detail how to find the nearest neighbors and measure the divergence of their predictions.

1) *Selection of the Top K Nearest Neighbors* : To accurately measure the distance of samples, we use a conventional neural network to extract powerful feature. Specifically, the output of its penultimate layer are used as the representation of samples. Then, for the j -th unlabeled sample X_j^u , we search for its K nearest labeled neighbors $\{X_1^l, \dots, X_K^l\}$ in the feature space. Specifically, we compute the distance between an unlabeled sample and each labeled sample using a metric function $d(f(X_j^u), f(X_i^l))$, where $f(X_j^u)$, $f(X_i^l)$ denote the feature vector of unlabeled sample X_j^u and i -th labeled sample X_i^l , respectively. The distance metric function $d()$ can be the Euclidean distance or cosine distance function. After that, we rank the distances of all labeled samples and select K labeled samples with the smallest distance. Here, K is a hyper-parameter, whose influence will be shown in the experimental section.

2) *Significant Score of Unlabeled Samples*: After the above process, we obtained a set of labeled samples for each unlabeled sample. Intuitively, if these samples belong to different categories, the unlabeled sample is likely located around the classification boundary. Moreover, the greater the diversity in the neighbors' categories, the closer it is to the boundary. Therefore, we regard the label variation of the K nearest neighbors as the significant score for an unlabeled sample.

Specifically, we calculate the variation of predicted labels by current model $\mathcal{M}$ between each unlabeled sample and its selected K nearest neighbors $\{X_1^l, \dots, X_K^l\}$. Instead of the Kullback–Leibler divergence, we employ the variance function, which is more efficient for measuring prediction inconsistency across multiple categories. In detail, the calculation of the significant score can be represented as:

$$S(X_j^u) = \sum_{c=1}^C \text{Var}[P(\hat{Y} = c | X_j^u, \mathcal{M}), P(\hat{Y} = c | X_1^l, \mathcal{M}), \dots, P(\hat{Y} = c | X_K^l, \mathcal{M})], \quad (1)$$

where $P(\hat{Y} = c | X_1^l)$ represents the predicted likelihood that the sample X_1^l is classified as class c , C represents the total number of categories.

After that, each unlabeled sample is given a significant score. A higher score indicates a notable difference between the predicted labels of an unlabeled sample and those of its nearest neighbors, despite their proximity in the feature space. In other words, this unlabeled sample is more likely located around the decision boundary. Thus, we first rank the scores of all unlabeled samples in descending order. Then, the top b samples are selected and sent to oracles for annotation. After obtaining these labels, they will be merged with old labeled set L_{t-1} , leading to a new labeled set L_t for next cycle.

In theory, choosing such boundary samples is consistent with the core principle of version space reduction. As per our definition, samples with higher label variability are those for which the model experiences the most uncertainty. Labeling these samples provides the greatest information gain, significantly restricting the model's hypothesis space and resulting in the quicker decrease in overall model uncertainty.

Algorithm 1 : An active learning method based on the label diversity of nearest neighbors.

Input: An initial labeled dataset L_0 , unlabeled set U , total budget B , label budget b for one cycle, the number of classes C , the number of nearest neighbors K , random initialized model $\mathcal{M}$.

Output: well-trained classification network: $\mathcal{M}^*$.

```

1: for  $t$  in  $(B / b)$  AL cycles
2:   Update  $\mathcal{M}$  using  $L_{t-1}$ 
3:   Calculate the feature vector  $f(X)$  for  $X \in L_{t-1} \cup U$ ;
4:   for the  $j$ -th sample  $X_j^u \in U$ 
5:     Calculate  $d(f(X_j^u), f(X^l))$  for all  $X^l \in L_{t-1}$ 
6:     Select  $K$  samples  $\{X_i^l\}_{i=1}^K$  based on the distance;
7:     Calculate  $P(\hat{Y} = c | X_i^l, \mathcal{M})$  for  $i = 1, 2, \dots, K$ ;
8:     Calculate  $P(\hat{Y} = c | X_j^u, \mathcal{M})$ ;
9:     Calculate the significant score  $S(X_j^u)$  using Eq. (1);
10:  end
11:  Select  $b$  unlabeled samples  $L_q$  with the maximal scores;
12:  Annotate all samples in  $L_q$  by oracle;
13:   $L_t \leftarrow L_{t-1} \cup L_q$ ;
14:   $U \leftarrow U \setminus L_q$ ;
15: end
16: Train the model  $\mathcal{M}^*$  on  $L_t$  from random initialization;
17: return  $\mathcal{M}^*$ ;

```

Consequently, our criterion can effectively lower the total labeling budget needed to reach a desired level of accuracy, as confirmed by the experiments.

C. Nearest Neighbor Classifier

In active learning, the severe distribution bias between labeled set and unlabeled set has a bad impact on the performance. To solve this problem, Wan et al. [9] employed the nearest neighbor classifier. Inspired by its success, we also replace the traditional softmax with this classifier. Specifically, this classifier first learns n prototype vectors for each class using the set of labeled samples. Then, for an unlabeled sample X_j^u , its likelihood for a class is calculated based on the distance between this sample and its corresponding prototype vectors:

$$P(\hat{Y} = c | X^u) = \max_n \delta(-d(f(X_j^u), m_{c,n})), \quad (2)$$

where $m_{c,n}$ is the n -th prototype vector for a class c . δ is an activate function which maps the distance between the feature and prototype vector to the range of $[0, 1]$. As described in Eq. (2), the smallest distance between a candidate sample and all prototype vectors of a class c is considered as its likelihood for belonging to this class.

To train the parameters, we employ the popular CE loss:

$$\begin{aligned} \mathcal{L} = & - \sum_c Y_c \log P(\hat{Y} = c | X) \\ & + (1 - Y_c) \log (1 - P(\hat{Y} = c | X)), \end{aligned} \quad (3)$$

where $Y_c \in \{0, 1\}$ denotes the c -th component of the true label Y using one-hot encoding. Compared with traditional

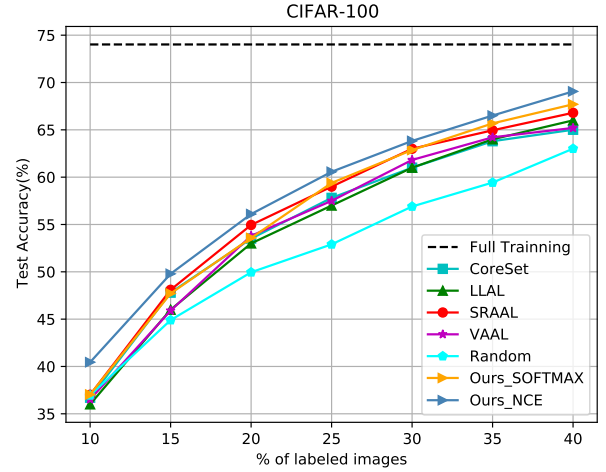
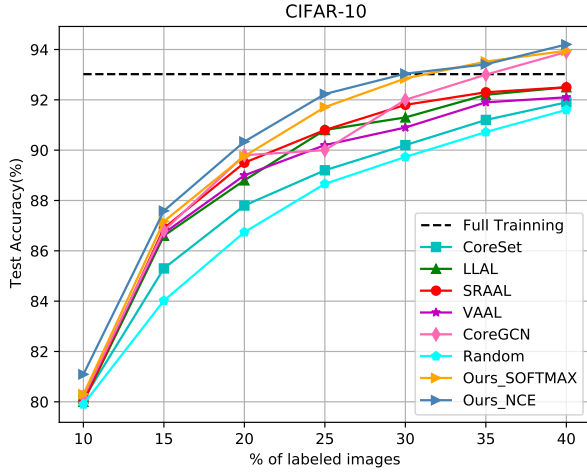


Fig. 2. Comparison of accuracy curve over varying number of labeled images on the CIFAR-10 and CIFAR-100 dataset.

softmax classifier, the nearest neighbor classifier can output more reliable probability prediction values. The reason behind this phenomenon is that the decision boundary obtained by the nearest neighbor classifier is closer to the real classification boundary [9]. Consequently, the instances identified by our method are situated in closer proximity to the true classification boundary, thereby possessing higher informational value. Thus, we use Eq. (2) to calculate the likelihood in Eq. (1). The complete workflow of our proposed method is delineated in Algorithm 1.

IV. EXPERIMENTS

A. Dataset and Implementation Details

Like previous works [20], [26], we employ two commonly used datasets for active learning based image classification: CIFAR-10 and CIFAR-100 [10]. The former comprises 10 distinct categories, each consisting of 6,000 color samples. The latter divides the same images into 100 categories, with each category comprising 600 images. Similar to earlier approaches, 50000 images are regarded as the unlabeled pool U while the other images are used for testing.

Specifically, we first randomly sample 5000 images from U as the labeled dataset L_0 . The total number of cycles is set to 6. In each cycle, the budget size b is 2500. We execute 200 epochs with a batch size of 128. Furthermore, to maintain consistency with prior baselines, we utilize ResNet-18 as the backbone network. Horizontally flipping and random image cropping are used to augment the dataset. Then, these images are normalized using channel mean and standard deviation vectors derived from the training set. The SGD optimizer is utilized to update the network. The learning rate starts at 0.1 and is subsequently decreased to 0.01 following the completion of the 160th epoch. Furthermore, the momentum and weight decay hyper-parameters are fixed at 0.9 and 5×10^{-4} , respectively. To mitigate the impact of random sampling, each experiment is conducted thrice, and we present

the mean accuracy. The experiments are conducted on a server equipped with a single GeForce RTX 2080 Ti GPU.

B. Performance Comparison

We first compare our model with some typical methods, whose results are shown in Fig. 2. In this figure, “Random” means selecting samples randomly. “CoreSet” [3] is one of the best performing geometric techniques. In “LLAL” [18], the samples are chosen according to predicted loss value. “SRAAL” [21] searched for the most valuable samples by utilizing both the annotation data and the labeled/unlabeled status. “VAAL” [20] is a variational adversarial active learning method. “CoreGCN” [26] uses the graph convolutional network. “Full Training” means all training set are used. The “Ours_SOFTMAX” and “Ours_NCE” are two proposed methods with different classifier.

As shown in Fig. 2, our approach surpasses all other methods on the two datasets. In any cycles, we obtain the best performance. This means that our method is more practical since the total annotation budget is various in real application. And the gain is more obvious when the NCE classifier is used. Besides, on the CIFAR-10 dataset, the proposed approach attains 94.12% accuracy using 20000 images (40% of the whole training set), surpassing the accuracy of 93.02% achieved when using the full set of images. We think this is because not all data are useful for model learning and some are even harmful, which is in accord with the observations in [27]. Moreover, compared with other baselines, our method requires fewer samples to achieve the same performance. For instance, “SRAAL” and “LLAL” get 90.8% accuracy with about 12500 annotated samples on CIFAR-10. To achieve the similar accuracy, our method with NCE classifier only requires about 10000 labeled images. All these demonstrate the advantages of our proposed method.

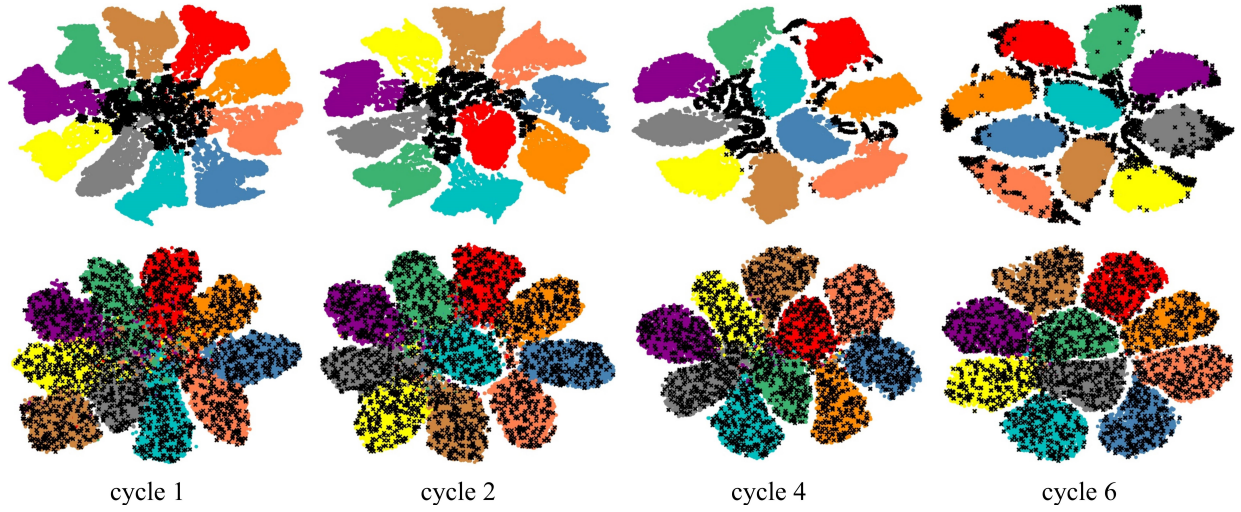


Fig. 3. T-SNE visualization of sample distributions on CIFAR-10. Black cross denotes the selected samples for annotation while other colored points represent unlabeled samples from each category. The top line represents the results of our approach, whereas the bottom line indicates the outcomes of a baseline method employing a random sampling strategy.

TABLE I
INFLUENCE OF THE K VALUE ON CIFAR-10

K	Proportions of Labeled Samples						
	10%	15%	20%	25%	30%	35%	40%
10	81.68	87.21	90.00	92.20	93.22	93.64	93.70
15	81.20	87.67	90.63	92.07	93.23	93.71	93.54
20	81.90	87.35	90.37	92.18	93.19	93.89	93.00
30	81.09	87.59	90.34	92.23	93.04	93.41	94.20
40	81.41	87.88	89.89	92.19	93.29	93.50	94.12

TABLE II
INFLUENCE OF THE K VALUE ON CIFAR-100

K	Proportions of Labeled Samples						
	10%	15%	20%	25%	30%	35%	40%
40	39.75	49.52	55.61	60.77	63.44	66.24	68.5
50	40.45	49.79	56.1	60.57	63.83	66.51	69.06
60	40.53	49.51	55.78	60.25	63.60	66.37	68.61
80	40.84	49.48	55.32	59.52	63.16	65.80	67.99
100	40.53	49.26	54.28	58.48	62.31	65.19	67.64

C. Ablation Studies

First, we show the influence of K , the number of the nearest neighbors introduced in Section III-B1. To select the best K , we conduct several experiments using different K on CIFAR-10 and CIFAR-100, whose results are given in Table I and II. The best value for CIFAR-10 is different for each cycle. However, the discrepancy of accuracy is not obvious. In contrast, the performance on CIFAR-100 achieves the best when K equals to 50. When it continues to increase, the accuracy become worse. If an unlabeled sample just locates at the boundary, a larger K value does not affect its variance. However, for a sample located around boundary, if a larger K is used, more neighbors belonging to other categories may be selected, which will lead to larger variance. In other words, a large K means samples further from the boundary will be

TABLE III
IMPACT OF THE DEFINITIONS FOR NEAREST NEIGHBORS ON CIFAR-100

Neighbors	Proportions of Labeled Samples						
	10%	15%	20%	25%	30%	35%	40%
Baseline	40.45	49.79	56.10	60.57	63.83	66.51	69.06
TrueLabel	40.43	49.98	55.94	60.50	63.71	66.31	68.49
NearestK	40.35	49.83	56.24	60.56	63.84	66.37	68.50

selected. However, they may be not the best for annotation especially when the budget is limited. The varying impact can be ascribed to the different quantity of image categories. Since the CIFAR-10 includes less classes, its variance will not become bigger when K keeps increasing. Therefore, we set K to 30 and 50 for CIFAR-10 and CIFAR-100, respectively.

In Sec. III-B, only the labeled samples are employed to calculate the label diversity. Actually, we have two different choices for nearest neighbors selection: (1) select K nearest labeled samples around a candidate sample; (2) select K nearest samples, which may include labeled and unlabeled samples. Besides, for labeled samples in case (1), we can use the true label or the label predicted by the current model. To demonstrate their impact, we conducted experiments on CIFAR-100, with the outcomes presented in Table III. In this table, the “Baseline” and “TrueLabel” denote the performance using the predicted label and true label respectively. The second case is denoted as the “NearestK”. From this table, we can see that the performance of different choices has little difference in most cases. We use the labeled samples and predicted labels for other experiments since its average score is a little higher.

To further verify the superiority of our sampling strategy, we designed a model variant “Margin+NCE”. Specifically, we just substitute the proposed strategy with a margin-based

TABLE IV
INFLUENCE OF SAMPLING STRATEGY, WHERE C10 AND C100 DENOTES
THE CIFAR-10 AND CIFAR-100 RESPECTIVELY.

Data	Method	Proportions of Labeled Samples						
		10%	15%	20%	25%	30%	35%	40%
C10	Margin+NCE	81.20	87.44	89.79	91.38	92.55	92.87	93.35
	Ours_NCE	81.09	87.59	90.34	92.23	93.04	93.41	94.20
C100	Margin+NCE	37.88	49.10	55.46	60.07	63.13	65.65	67.37
	Ours_NCE	40.45	49.79	56.10	60.57	63.83	66.51	69.06

uncertainty sampling technique while other things are kept the same. Table IV display the experimental results on CIFAR-10 and CIFAR-100. Obviously, Ours_NCE outperforms the Margin+NCE across all labeling budgets, confirming that our variance-based sampling strategy is superior to the classic margin-based uncertainty approach.

D. Qualitative Comparison

Fig. 3 shows the distribution of selected samples via the t-SNE technique [28]. From this figure, we can find the samples selected by random sampling are evenly distributed, with only a minor portion containing important information. In contrast, most of our selected samples are in the boundary. In the initial iteration, most of them locate at the junction of multiple classification boundaries, which can make the model grasp the main differences of samples. In the later stages, our method pays more attention to the edges of different classes. This coarse to fine strategy for sample selection enhances model training.

V. CONCLUSION

We develop an effective sample selecting strategy for active learning by exploring the diversity of the predicted labels of nearby samples. Specifically, we compute the variance of class distribution of nearest neighbors as the significant scores for unlabeled samples. Besides, we replace the traditional softmax classifier with the nearest neighbor classifier. To evaluate this method, we have done extensive experiments on two popular dataset, demonstrating the efficacy of the proposed methodology. In upcoming research, we intend to integrate our uncertainty strategy with the diversity of selected samples to further enhance the performance.

REFERENCES

- [1] D. Li, Z. Wang, Y. Chen, R. Jiang, W. Ding, and M. Okumura, "A survey on deep active learning: Recent advances and new frontiers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 4, pp. 5879–5899, 2024.
- [2] P. Liu, L. Wang, R. Ranjan, G. He, and L. Zhao, "A survey on active deep learning: from model-driven to data-driven," *ACM Computing Surveys (CSUR)*, 2021.
- [3] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," in *Proc. of Int. Conf. Learn. Represent.*, 2018.
- [4] W. Tan, L. Du, and W. Buntine, "Bayesian estimate of mean proper scores for diversity-enhanced active learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 5, pp. 3463–3479, 2024.
- [5] M. Thoma, T. Preintner, E. Aghajanzadeh, S. B. Sampath, P. Mori, N. Fafous, M.-R. Vempala, A. Frickenstein, D. Mueller-Gritschneider, and U. Schlichtmann, "Uncertainty aware training to improve uncertainty active learning for semantic segmentation," in *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2025, pp. 4401–4411.

- [6] B. Xie, L. Yuan, S. Li, C. H. Liu, X. Cheng, and G. Wang, "Active learning for domain adaptation: An energy-based approach," in *Proc. of AAAI Conf. Artif. Intell.*, vol. 36, no. 8, 2022, pp. 8708–8716.
- [7] X. Li, P. Yang, Y. Gu, X. Zhan, T. Wang, M. Xu, and C. Xu, "Deep active learning with noise stability," in *Proc. of AAAI Conf. Artif. Intell.*, vol. 38, no. 12, 2024, pp. 13 655–13 663.
- [8] M. Ducoffe and F. Precioso, "Adversarial active learning for deep networks: a margin based approach," *arXiv preprint arXiv:1802.09841*, 2018.
- [9] F. Wan, T. Yuan, M. Fu, X. Jib, Q. Huang, and Q. Ye, "Nearest neighbor classifier embedded network for active learning," in *Proc. of AAAI Conf. Artif. Intell.*, vol. 35, no. 11, 2021, pp. 10041–10048.
- [10] A. Krizhevsky, V. Nair, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," University of Toronto, Tech. Rep., Apr. 2009.
- [11] H. T. Nguyen and A. Smeulders, "Active learning using pre-clustering," in *Proc. of Int. Conf. Mach. Learn.*, 2004, p. 79.
- [12] P. Zhou, T. Zhang, L. Zhao, Y. Qi, Y. Chang, and L. Bai, "Pre-clustering active learning method for automatic classification of building structures in urban areas," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106382, 2023.
- [13] Y. Guo, "Active instance sampling via matrix partition," in *Proc. of Adv. Neural Inf. Process. Syst.*, vol. 23, 2010.
- [14] B. Zhang, L. Li, S. Wang, S. Cai, Z.-J. Zha, Q. Tian, and Q. Huang, "Inductive state-relabeling adversarial active learning with heuristic clique rescaling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 9780–9796, 2024.
- [15] J. Lin, Z. Liang, S. Deng, L. Cai, T. Jiang, T. Li, K. Jia, and X. Xu, "Exploring diversity-based active learning for 3d object detection in autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 11, pp. 15 454–15 466, 2024.
- [16] J. Elenter, N. NaderiAlizadeh, and A. Ribeiro, "A lagrangian duality approach to active learning," in *Proc. of Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 37 575–37 589.
- [17] S. Gul, M. Elmahallawy, S. Madria, and A. Tripathy, "Lplgrad: Optimizing active learning through gradient norm sample selection and auxiliary model training," in *IEEE BigData*, 2024, pp. 900–909.
- [18] D. Yoo and I. S. Kweon, "Learning loss for active learning," in *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 93–102.
- [19] D. Gissin and S. Shalev-Shwartz, "Discriminative active learning," *arXiv preprint arXiv:1907.06347*, 2019.
- [20] S. Sinha, S. Ebrahimi, and T. Darrell, "Variational adversarial active learning," in *Proc. of IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 5972–5981.
- [21] B. Zhang, L. Li, S. Yang, S. Wang, Z.-J. Zha, and Q. Huang, "State-relabeling adversarial active learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 8756–8765.
- [22] X. Li and Y. Guo, "Adaptive active learning for image classification," in *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2013, pp. 859–866.
- [23] M. Gao, Z. Zhang, G. Yu, S. Ö. Anık, L. S. Davis, and T. Pfister, "Consistency-based semi-supervised active learning: Towards minimizing labeling cost," in *Proc. of Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2020, pp. 510–526.
- [24] X. Yan, S. Nazmi, B. Gebru, M. Anwar, A. Homaifar, M. Sarkar, and K. D. Gupta, "A clustering-based active learning method to query informative and representative samples," *Applied Intelligence*, vol. 52, no. 11, pp. 13 250–13 267, 2022.
- [25] S. Matiz and K. E. Barner, "Inductive conformal predictor for convolutional neural networks: Applications to active learning for image classification," *Pattern Recognition*, vol. 90, pp. 172–182, 2019.
- [26] R. Caramalau, B. Bhattarai, and T.-K. Kim, "Sequential graph convolutional network for active learning," in *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 9583–9592.
- [27] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *Proc. of Int. Conf. Mach. Learn.*. PMLR, 2017, pp. 1885–1894.
- [28] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 11, p. 2579–2605, 2008.