State-Relabeling Adversarial Active Learning

Beichen Zhang[1], Liang Li[2]*, Shijie Yang[1, 2], Shuhui Wang[2], Zheng-Jun Zha[3], Qingming Huang[1, 2, 4]
[1]University of Chinese Academy of Sciences, Beijing, China
[2]Key Lab of Intell. Info. Process., Inst. of Comput. Tech., Chinese Academy of Sciences, China
[3]University of Science and Technology of China, China, [4]Peng Cheng Laboratory, Shenzhen, China,
{beichen.zhang, shijie.yang}@vipl.ict.ac.cn, {liang.li, wangshuhui}@ict.ac.cn,
zhazj@ustc.edu.cn, qmhuang@ucas.ac.cn

Abstract

Active learning is to design label-efficient algorithms by sampling the most representative samples to be labeled by an oracle. In this paper, we propose a state relabeling adversarial active learning model (SRAAL), that leverages both the annotation and the labeled/unlabeled state information for deriving the most informative unlabeled samples. The SRAAL consists of a representation generator and a state discriminator. The generator uses the complementary annotation information with traditional reconstruction information to generate the unified representation of samples, which embeds the semantic into the whole data representation. Then, we design an online uncertainty indicator in the discriminator, which endues unlabeled samples with different importance. As a result, we can select the most informative samples based on the discriminator's predicted state. We also design an algorithm to initialize the labeled pool, which makes subsequent sampling more efficient. The experiments conducted on various datasets show that our model outperforms the previous state-of-art active learning methods and our initially sampling algorithm achieves better performance.

1. Introduction

Although deep neural network models have made great success in many areas, they still heavily rely on large-scale labeled data to train large number of parameters. Unfortunately, it is very difficult, time-consuming, or expensive to obtain labeled samples, which becomes the main bottleneck for deep learning methods [10]. To reduce the demand of labeled data, learning methods like unsupervised learning [6, 35], semi-supervised learning [19, 30], weakly supervised learning [46, 28] and active learning have attracted a lot of attention. Unsupervised and semi-supervised methods aim to fully utilize the unlabeled samples while active learning is to select as few samples to be labeled as possible for efficient training. This paper focuses on active learning, which is widely used in computer vision tasks such as classification [37, 2] and segmentation [42, 17].

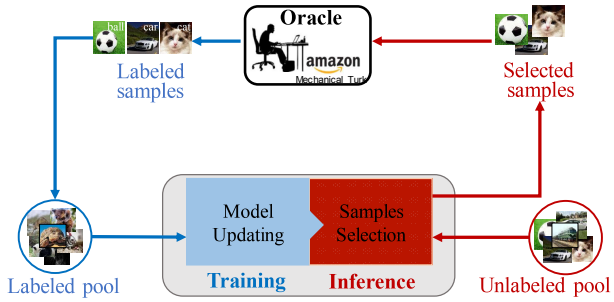

Figure 1. A traditional pool-based active learning cycle. At each iteration, the sampling model is trained with labeled data. After training, a subset of unlabeled samples is selected based on the model inference and then labeled by an oracle. The active learning system will repeat this iteration until the model performance meets user's requirements or the label budget runs out.

In active learning, this means the selected samples should be the most informative ones. As shown in Fig. 1, active learning algorithm is typically an iterative process in which a set of samples is selected to be labeled from an unlabeled pool at each iteration. These selected unlabeled subsets are labeled by an oracle, integrated into the labeled data pool. How to select the most informative samples from the unlabeled pool is the key problem in active learning.

To solve this problem, previous works have made full use of the annotation information of labeled data from the oracle. Many methods [2, 37, 14, 42] built deep learning models, trained them under the supervision of the labeled

*Corresponding author

2575-7075/20/$31.00 ©2020 IEEE
DOI 10.1109/CVPR42600.2020.00878

samples, and then implemented the inference for sampling. A recent work [7] (LL4AL) designed a deep network with an auxiliary loss prediction for labeled data, then selected samples based on the predicted loss. As the supervised information for network training, both the size and quality of labeled instances determines the performance of models. Further, the above approaches usually require a certain amount of labeled samples to achieve a high accuracy. However, in the early iterations of sampling, the labeled pool is usually small, so that it restricts the ability to choose samples with high quality.

Beside the above annotation information, some recent works focused on utilizing the state information of samples which indicate a sample is labeled (0) or unlabeled (1). As the key of active learning is to select unlabeled samples and label them, this state information can be directly used as the supervised information in this process. Some recent works [8, 39] regarded the state information as a kind of adversarial label. They built a discriminator to map the data point to a binary label which is 1 if the sample is unlabeled and is 0, otherwise. The above works consider all the unlabeled samples of the same quality. In fact, different samples in unlabeled pool have different importance for target task, and an unlabeled sample has lower priority to be labeled if it is more similar to samples in labeled pool. Thus, this state information should be deeply explored to leverage the sample selection.

In this paper, we propose a state relabeling adversarial active learning model (SRAAL) that considers both the annotation and the state information for deriving most informative unlabeled samples. Our model consists of a unified representation generator and a labeled/unlabeled state discriminator. For the generator, we first build an unsupervised image reconstructor based on VAE architecture to learn the rich representation. Secondly, we design a supervised target learner to predict annotations for labeled samples, where the annotation information is embedded into the representation. Then we concatenate the above representations. For the state discriminator, both labeled state and relabeled unlabeled state are used to optimize the discriminator. We propose an online uncertainty indicator to do the state relabeling for unlabeled data, which endues the unlabeled samples with different importance. Specifically, the indicator calculates the uncertainty score for each unlabeled sample as its new state label. State relabeling helps the discriminator to select more instructive samples.

It is notable that most previous works mainly concentrated on the selection strategy in later iterations while the initialization is usually random. However, the initialization of labeled pool has a large influence for subsequent sample selection and performance of active learning. Here we introduce the k-center [15] approach to initialize the labeled pool, where selected samples is a diverse cover for the dataset under the minimax distance.

Experiments on four datasets at image classification and segmentation tasks show that our method outperforms previous state-of-the-art methods. Further, we implement the ablation study to evaluate the contribution of online uncertainty indicator, supervised target learner and our initial sampling algorithm in SRAAL.

The main contributions of this paper are summarized as:

(i) This paper proposes a state relabeling adversarial active learning model to select most informative unlabeled samples. An online uncertainty indicator is designed to relabel the state of unlabeled data with different importance.

(ii) We build an unsupervised image recontructor and a supervised target learner to generate a unified representation of image, where the annotation information is embedded iteratively.

(iii) We propose the initially sampling algorithm based on the k-center approach, which makes subsequent sampling more efficient.

2. Related work

Active learning has been widely studied for decades and most of the classical methods can be grouped into three scenarios: membership query synthesis [1], stream-based selective sampling [5, 23] and pool-based sampling. As the acquirement of abundant unlabeled samples becomes easy, most of recent works [14, 37, 2, 7, 39] focus on the last scenarios. Current active learning methods can be divided into two categories: pool-based approaches and synthesizing approaches.

Instead of querying most informative instances from an unlabeled pool, the synthesizing approaches [31, 32, 47] use generative models to produce new synthetic samples that are informative for the current model. These methods typically introduce various GAN models [16, 33] or VAE models [22, 40] into their algorithm to generate informative data with high quality. However, the synthesizing approaches still has some disadvantages to overcome, such as high computational complexity and instability of performance [47]. For this reason, this paper mainly focuses on research of the pool-based approaches.

The pool-based approaches can be categorized as distribution-based and uncertainty-based methods. The distribution approach chooses data points that will increase the diversity of labeled pool. To do so, the model should extract the representation of the data and calculate the distribution based on it. Previous works [26, 41, 43] have provided various method to learning the representation. Some active learning models [11, 44] optimize the data selection in a discrete space, and [34] clusters the informative data

points to be selected. Some works [3, 18, 29] focus on how to map the distance of distributions to the informativeness of a data point. Besides, some works estimate the distribution diversity by observing gradient [38], future errors [36] or output changes of trained model [13, 20]. Sener et al. [37] introduce core-set technique into active learning. This method calculates the core-set distance by intermediate features rather than the task-specific outputs, which makes the method applicable to any task and network. The core-set technique has a good performance on datasets with small number of classes. However, the core-set method performs ineffective when the number of classes is big or the data points are in high-dimensions [39].

Uncertainty-based approaches do selection by estimating the uncertainties of samples and sampling top-K data points at each iteration. For Bayesian frameworks, [21, 36] estimate uncertainty by Gaussian processes and [9] adopts Bayesian neural networks. [45] propose a novel active learning approach based on the optimum experimental design criteria in statistics. These traditional methods perform well in some specific tasks but do not scale to deep learning network and large-scale datasets. The ensemble model method was proposed by [2] and applied to some specific tasks [42]. [14] introduces Monte Carlo Dropout to build multiple forward passes, which is a general method for various tasks. However, both the ensemble method and dropout method are computationally inefficient for current deep network and large-scale datasets. Yoo et al. [7] propose a Learning-Loss method and has shown the state-of-the-art performance. Their model consists of a task module and a loss prediction module that predicts the loss of the task module. The two modules learn together and the target loss of task module is regarded as a ground-truth loss for the loss prediction module. This method only utilizes the annotation information in labeled samples and the loss prediction accuracy is affected by the performance of task module. If the task module is inaccurate, the predicted loss cannot reflect how informative the sample is.

Some recent works combine uncertainty and distribution to select data points using a two-step process. Distribution of data points can represent the labeled or unlabeled pool and uncertainty estimation based on the distribution can be more generalized and accurate. A two-step model calculating uncertainty based on information density was proposed in [27]. DFAL [8] and VAAL [39] introduces adversarial learning into their models and build a module to learn the representation of data points. The former method extract representation by learning labeled sample's annotation, while the VAAL method builds a latent space by a VAE that learns together with the discriminator. Both of these models map the representation to labeled/unlabeled in a brute force way and the labeled/unlabeled information are not equivalent to informativeness. For this reason, the results of this method may be unreliable.

3. Method

3.1. Overview

In this section, we formally define the scenario of active learning (AL) and set up the notations for the rest of the paper. In the AL, we have a target task and a target model Θ for the task. At the initial stage, there exists a large unlabeled data pool from which we can randomly select $\mathcal{M}$ samples and obtain annotations of them via an oracle. Let us denote the initial unlabeled pool by D_U and the initial labeled pool by D_L. (x_U) denotes that a data point in unlabeled pool and (x_L, y_L) denotes a data point and its annotation in labeled pool.

The key of the AL algorithm is to select the most informative samples from the unlabeled pool D_U. Once the labeled and unlabeled data pools are initialized, a fixed number of samples will be iteratively selected, labeled and transferred from the unlabeled pool to the labeled pool. Then the unlabeled and labeled pool are updated. As illustrated in Fig. 1, this procedure will repeat until the performance of the target model meets user's requirements, or the budget for annotation runs out.

Fig. 2 shows our state relabeling adversarial active learning model (SRAAL), which uses the annotation and labeled/unlabeled state information for selecting the most informative samples. The SRAAL consists of a unified representation generator (Section 3.2) and a labeled/unlabeled state discriminator (Section 3.3). The former learns the annotation-embedded image feature, and the latter selects more representative samples to be labeled with the help of the online uncertainty indicator. Sampling strategy based on the generator and discriminator is introduced in Section 3.4, and the proposed initially sampling algorithm with k-center is detailed in Section 3.5.

3.2. Unified representation generator

The image representation learning is in the charge of the unified representation generator which consists of the unsupervised image reconstructor (UIR) and the supervised target learner (STL). The image encoder consists of a CNN and two FC modules. The CNN extracts image feature and then FC individually learns the two latent variables for STL and UIR. The UIR module is a variational autoencoder (VAE) in which a low dimensional latent space is learned based on a Gaussian prior. As this process does not require annotations and the reconstruction target is the image itself, samples from both the labeled pool and unlabeled pool contribute to this module. As it's a VAE, the objective function

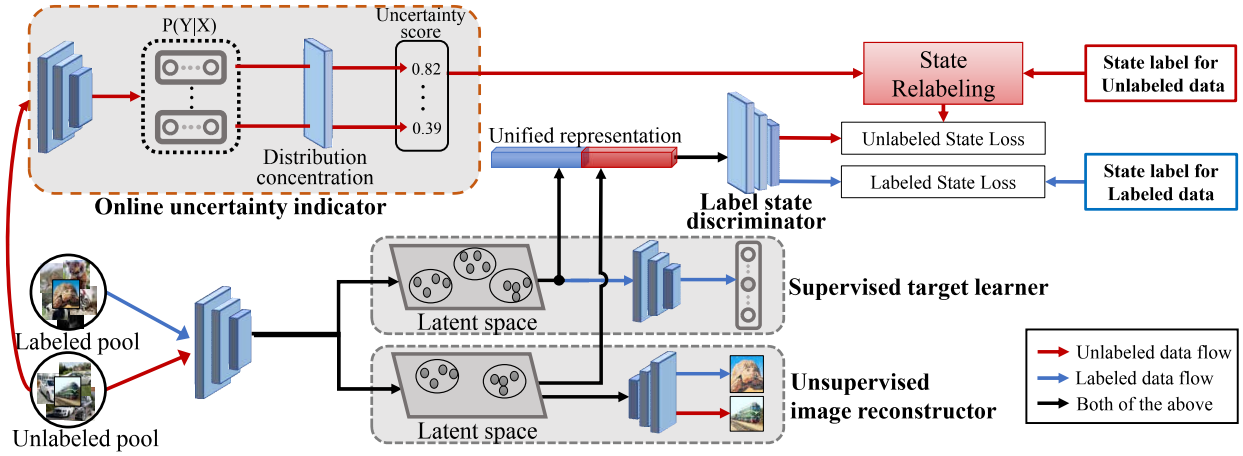

Figure 2. Network architecture of our proposed SRAAL. It consists of a unified representation generator and a labeled/unlabeled state discriminator. The generator embeds the annotation information into the final image features via the supervised target learner and unsupervised image reconstructor. Online uncertainty indicator is introduced to relabel the state of unlabeled samples and endues them with different importance. Finally, the state discriminator is updated through the labeled and unlabeled state losses, and helps select the more informative samples.

of this module can be formulated as,

$$\begin{aligned}\mathcal{L}^{UIR} &= \mathcal{L}_U^{UIR} + \mathcal{L}_L^{UIR} \\ \mathcal{L}_U^{UIR} &= E[log[p_\phi(x_U \mid z_U)] - D_{KL}(q_\theta(z_U \mid x_U) \parallel p(z))] \\ \mathcal{L}_L^{UIR} &= E[log[p_\phi(x_L \mid z_L)] - D_{KL}(q_\theta(z_L \mid x_L) \parallel p(z))]\end{aligned} \tag{1}$$

where $\mathcal{L}_U^{UIR}$ is the objective function for unlabeled data points and L_L^{UIR} is for labeled ones, z is the latent variables, ϕ parametrizes the decoder p_ϕ and θ parametrizes the encoder q_θ. The UIR module finally learns the rich representation by reconstructing both the labeled and unlabeled samples.

To embed the annotation information into the representation, we build a supervised target learner to predict annotations of the samples based on the representation in latent space. The STL is also a VAE network and its decoder does not decode the representation for reconstruction. The decoder of STL varies with different tasks. For example, the decoder is a classifier for image classification, or a segmentation model for semantic segmentation. The STL is similar to VAE but only labeled samples can provide loss for STL's training. We can formulate the objective function for STL as follows,

$$\mathcal{L}_L^{STL} = E[log[p_\phi(y_L \mid z_L)]] - D_{KL}(q_\theta(z_L \mid x_L) \parallel p(z))] \tag{2}$$

where z_L is the latent variables from the latent space for labeled data, ϕ parametrizes the decoder p_ϕ and θ parametrizes the encoder q_θ. The STL module will embed the annotation information into the representation.

The two above representations are concatenated together as the unified image representation.

3.3. State discriminator and state relabeling

To make full use of the state information, we introduce the adversarial learning into SRAAL, where a discriminator is built to model the state of samples. The previous works utilize the binary state information, where the state of unlabeled samples is set to 1 and that of labeled samples is set to 0. In fact, different samples in unlabeled pool have different contribution for target task, and an unlabeled sample has lower priority to be labeled if it is more similar to samples in labeled pool. To better use the state information, we propose the online uncertainty indicator (OUI) to calculate an uncertainty score to relabel the state of unlabeled data. The uncertainty score measures the distribution concentration of the unlabeled data and is bound to [0,1]. After the state relabeling, the state of unlabeled samples changes from the fixed binary label 1 to a new continuous state.

The OUI calculates the uncertainty score based on the prediction vector of the target model(such as image classifier, semantic segmentation model). Before each iteration, the target model is trained with labeled data and then produce a prediction vector for each unlabeled sample. Specifically, for image classification, the prediction is a probability vector for each category. For segmentation, each pixel has a probability vector and the prediction vector is the mean of each probability vector. Assume that the number of classes is C and the samples are labeled with $y_i \in \mathcal{R}^C$. The calcu-

lation of the uncertainty score is formulated as,

$$Indicator(x_U) = 1 - \frac{MINVar(V)}{Var(V)} \times max(V) \quad (3)$$

where x_U is an unlabeled sample, $V = p(x_U|D_L)$ is the probability vector of x_U based on the target model trained with current labeled pool D_L.

The $MINVar(V)$ can be formulated as,

$$\begin{aligned} &MINVar(V) = Var(V') \\ &= \frac{1}{C}((\frac{1}{C} - max(V))^2 + (C-1)(\frac{1}{C} - \frac{1-max(V)}{1-C})^2) \end{aligned} \quad (4)$$

$MINVar(V)$ is the variance of the vector V', whose maximum element is the same with the V's and other elements have the same value $\frac{1-max(V)}{C-1}$. $MINVar(V)$ is the smallest variance among vectors whose maximum are same with V's, so that $\frac{MINVar(V)}{Var(V)}$ measures the concentration of the probabilities distribution. According to Eq. 3, we can prove that the uncertainty score has three properties: (1) it has a boundary of [0,1); (2) it has a negative correlation with the value of maximum probability; (3) it has a positive correlation with the concentration of the probabilities distribution. Due to these properties, the uncertainty score has a good response to the informativeness of samples.

The objective function of the discriminator is defined as follows,

$$\begin{aligned} \mathcal{L}^D = &- E[log(D(q_\theta(z_L \mid x_L)))] \\ &- E[log(Indicator(x_U) - D(q_\theta(z_U \mid x_U)))] \end{aligned} \quad (5)$$

where the indicator relabels the unlabeled data's label.

As adversarial learning, the objective function of the unified representation generator in SRAAL is

$$\begin{aligned} \mathcal{L}^G_{adv} = &- E[log(D(q_\theta(z_L \mid x_L)))] \\ &- E[log(D(q_\theta(z_U \mid x_U)))] \end{aligned} \quad (6)$$

The total objective function combined with Eq. 1, Eq. 2 and Eq. 6 for the latent variable generator is also given as follows,

$$\mathcal{L}^G = \lambda_1 \mathcal{L}^{UIR} + \lambda_2 \mathcal{L}^{STL}_L + \lambda_3 \mathcal{L}^G_{adv} \quad (7)$$

3.4. Sampling strategy in active learning

The algorithm for training the SRAAL at each iteration is shown in Fig. 2. In the sampling step, the generator generates the unified representation for each unlabeled sample. The discriminator predicts its state value, and the top-K samples are selected to be labeled by the oracle.

Algorithm 1 Initialization of labeled pool

Input: labeled data pool D_L, unlabeled pool D_U, the size of initial labeled pool $\mathcal{M}$, latent variables z for all the data points

Hyperparameters: Randomly select I ($I \ll \mathcal{M}$) data points in D_U and move them to D_L

repeat

$u = argmax_{x_U \in D_U}[min_{x_L \in D_L} Distance(x_U, x_L)]$

$D_L = D_L \cup \{u\}$

$D_U = D_U - \{u\}$

until $size(D_L) = \mathcal{M}$

return the initialized labeled pool D_L

3.5. Initially sampling algorithm

It is worth noting that most AL methods mainly study the selection strategy, while the initialization of labeled pool is usually random. However, the initialization of labeled pool can heavily affect subsequent sample selection and performance of active learning. Thus, we propose an initially sampling algorithm in which the problem of initially sampling is defined as a set cover problem. The goal cover problem is to find a subset of data points that the largest distance of any point to the subset is minimum. To measure the distance between samples, first we train the unsupervised image reconstructor so that it learns the latent variables of all the samples, then we apply a greedy k-center algorithm where the distance between two samples is measured by the Euclidean distance between their latent variables. The final output is a subset with $\mathcal{M}$ samples that is labeled by an oracle and sent to the labeled pool. Alg. 1 shows the detail of the algorithm.

4. Experiment

In this section, we evaluate SRAAL against state-of-the-art active learning approaches on image classification and segmentation task.

For both tasks, we initialize the labeled pool D^0_L by randomly sampling M = 10% samples from the entire dataset and the rest 90% samples make up the initial unlabeled pool D^0_U. The unlabeled pool contains the rest of the training set form which samples are selected to be annotated by the oracle. We iteratively train the current model and select K = 5% samples from the unlabeled pool until the portion of labeled samples reaches 40%. For each active learning method, we repeat the experiment 5 times with different initial labeled pool and report the mean performance. When we compare the performance with our methods, they both start with the same initial labeled pool.

To verify the performance of our initial sampling algorithm, we also set an experiment to compare the designed

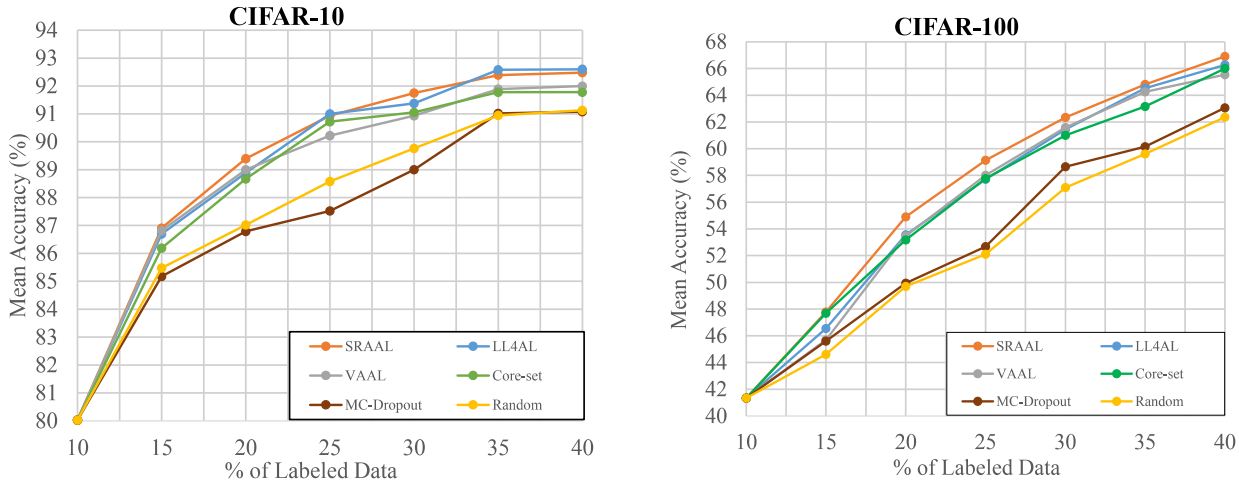

Figure 3. Active learning results of image classification over CIFAR-10 and CIFAR-100.

initialization with the random style. Besides these experiments, we also set an experiment for ablation study to evaluate some main modules in our model.

4.1. Active learning for image classification

Dataset. For the classification task, We choose CIFAR-10, CIFAR-100 [24] and Caltech-101 [12] as they are classical for image recognition and some recent works evaluate on them. Both of CIFAR-10 and CIFAR-100 have 60,000 images of 32×32×3 where 50000 is training images and 10000 is test images. The CIFAR-10 with 10 categories has 6000 images per class, while the CIFAR-100 has 100 classes containing 600 images each. The Caltech-101 consists of a total of 9,146 images, split between 101 different categories, as well as a background category, and each category in Caltech-101 has about 40 to 800 images. These datasets simulate different real-world situations for the number of images per class.

Compared methods. For image classification,we compare the performance of SRAAL with some recent state-of-the-art approaches, including Core-set [37], Monte-Carlo Dropout [14], VAAL [39] and LL4AL. We also introduce the random selection method as a baseline. We reproduce the results of these works with the official released codes and adopt the original hyperparameters. When evaluating, these methods are evaluated by the same target model.

Performance measurement. We evaluate the performances of these methods in image classification by measuring the average accuracy of 5 trials. In each trial, all the methods begin with a same initial labeled pool. The target model used to evaluate the accuracy is a 18-layer residual network (ResNet-18). We utilize a specified Resnet-18 model for CIFAR-10/100 and a classical Resnet-18 for Caltech-101. Besides, images from Caltech-101 are resized to 220×220 for convenience.

4.1.1 Performance on CIFAR-10

The left of Fig. 3 shows the performances on CIFAR-10. We can observe that, first, our method achieves an accuracy over 90% by using 25% of the data and the performance in last iteration reaches 92.48%. The highest accuracy of the Resnet-18 with full dataset reaches 93.5%, which is only 1.02% better than SRAAL with 40% samples. This shows that on CIFAR-10 SRAAL performs closely to the full-data trained model. Second, our method evidently outperforms MC-Dropout, Random sample, core-set and VAAL and is on par with the LL4AL. LL4AL outperforms our SRAAL at 25%, 35% and 40% with very slight lead, but underperforms our method at 15%, 20% and 30%. This also demonstrates that our SRAAL has selected more informative samples, and it benefits from the use of annotation and labeled/unlabeled state information.

4.1.2 Performance on CIFAR-100

CIFAR-10 dataset has 50000 training images categorized into 10 classes, while the CIFAR-100 has 50000 images in 100 classes. Thus, this dataset is much more challenging. The right of Fig. 3 shows the performances, we find that, first, all the AL methods have better results than random selection method. Second, on CIFAR-10, LL4AL performs continuously better than most methods. However, the LL4AL on CIFAR-10 becomes not as competitive as on CIFAR-10. Especially in first iteration, the core-set and LL4AL achieve better performance than LL4AL. The LL4AL trains its model only with the labeled samples. The inadequate labeled samples restrain the accuracy of its main module, which makes the sample selection inefficient. Third, for all iterations, our SRAAL achieves the better performance than the state-of-the-art methods, such as VAAL, LL4AL. Although the VAAL also uses the label state in-

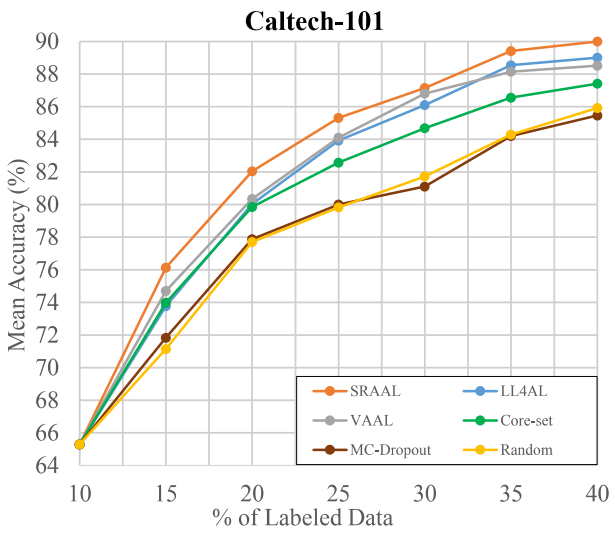

Figure 4. Active learning results of image classification over Caltech-101.

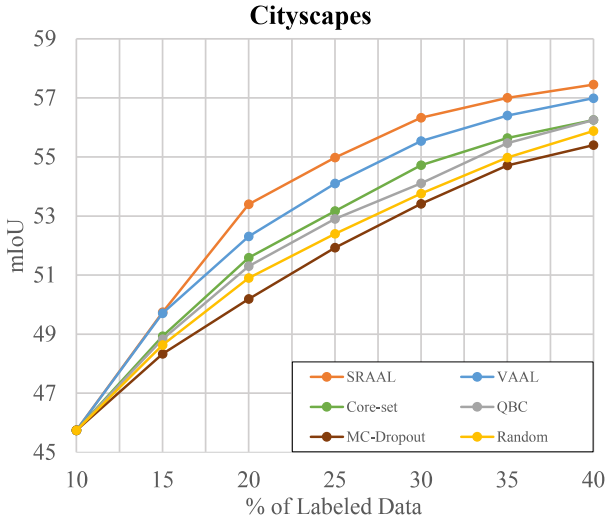

Figure 5. Active learning results of image semantic segmentation over Cityscapes.

formation, the state relabeling in the discriminator from our SRAAL has a better use of it. Besides, the annotation embedded unified representation from the generator provides the richer feature for samples.

4.1.3 Performance on Caltech-101

To further explore the influence of image amount per class to AL model, we conduct the comparison experiment on the Caltech-101 dataset. Fig. 4 shows the performance of these methods. Caltech-101 has less images per class and the image amount for each class is also different. We find that SRAAL outperforms all previous methods from the first iteration to last, and the gap between SRAAL and second-best method becomes larger than that over CIFAR100. This phenomenon provides a further proof that our method can better resist the impact from adequate labeled samples. Besides, the core-set method and LL4AL method perform worse than VAAL because they only utilize annotation information. This verifies that the label state information is useful to help sample the representative data again.

4.2. Active learning for semantic segmentation

Dataset. For semantic segmentation, it is a popular task to evaluate the active learning model. Semantic segmentation is more challenging than image classification, so that this experiment evaluates the performance of AL methods on a difficult task. Here we also conduct the comparison experiment on the Cityscapes dataset [4]. This dataset has 3475 frames with instance segmentation annotations recorded in street scenes. Following [39], we convert this dataset into 19 classes.

Compared methods. We evaluate our SRAAL against some active learning approaches for semantic segmentation. The compared works include Core-set [37], MC-Dropout [14], Query-By-Committee (QBC) [25], suggestive annotation (SA) [42] and VAAL [39].

Performance measurement. For this task, the target model is DRN, and the mean IoU is used to evaluate the performances. All the methods are evaluated with a same initial labeled pool and a same selection budget for each iteration.

Fig. 5 shows our results of semantic segmentation about different methods. We can observe that, first, the VAAL and our SRAAL obtain better performance than other methods, such as SA, QBC, MC-Dropout. This is because both VAAL and SRAAL introduce the label state information to model the sample selection. Second, our SRAAL outperforms the VAAL with a large margin. This benefits from the state relabeling of the proposed online uncertainty indicator. This relabeled state can better guide the discriminator to choose the most informative samples.

4.3. Initialization algorithm comparison

As mentioned in Section 3.5, we introduce the k-center approach to initialize the labeled pool. We evaluate the initialization algorithm on the CIFAR-10 dataset. The AL model is our proposed SRAAL and the target model is the ResNet-18 image classifier, which is the same with that in Section 4.1.

As shown in Fig. 6, we can observe that the mean accuracy of our initialization algorithm is significantly higher than the random initialization. The higher accuracy proves that our initial labeled samples are more informative than random selected ones. Further, the dotted lines show that the standard deviation of our initialization is also less than that of the random initialization. To sum up, our initializa-

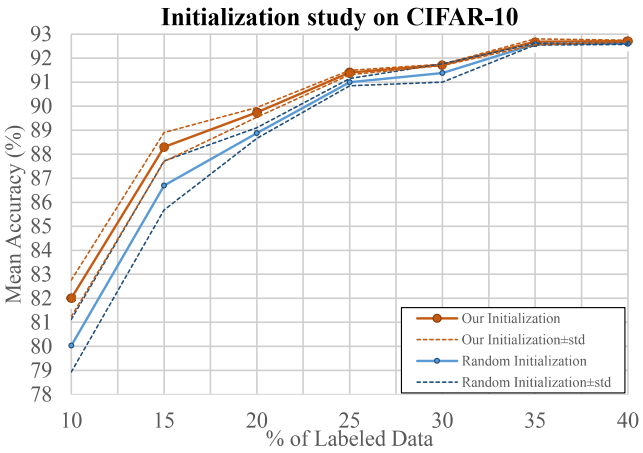

Figure 6. Experiment result for initial sampling algorithm.

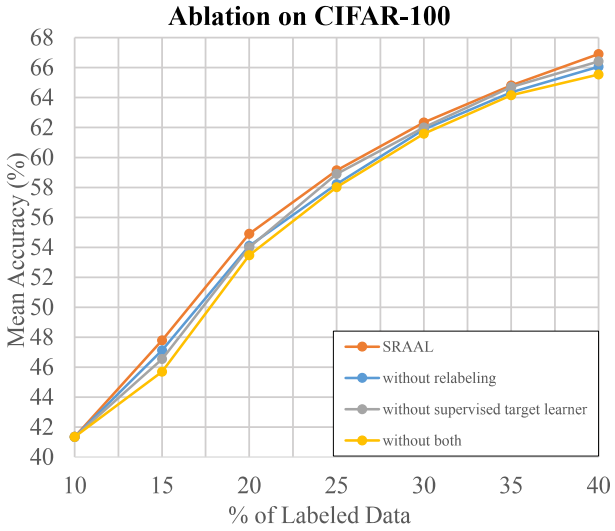

Figure 7. Experiment result for ablation study on CIFAR-100.

tion makes subsequent sampling more efficient.

4.4. Ablation study

To evaluate the contribution of different modules in our model, we conduct this experiment for ablation study on CIFAR-100 dataset. As the state relabeling is the key of our work, we first verify the role of it by eliminating the online uncertainty indicator and using the original binary label state. We also perform an ablation on the supervised target learner to explore the importance of the annotation embedded unified representation. Besides, an ablation to both two modules is performed as the control group.

Fig. 7 shows the results for the ablation study. The complete SRAAL consistently outperforms all the ablations, and the ablation to two modules performs yields lowest accuracy among the ablations. The above phenomenon illustrates that either the relabeling or the annotation-embedded representation helps to improve the AL performance.

4.5. Comparison on different uncertainty estimators

To accurately relabel the state of unlabeled data with different importance, we design an uncertainty score(Eq. 3) in the online uncertainty indicator module. To verify the superiority of our score and prove that our uncertainty score is more suitable for state relabeling, we compare some common uncertainty acquisition functions with ours by replacing our uncertainty score with them. The experiment result in Tab. 1 shows that our indicator outperforms these uncertainty acquisition functions under different sampling ratios, which verifies that our uncertainty score can better reflect the importance of unlabeled data.

(%)	20%	25%	30%	35%	40%
Ours	55.0	59.1	62.3	65.0	66.9
Entropy	54.0	58.2	61.1	64.5	65.7
SD	54.1	57.0	59.4	63.3	64.1

Table 1. Comparison with entropy and standard deviation(SD) under different sampling ratios.

5. Conclusion

In this paper, we study the active learning and propose a state-relabeling adversarial active learning model (SRAAL) that makes full use of both the annotation and the state information for deriving most informative unlabeled samples. The model consists of a unified representation generator that learns the annotation-embedded image feature, and a labeled/unlabeled state discriminator that selects most informative samples with the help of online updated indicator. Further, we introduce the k-center approach to initialize the labeled pool, which makes subsequent sampling more efficient. The experiments on image classification and segmentation demonstrate that our model outperforms previous state-of-the-art methods and the initially sampling algorithm significantly improve the performance of our model.

Acknowledgement. This work was supported in part by the National Key R&D Program of China under Grand:2018AAA0102003, in part by National Natural Science Foundation of China: 61771457, 61732007, 61772497, 61772494, 61931008, 61620106009, U1636214, 61622211, U19B2038, and in part by Key Research Program of Frontier Sciences, CAS: QYZDJ-SSW-SYS013 and the Fundamental Research Funds for the Central Universities under Grant WK2100100030.

References

[1] Dana Angluin. Queries and concept learning. *Machine learning*, 2(4):319–342, 1988.

[2] William H Beluch, Tim Genewein, Andreas Nürnberger, and Jan M Köhler. The power of ensembles for active learning in image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9368–9377, 2018.

[3] Mustafa Bilgic and Lise Getoor. Link-based active learning. In *NIPS Workshop on Analyzing Networks and Learning with Graphs*, 2009.

[4] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[5] Ido Dagan and Sean P Engelson. Committee-based sampling for training probabilistic classifiers. In *Machine Learning Proceedings 1995*, pages 150–157. Elsevier, 1995.

[6] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1422–1430, 2015.

[7] In So Kweon Donggeun Yoo. Learning loss for active learning. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2019.

[8] Melanie Ducoffe and Frederic Precioso. Adversarial active learning for deep networks: a margin based approach. *arXiv preprint arXiv:1802.09841*, 2018.

[9] Sayna Ebrahimi, Mohamed Elhoseiny, Trevor Darrell, and Marcus Rohrbach. Uncertainty-guided continual learning with bayesian neural networks. *arXiv preprint arXiv:1906.02425*, 2019.

[10] Sayna Ebrahimi, Anna Rohrbach, and Trevor Darrell. Gradient-free policy architecture search and adaptation. *arXiv preprint arXiv:1710.05958*, 2017.

[11] Ehsan Elhamifar, Guillermo Sapiro, Allen Yang, and S Shankar Sasrty. A convex optimization framework for active learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 209–216, 2013.

[12] Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Pattern Recognition Workshop*, 2004.

[13] Alexander Freytag, Erik Rodner, and Joachim Denzler. Selecting influential examples: Active learning with expected model output changes. In *European Conference on Computer Vision*, pages 562–577. Springer, 2014.

[14] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.

[15] Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38(none):293–306.

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.

[17] Marc Gorriz, Axel Carlier, Emmanuel Faure, and Xavier Giro-i Nieto. Cost-effective active learning for melanoma segmentation. *arXiv preprint arXiv:1711.09168*, 2017.

[18] Mahmudul Hasan and Amit K Roy-Chowdhury. Context aware active learning of activity recognition models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4543–4551, 2015.

[19] Armand Joulin, Laurens van der Maaten, Allan Jabri, and Nicolas Vasilache. Learning visual features from large weakly supervised data. In *European Conference on Computer Vision*, pages 67–84. Springer, 2016.

[20] Christoph Käding, Erik Rodner, Alexander Freytag, and Joachim Denzler. Active and continuous exploration with deep neural networks and expected model output changes. *arXiv preprint arXiv:1612.06129*, 2016.

[21] Ashish Kapoor, Kristen Grauman, Raquel Urtasun, and Trevor Darrell. Active learning with gaussian processes for object categorization. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[22] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[23] Vikram Krishnamurthy. Algorithms for optimal scheduling and management of hidden markov model sensors. *IEEE Transactions on Signal Processing*, 50(6):1382–1397, 2002.

[24] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.

[25] Weicheng Kuo, Christian Häne, Esther Yuh, Pratik Mukherjee, and Jitendra Malik. Cost-sensitive active learning for intracranial hemorrhage detection.

[26] Liang Li, Shuqiang Jiang, and Qingming Huang. Learning hierarchical semantic description via mixed-norm regularization for image understanding. *IEEE Transactions on Multimedia*, 14(5):1401–1413, 2012.

[27] Xin Li and Yuhong Guo. Adaptive active learning for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 859–866, 2013.

[28] Xuejing Liu, Liang Li, Shuhui Wang, Zheng-Jun Zha, Dechao Meng, and Qingming Huang. Adaptive reconstruction network for weakly supervised referring expression grounding. In *IEEE/CVF International Conference on Computer Vision, ICCV*, pages 2611–2620. IEEE, 2019.

[29] Oisin Mac Aodha, Neill DF Campbell, Jan Kautz, and Gabriel J Brostow. Hierarchical subquery evaluation for active learning on a graph. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 564–571, 2014.

[30] Dhruv Mahajan, Ross Girshick, Vignesh Ramanathan, Kaiming He, Manohar Paluri, Yixuan Li, Ashwin Bharambe, and Laurens van der Maaten. Exploring the limits of weakly supervised pretraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 181–196, 2018.

[31] Dwarikanath Mahapatra, Behzad Bozorgtabar, Jean-Philippe Thiran, and Mauricio Reyes. Efficient active learning for

image classification and segmentation using a sample selection and conditional generative adversarial network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 580–588. Springer, 2018.

[32] Christoph Mayer and Radu Timofte. Adversarial sampling for active learning. *arXiv preprint arXiv:1808.06671*, 2018.

[33] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[34] Hieu T Nguyen and Arnold Smeulders. Active learning using pre-clustering. In *ICML*, page 79. ACM, 2004.

[35] Mehdi Noroozi, Hamed Pirsiavash, and Paolo Favaro. Representation learning by learning to count. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5898–5906, 2017.

[36] Nicholas Roy and Andrew McCallum. Toward optimal active learning through monte carlo estimation of error reduction. *ICML, Williamstown*, pages 441–448, 2001.

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

[38] Burr Settles, Mark Craven, and Soumya Ray. Multiple-instance active learning. In *Advances in neural information processing systems*, pages 1289–1296, 2008.

[39] Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational adversarial active learning. *arXiv preprint arXiv:1904.00370*, 2019.

[40] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in neural information processing systems*, pages 3483–3491, 2015.

[41] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 5693–5703, 2019.

[42] Lin Yang, Yizhe Zhang, Jianxu Chen, Siyuan Zhang, and Danny Z Chen. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 399–407. Springer, 2017.

[43] Shijie Yang, Liang Li, Shuhui Wang, Weigang Zhang, Qingming Huang, and Qi Tian. Skeletonnet: A hybrid network with a skeleton-embedding process for multi-view image representation learning. *IEEE Transactions on Multimedia*, 21(11):2916–2929, 2019.

[44] Yi Yang, Zhigang Ma, Feiping Nie, Xiaojun Chang, and Alexander G Hauptmann. Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision*, 113(2):113–127, 2015.

[45] Zheng-Jun Zha, Meng Wang, Yan-Tao Zheng, Yi Yang, Richang Hong, and Tat-Seng Chua. Interactive video indexing with statistical active learning. *IEEE Trans. Multimedia*, 14(1):17–27, 2012.

[46] Zhi Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, v.5(1):48–57.

[47] Jia-Jie Zhu and José Bento. Generative adversarial active learning. *arXiv preprint arXiv:1702.07956*, 2017.