Semi-supervised Active Learning for Semi-supervised Models: Exploit Adversarial Examples with Graph-based Virtual Labels

Jiannan Guo[1,2*] Haochen Shi[3*] Yangyang Kang[2] Kun Kuang[1] Siliang Tang[1†]
Zhuoren Jiang[1] Changlong Sun[1,2] Fei Wu[1] Yueting Zhuang[1]
[1]Zhejiang University [2]Alibaba Group [3]Université de Montréal
{jiannan, kunkuang, siliang, jiangzhuoren, wufei, yzhuang}@zju.edu.cn
yangyang.kangyy@alibaba-inc.com, changlong.scl@taobao.com, haochen.shi@umontreal.ca

Abstract

The performance of computer vision models significantly improves with more labeled data. However, the acquisition of labeled data is limited by the high cost. To mitigate the reliance on large labeled datasets, active learning (AL) and semi-supervised learning (SSL) are frequently adopted. Although current mainstream methods begin to combine SSL and AL (SSL-AL) to excavate the diverse expressions of unlabeled samples, these methods' fully supervised task models are still trained only with labeled data. Besides, these method's SSL-AL frameworks suffer from mismatch problems. Here, we propose a graph-based SSL-AL framework to unleash the SSL task models' power and make an effective SSL-AL interaction. In the framework, SSL leverages graph-based label propagation to deliver virtual labels to unlabeled samples, rendering AL samples' structural distribution and boosting AL. AL finds samples near the clusters' boundary to help SSL perform better label propagation by exploiting adversarial examples. The information exchange in the closed-loop realizes mutual enhancement of SSL and AL. Experimental results show that our method outperforms the state-of-the-art methods against classification and segmentation benchmarks.

1. Introduction

The development of deep learning brings prosperity to the field of computer vision [23, 39, 3, 5, 57, 53, 54, 55, 56, 26, 25, 27]. However, these data-hungry models still need to be fed with a large amount of labeled data, the acquisition of which is limited by the expensive cost of annotation [58]. This dilemma between performance and cost brings massive research and practical value to achieve a higher performance of the task models with limited labeled data.

With awareness of this problem, active learning (AL) [52, 50, 29, 34] is introduced to unleash the potential of labeling procedures for budget-limited annotation. Despite the progress achieved, most AL algorithms suffer from data wasting problems as they ignore that utilizing AL in real-world scenarios means that the majority of data remains unlabeled, which could further empower AL with semi-supervised learning (SSL) in the following three ways: 1) (SSL Task Model) When numerous unlabeled data are used in conjunction with a bunch of labeled data, it is very natural and practical to further improve the performance of the task models by SSL without introducing any further annotation cost. 2) (SSL→AL) The success of SSL [20, 41, 4, 42, 46] proves that it is feasible to improve performance by modeling the relation (intra-class similarity and inter-class distinguishability) among samples with SSL. Thus AL could assess samples' annotation value more accurately, being guided by the prior knowledge of SSL-modeled relation. 3) (AL→SSL) Since the initial samples' relations modeled by SSL can hardly be completely correct. AL could confirm precise relations and rectify wrong relations by labeling specific samples. Finally, the mutual enhancement of AL and SSL is achieved in such a loop.

Although several works combine SSL and AL (SSL-AL) [52, 19, 40], these methods suffer from the two problems: 1) Their fully supervised task models are subjected to data-wasting problems. 2) These works [40, 19, 52] are based on the VAE-GAN structure and conduct AL on the latent representation of samples, which is learned through a mini-max game on samples' labeling states. As is illustrated in Figure 1, this kind of method suffers from the mismatch problem, which decreases AL's efficiency.

Based on the above insights, we propose a novel g**R**aph-bas**E**d **VI**rtual ad**V**ersarial **A**ctive **L**earning (REVIVAL) framework for semi-supervised models. The REVIVAL is free of the above two deficiencies and realizes the mutual enhancement of AL and SSL. This framework mainly con-

*Equal contribution.
†Corresponding author.

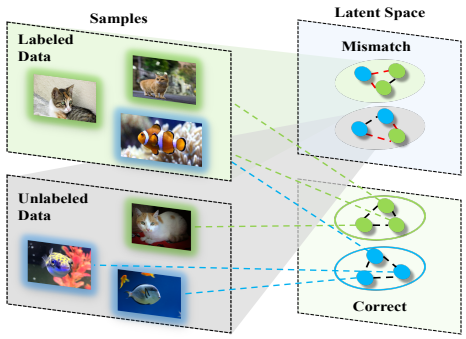

Figure 1: Since samples' labeling states and class labels are uncorrelated, these methods tend to project the representation of samples with different class labels (e.g., cat and fish) to the same class (e.g., labeled or unlabeled), and vice versa.

sists of three components: 1) the **label propagator**. To explicitly model the SSL-inferred samples' relations without losing generality, we adopt the well-explored graph-based label propagation to conduct SSL and boost AL. Specifically, the label propagator first explicitly models the clusters formed by samples with their relations as a graph, then propagates label information on the graph. Finally, the passed messages merge into graph-based virtual labels. Consequently, SSL enhances AL by rendering samples' structural distribution and reducing unlabeled pool's uncertainty with the virtual labels. 2) the **virtual adversarial generator**. To boost SSL by refining the graph, we introduce the adversarial generator to select samples lying on the clusters' boundaries as annotation candidates. Specifically, since the label propagator's message passing would be weaker or even inconsistent on clusters' boundaries, the virtual adversarial generator could identify samples near the boundaries by measuring the inconsistency between samples and the corresponding generated adversarial examples. 3) the **boundary limitator**. To identify the hard cases in the unlabeled pool, we further propose a boundary limitator to re-rank the annotation candidates based on the entropy of the graph-based virtual labels.

We conduct experiments on classification and segmentation benchmarks. The experimental results demonstrate that REVIVAL leads to consistent improvement over the previous state-of-the-art (SOTA) methods. The major contributions of this paper are summarized as follows:

1. We propose a novel graph-based SSL-AL framework for SSL task models. This framework achieves closed-loop mutual enhancement between SSL and AL, which further unleashes the power of SSL task models.

2. We propose the AL enhanced with the virtual adversarial generator, which cooperates with SSL effectively by exploiting the SSL boosted adversarial examples.

3. Extensive experiments on standard benchmarks demonstrate the effectiveness of the proposed method, with significant improvement over several SOTA methods (up to 10% labeling demand reduced).

2. Related Work

Pool-based active learning In practice, it is easy to acquire abundant unlabeled samples. Thus AL in the pool-based [24, 52, 50, 36, 40, 9] scenario is more popular than the other two scenarios: steam-based [7] and membership query synthesis [2, 49, 44, 59, 29].

Uncertainty-based sampling [17, 13, 8, 10, 21] and distribution-based sampling [18, 32, 47, 33] are common methods in the pool-based scenario. Our method considers both uncertainty and distribution. For uncertainty-based frameworks, some past heuristic algorithms [35, 38, 16, 28, 37] have built the foundation for the field of AL. In the era of deep-learning-based active learning, [50] predicts each sample's training loss to measure uncertainty and selects samples with the biggest predicted loss. Nevertheless, the selected samples' quality highly relies on the task model's loss information, which is unstable in the early selection stage. In distribution-based methods [36, 48], the relation between the subset and the whole set is shown intuitively from the perspective of geometry. When scaling to large scale datasets or high-dimension input data, greedy procedures become computationally infeasible. Thus these algorithms will suffer from inefficient computing.

Semi-supervised active learning Recently, [40, 19, 52] leverage the VAE-GAN structure to learn the representation of labeled and unlabeled samples in latent space together with the discriminator. The discriminator aims to discriminate samples belonging to the unlabeled pool through a mini-max game with VAE-GAN. However, these methods have a concern that they do not take class labels into account. Specifically, some samples with the different true labels (*e.g.* fish and cat) are classified into the same class (labeled or unlabeled), and samples with the same true labels may fall into opposite classes. As a consequence, the learning process may harm the semantic distribution of samples in latent space. In fact, the annotation state and the feature representation are orthogonal, while the feature representation is highly correlated with class labels of different classes. Compared to these VAE-GAN based methods, our method considers the relation between class labels and feature representation to combine AL and SSL smoothly. The latest works [43, 11] combine AL and SSL based on prediction consistency given a set of data augmentations. However, these methods only use a limited number of ways of data augmentation to estimate inconsistency. In contrast, our method can explore the continuous local distribution of unlabeled samples in feature space and obtain more semantic distribution information. Furthermore, we consider the uncertainty of samples and the feature distribution in the latent space simultaneously.

3. Method

In this section, we formulate the proposed g**R**aph-bas**E**d **VI**rtual ad**V**ersarial **A**ctive **L**earning (REVIVAL) algorithm. We first provide an overview of the whole AL system in Section 3.1, then introduce three main components of REVIVAL: a label propagator in Section 3.2, a virtual adversarial generator in Section 3.3 and a boundary limitator in Section 3.4.

3.1. Overview

In this subsection, we formally define the pool-based AL loop with the proposed REVIVAL. As is demonstrated in Figure 2, the initial labeled pool $\mathcal{D}^l = \{(x_1^l, y_1^l), \ldots, (x_{N_l}^l, y_{N_l}^l)\}$ and the unlabeled pool $\mathcal{D}^u = \{x_1^u, \ldots, x_{N_u}^u\}$ (where N_l and N_u are the numbers of labeled samples and unlabeled samples respectively) are given. The algorithm feeds them into the task model to obtain corresponding representations $\mathbf{R}^l = \{r_1^l, \ldots, r_{N_l}^l\}$ and $\mathbf{R}^u = \{r_1^u, \ldots, r_{N_u}^u\}$. Based on extracted representations, the graph structure and an adjacency matrix are constructed. Then a graph convolutional network (GCN) [20] based label propagator is trained using the constructed graph structure and the adjacency matrix to propagate label information. Once the training is finished, we will feed unlabeled samples' virtual labels $\bar{\mathbf{Y}}^u = \{\bar{y}_1^u, \ldots, \bar{y}_{N_u}^u\}$ inferred by the label propagator to the virtual adversarial generator to generate adversarial samples. After that, we select out top-$\mathcal{M}$ samples based on the divergence between unlabeled samples and their adversarial samples. Furthermore, we apply the boundary limitator to these candidates to select top-$\mathcal{K}$ candidates with the largest uncertainty and provide them to the human oracle for annotation. Consequently, the sizes of the labeled pool and unlabeled pool will be updated to $N_l + \mathcal{K}$ and $N_u - \mathcal{K}$ respectively. The loop will be repeated until the performance of the task model meets requirements or the budget for annotation is run out.

3.2. Label Propagator

We propose a semi-supervised label propagator to excavate class labels information from unlabeled samples and assist AL algorithms to evaluate correlative metrics of unlabeled samples. Specifically, We construct the pre-extracted feature representation set $\mathbf{R} = \{r_1^l, \ldots, r_{N_l}^l, r_1^u, \ldots, r_{N_u}^u\}$, where r_i is the hidden state from task model's feature extractor. A sparse cosine similarity distance matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ with elements

$$s_{ij} = \begin{cases} 1 - \frac{{r_i}^T \cdot r_j}{||r_i|| \times ||r_j||}, & \text{if } i \neq j \wedge r_i \in \text{NN}_k(r_j) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

is constructed using $\mathbf{R}$, where $\text{NN}_k(r_j)$ denotes the set of k nearest neighbors of r_j, and $n = N_l + N_u$. Note that constructing the distance matrix of the nearest neighbor graph can be efficient even for a large n [14]. Based on $\mathbf{S}$, we get adjacency matrix $\mathbf{A}$ to explore class labels of unlabeled samples via their feature representation in the latent space. In fact, samples with their neighbors naturally form clusters in feature space, and samples belonging to the same cluster are more likely to share the same true labels. Given this structural prior, we further utilize a multi-layer GCN to propagate intra-cluster samples' feature and label information to unlabeled samples under the guidance of the adjacency matrix $\mathbf{A}$, which can be formulated:

$$\mathbf{H}^{(l+1)} = g^{(l)}(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}) \quad (2)$$

where $l = 0, \ldots, L-1$ is the index of current GCN layer, $g^{(l)}$ is the activation function for layer l, $W^{(l)}$ is the trainable weight matrix for l_{th} layer, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix normalized by adding self-loop, $\mathbf{I}$ is the identity matrix, $\tilde{\mathbf{D}}_{ii} = \Sigma_j \tilde{\mathbf{A}}_{ij}$ is the degree matrix of $\tilde{\mathbf{A}}$. In this paper, we set the number of layers $L = 2$, $g^{(0)} = ReLU(\bullet)$, $g^{(1)} = softmax(\bullet)$, $\mathbf{H}^{(0)} = \mathbf{R}$, and take the output of the last layer as the semi-supervised posterior probability inferred by the label propagator: $\bar{\mathbf{Y}} = \mathbf{H}^{(2)}$. $\bar{\mathbf{Y}}$ can serve as the graph-based virtual label, which will further boost the AL algorithm in following sections.

Unlabeled samples acquire propagated label information by aggregating their neighbor nodes' features. Since the feature matrix should multiply the adjacency matrix $\tilde{\mathbf{A}}$ during the inferring process of GCN, label and feature flow into unlabeled samples through the adjacency matrix $\tilde{\mathbf{A}}$. GCN will output smoothed predictions with the impact of neighbors' features for unlabeled samples. These 'virtual' predictions can provide more supervised information for AL.

Finally, we train this label propagator in each AL cycle to assimilate the adjacency relation and perform correct message passing using the following cross-entropy loss:

$$L_{semi} = -\sum_{x \in \mathcal{D}^l} \sum_{c=1}^{C} \mathbf{Y}_c^l \, ln \bar{\mathbf{Y}}_c^l \quad (3)$$

where l indicates labeled samples, C is the number of classes, $\mathbf{Y}_c^l$ denotes the sample has ground-truth label c.

3.3. Virtual Adversarial Generator

Since the constructed SSL graph can hardly be perfect, we introduce the virtual adversarial generator to refine the graph (*i.e.*, confirm precise relations and correct wrong relations in the SSL graph) and further improve the performance of the SSL models. Specifically, the model predicts more sharply on the clusters' boundaries, as the label propagator's message passing would be more consistent and stronger within clusters but weaker or even contradictory on clusters' boundaries and inter-clusters, which is caused by

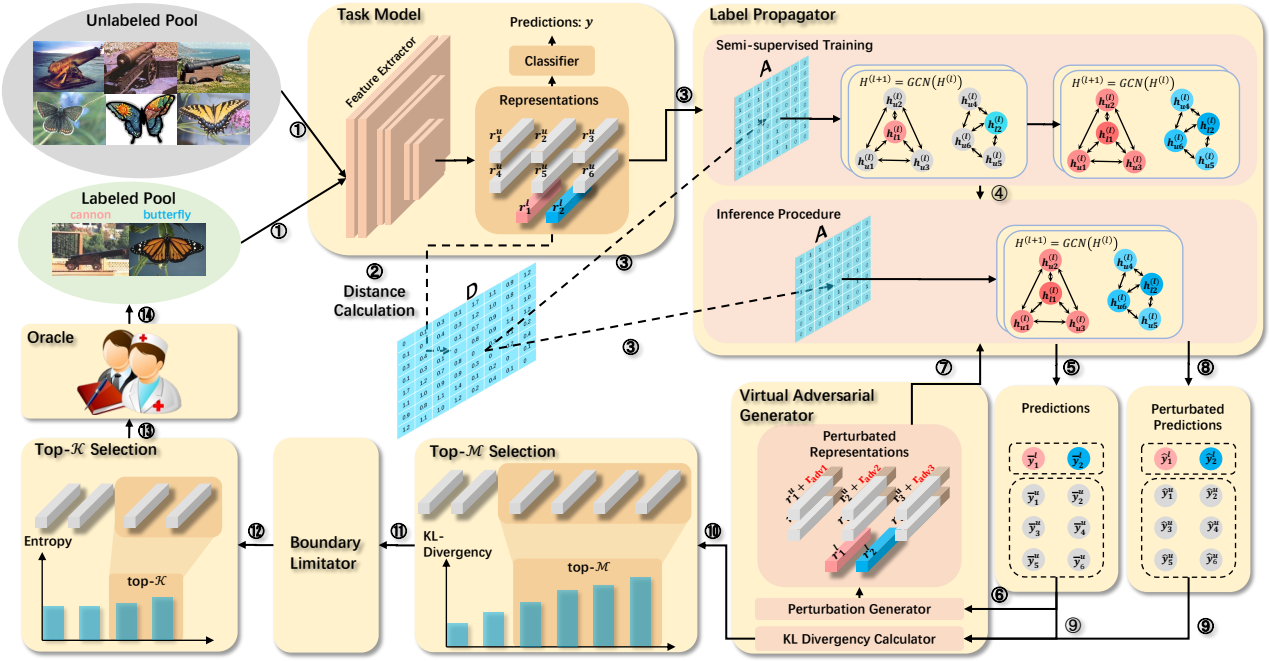

Figure 2: Overview of our method. It consists of three modules: (a) *Label Propagator* constructs the adjacent matrix using extracted feature embedding from the task model and propagates virtual labels to unlabeled samples. (Section 3.2); (b) *Virtual Adversarial Generator* produces adversarial examples for unlabeled samples and measures the difference of posterior probability between unlabeled samples and their adversarial examples (Section 3.3); (c) *Boundary limitator* further imposes uncertainty restrictions on samples (Section 3.4).

the unconfidently and improperly modeled relations. Therefore, labeling these samples could improve SSL models' ability to distinguish different classes and boost SSL by refining the graph.

To find samples near the boundary, the virtual adversarial generator estimates the smoothness of the model's prediction on samples by calculating the inconsistency between prediction on samples and on corresponding adversarial examples. Specifically, given the extracted representation r_i^u of unlabeled sample x_i^u, we first feed r_i^u into the propagator to get its prediction $\bar{y}_i^u$. Then we feed r_i^u and $\bar{y}_i^u$ simultaneously into the generator to get adversarial perturbation r_{adv}^i. After that, we feed perturbed representation $\hat{r}_i^u = r_i^u + r_{adv}^i$ into the label propagator again to get its perturbed prediction $\hat{y}_i^u$. Finally, the generator calculates the KL-divergence $KLD_i^u = \mathrm{KL}(\hat{y}_i^u, \bar{y}_i^u)$ between $\bar{y}_i^u$ and $\hat{y}_i^u$. Samples with larger KLD_i^u are closer to corresponding clusters' boundaries. Here, the adversarial perturbation is formulated as:

$$r_{adv} = \underset{\Delta r, ||\Delta r|| \le \epsilon}{\arg\max}\ \mathrm{KL}(p(\bar{y}^u|r^u, \theta), p(\hat{y}^u|r^u + \Delta r, \theta)) \quad (4)$$

where $p(y|x, \theta)$ represents the posterior probability distribution of the propagator. Under perturbation of the same norm ϵ, there is a higher probability for adversarial samples of unlabeled samples near the boundary to change their original class label and fall into the other cluster. Therefore, the generator computes the KL-divergence of the posterior probability of samples and their adversarial examples to measure unlabeled samples' distance to the boundary.

Since the computation of r_{adv} is intractable for many neural networks, [30] proposed to approximate r_{adv} using the second-order Taylor approximation and solved the r_{adv} via the power iteration method. Specifically, we can approximate r_{adv} by applying the following update:

$$r_{adv} \leftarrow \epsilon \overline{\nabla_{\Delta r} \mathrm{KL}(p(\bar{y}^u|r^u), p(\hat{y}^u|r^u + \Delta r))} \quad (5)$$

where Δr is a randomly sampled unit vector, $p(\bar{y}^u|r^u)$ is the graph-based virtual label, $p(\hat{y}^u|r^u + \Delta r)$ is the perturbed prediction, and the sign $\overline{v}$ means the unit vector of v. The computation of $\nabla_{\Delta r}\mathrm{KL}$ can be performed with one set of backpropagation for NNs. Once the r_{adv} is solved, we can estimate the distance of unlabeled samples to the clusters' boundary by computing KLD_i^u $(i = 1, \ldots, N_u)$ and select top-$\mathcal{M}$ samples with the largest KLD^u as the annotation candidates.

3.4. Boundary Limitator

Although the generator selects annotation candidates near the clusters' boundary, the task model still holds different prediction confidence for these annotation candidates.

To further maximize the effect of the limited annotation budget, we introduce the boundary limitator, which limits the finally selected samples with high uncertainty. These finally selected samples can bring significant model uncertainty reduction and model information gain indeed. In practice, we use the graph-based virtual labels' entropy to estimate the uncertainty of samples and select top-$\mathcal{K}$ samples with the largest entropy for human annotation. The entropy for the unlabeled sample x_i^u can be calculated with the following formulation:

$$E_i^u = -\sum_{c}^{C} P(\bar{y}_{ic}^u | r_i^u)\, log P(\bar{y}_{ic}^u | r_i^u) \tag{6}$$

where $P(\bar{y}_{ic}^u | r_i^u)$ is the probability of the unlabeled sample x_i^u belonging to the c-th class.

Above all, on the one hand, the SSL label propagator could exploit labeled samples' feature and provide AL with graph-based virtual labels that integrate labeled samples' supervised information. On the other hand, the virtual adversarial generator and the boundary limitator will let the oracle annotate samples near the cluster's boundary and with the largest uncertainty. These samples could assist the semi-supervised label propagator in distilling correct boundary information for better label propagation and decreasing the unlabeled pool's uncertainty. Consequently, the information exchange in the closed-loop realizes the mutual enhancement between SSL and AL. Furthermore, SSL task models also benefit from label propagation to improve the performance by using unlabeled data, which is why our SSL-AL framework can unleash the power of SSL task models.

4. Experiments

In this section, we first conduct experiments for different tasks (image classification and semantic segmentation) on several benchmarks under different settings (supervised and semi-supervised task models), and then discuss REVIVAL's property with controlled experiments.

Dataset For image classification, we use CIFAR-10 [22] and CIFAR-100 [22] as our benchmarks. They both contain 60,000 images, of which 50,000 images are for training. The CIFAR-10 has ten categories with 6000 images per category, while the CIFAR-100 has 100 classes with 600 images per class. In the segmentation task, methods are evaluated on the Cityscapes dataset [6] for pixel-level segmentation annotation. The dataset collects 3475 pixel-level labeled urban street scene pictures, 2975 for training, and 500 for testing.

Baseline In the comparison of classification under SSL, we consider four baselines. ICAL [11] selects samples with the high inconsistency of predictions over a set of data augmentations. Core-set [36] is a distribution-based sampling algorithm that selects a subset to cover the whole set's distribution. Entropy [37] is commonly used as an uncertainty-based baseline. It selects samples with the maximum entropy of its prediction probabilities. Random sampling often serves as the lower bound of AL algorithms. In the supervised comparison of classification, we also consider three other baselines. SRAAL [52] and VAAL [40] leverage the VAE-GAN structure to learning embedding in latent space and find out samples that are more likely to fall into the unlabeled pool. LLAL [50] selects samples with the biggest training loss. Besides we adopt the current state-of-the-art method CDAL [1] as one of our baselines for the segmentation task. The method considers spatial neighborhoods of instances. It selects samples by modeling spatial co-occurrence and spatial context of instances.

Implementation details In the classification under SSL, we adopt Mixmatch [4] as our SSL task model, where Wide ResNet-28 [31] is the backbone of the task model. We keep the default hyper-parameters for different datasets following [4]. The initial training set is uniformly distributed over classes, and we set up the initial set by randomly sampling. We follow [11] for the setting of initial training set size and AL budget. In supervised classification tasks and the segmentation task, we follow [52] for AL settings. We use ResNet-18 [12] and DRN [51] as backbones of task models for classification and segmentation respectively. Different baselines are trained with the same initial training set and model parameters for a fair comparison. We plot the charts with the experiments conducted five times: solid lines indicate the results averaged over five trials, and shadows represent the standard deviation. We adopt a billion-scale similarity search tool called faiss [15] to build the KNN graph structure at a low cost and use Deep Graph Library[1] to conduct fast GCN computation.

4.1. Comparison under Supervised Learning

In this section, we conduct experiments to verify the proposed semi-supervised AL framework's effectiveness under supervised learning, where the contribution to the performance of the task model mainly comes from AL modules.

4.1.1 REVIVAL Performance on CIFAR-10

Figure 3 shows the performance of our method under supervised learning on CIFAR-10. We can observe that our method outperforms state-of-the-art methods in all selection cycles. Using the entire datasets, Resnet-18 can yield the highest accuracy of 93.5%, while our method outperforms the performance of fully supervised training only using 35% data. It can be explained that our method can select the most informative samples, and these samples' feature distribution benefits the task model's training the most.

[1] https://www.dgl.ai/

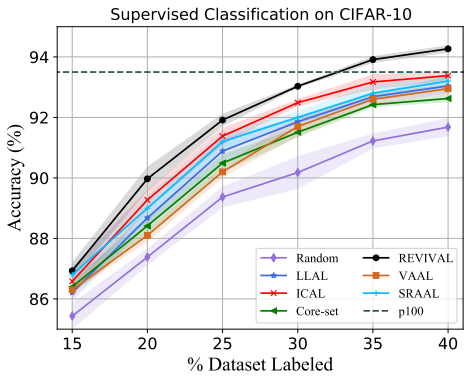

Figure 3: Performance comparison under supervised learning on the CIFAR-10. The dotted line (p100) at the top represents the performance with the entire training set labeled.

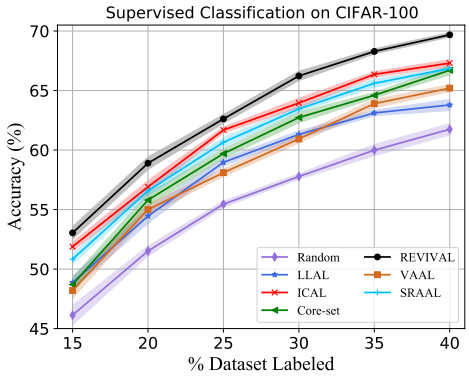

Figure 4: Performance comparison under supervised learning on the CIFAR-100.

4.1.2 REVIVAL Performance on CIFAR-100

Figure 4 shows the supervised performance of REVIVAL on the dataset CIFAR-100. In the figure, our method beats all other baselines in all selection cycles with a margin up to 2.38%. Without a SSL module, ICAL is equivalent to measuring samples' prediction stabilities. LLAL only uses predicted loss to estimate uncertainty. Similarly, core-set only considers distribution diversity to select samples. The incomplete consideration of sample information limits the performance of the LLAL, core-set and ICAL. Differently, our SSL-AL framework takes both uncertainty and distribution into account in SSL and AL's effective interaction. VAAL leverages VAE-GAN to build feature embedding in latent space and discriminate unlabeled data by its label state. The estimation of whether the sample is labeled is not equal to informativeness. Thus the method can not select the most informative samples. Compared to VAAL, our method considers the relation between class labels of samples and their representation similarity.

4.2. Comparison under Semi-supervised Learning

Most previous AL approaches focus on the setting where the task model is trained in a fully supervised manner, leaving massive unlabeled data wasted. To verify our intuition and illustrate the effectiveness of REVIVAL, we compare REVIVAL with other AL algorithms and demonstrate the performance gap between the SSL task model and the SL task model (up to 12.1%).

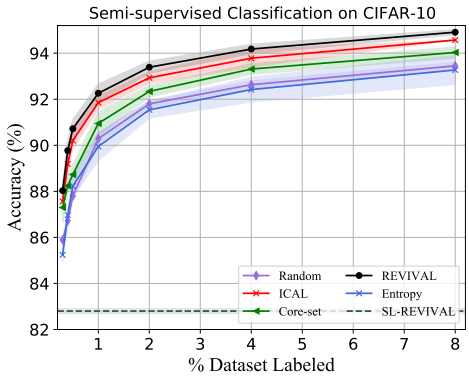

Figure 5: Performance comparison under semi-supervised learning on the CIFAR-10. The dotted line (SL-REVIVAL) at the bottom represents the performance of REVIVAL under supervised learning with 8% data labeled.

4.2.1 REVIVAL Performance on CIFAR-10

In Figure 5, all semi-supervised baselines that use only 0.3% labeled data outperform supervised REVIVAL, which uses 8% labeled data. Our method under semi-supervised learning consistently outperforms other baselines as the sample selection cycle progresses. It is worth noting that a fully entropy-based method's performance is even inferior to that of random sampling, which may be caused by the problem of overconfident mis-classification [11]. However, under the guide of graph-based SSL and the virtual adversarial generator, the boundary limitator in REVIVAL could use entropy to estimate the uncertainty effectively. Considering the samples' cluster structure in latent space, the core-set tends to select samples at centers of clusters to represent each cluster's feature, which may be limited by over-exploit and under-exploration. In contrast, REVIVAL selects samples near the clusters' boundary to accelerate structural information exploration.

4.2.2 REVIVAL Performance on CIFAR-100

Figure 6 demonstrates the results of REVIVAL's performance on CIFAR-100 under semi-supervised learning. We can see that semi-supervised REVIVAL outperforms all previous baselines throughout the whole sample selection stage. When using 25% labeled data, semi-supervised REVIVAL achieves 9.33% more accuracy than supervised REVIVAL. Core-set performs ineffectively when the number of categories gets bigger, compared to the performance on CIFAR-10. ICAL takes several data augmentation methods

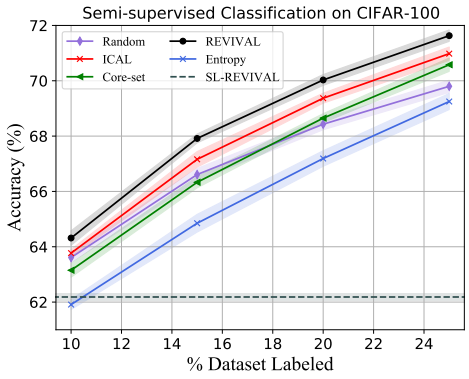

Figure 6: Performance comparison under semi-supervised learning on the CIFAR-100. The dotted line (SL-REVIVAL) at the bottom represents the performance of REVIVAL under supervised learning with 25% data labeled.

	Accuracy % on % Labeled data	
Method	15	25
REVIVAL	67.91±0.16	71.63±0.20
LP+VAG	67.71±0.18	71.28±0.20
VAG+BL	67.27±0.18	70.73±0.19
LP+BL	67.38±0.16	70.85±0.17
Random	66.61±0.22	69.82±0.22

Table 1: Ablation study under semi-supervised learning over CIFAR-100

to measure inconsistency loss and obtain an average accuracy of 70.98% in the last cycle. In contrast, REVIVAL reaches the accuracy of 71.63%, benefiting from the exploration of continuous local distribution around unlabeled samples and the consideration of the feature distribution in the latent space.

5. Model Analysis

5.1. Ablation Study

Table 1 demonstrates the ablation study to evaluate the contribution of critical modules of REVIVAL on the CIFAR-100 dataset under semi-supervised learning.

Effect of the label propagator (LP) The label propagator aggregates supervised information for unlabeled samples, and outputs smoothed prediction probabilities of unlabeled samples. Without smoothed signals from the propagator, the performance of REVIVAL degrades from 71.63% to VAG+BL's 70.73% with 25% labeled data. The Standalone boundary limitator is even inferior to random selection without smoothed signals in Figures 5 and 6. The label propagator can fully excavate the AL algorithm's great performance potential by utilizing abundant unlabeled data.

Effect of the virtual adversarial generator (VAG) Compared to all ablation results, the adversarial generator makes a significant contribution to the algorithm. The generator can work closely with the propagator exploring structural information and finding samples near the clusters' boundary. The generator improves the performance from 70.85% to 71.63% in the last selection stage.

Effect of the boundary limitator (BL) Although the semi-supervised AL structure, LP+VAG, can achieve an excellent performance (67.71% and 71.28%), ensuring high uncertainty of selected samples can still improve the model's performance (67.91% and 71.63%) further. The ablation results show that our complete method has the best performance.

5.2. Effect of Hyperparameter $\mathcal{M}$ on REVIVAL

Figure 7(a) illustrates the effect of parameter $\mathcal{M}$ on REVIVAL against the CIFAR-10 dataset. Besides, Top-$\mathcal{K}$ is not a hyperparameter as it is equal to the budget size of sample selection (5% of the whole set) at each stage. We can observe that when $\mathcal{M}$=2500 (budget size), the algorithm will degenerate into selecting samples only based on the generator. As the number of samples initially screened by the generator increases, the contribution of the boundary limitator to the performance becomes greater. Between the interval from 5500 to 14500, the performance is stable and nonsensitive to $\mathcal{M}$. After $\mathcal{M}$ exceeds 14500, the algorithm prefers to select samples with larger entropy and gradually ignores feature distribution of samples until the algorithm degenerates into selecting samples only based on the boundary limitator. In this phase, the algorithm only considers the uncertainty of samples.

5.3. Effect of Neighbor Number k for KNN Graph

The effect of the nearest neighbors' number in the KNN graph is shown in Figure 7(b). When k=1, the SSL-AL algorithm degenerates into supervised AL as no edge will be constructed in the graph, so that label information cannot flow among samples and no structural information will be provided to AL modules. As the number of k increases, the SSL-rendered structural and label information increases, and therefore the performance improves and reaches the peak when k=20. As the k continues to increase, the performance demonstrates a downward trend. This is because too much noise is introduced into a too-dense graph.

5.4. Analysis of Adjacency Graph Improvement

In order to demonstrate the enhancement effect of AL on SSL, we conduct experiments comparing the accuracy of the constructed graph for SSL on the CIFAR-100 benchmark. As illustrated in Figure 7(c), the accuracy of the adjacency graph constructed by REVIVAL outperforms all other methods consistently and significantly, demonstrating that the AL modules effectively accelerate the graph's evolutionary rate to boost SSL. Results in this subsection and 5.1 demonstrate that the mutual enhancement of AL and SSL is achieved within REVIVAL.

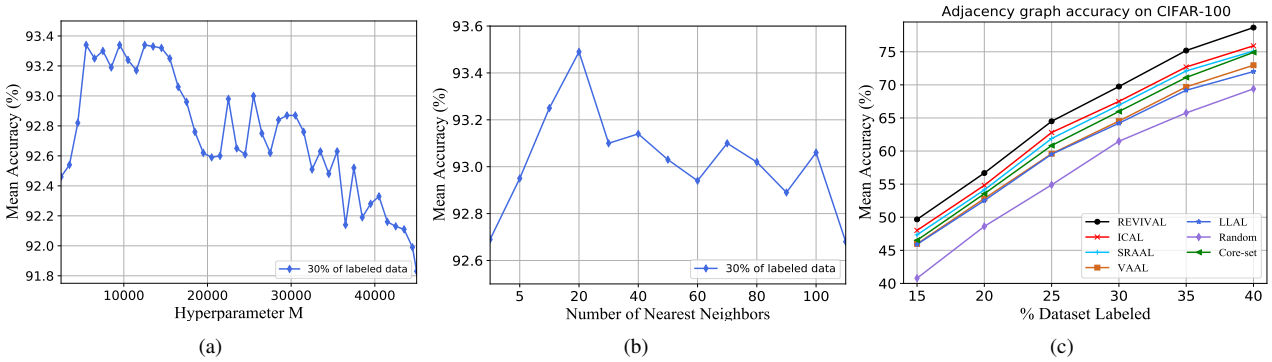

Figure 7: (a)Performance under different hyper-parameters $\mathcal{M}$. (b)Performance under different hyper-parameters k. (c)Results for analyses of adjacency graph construction.

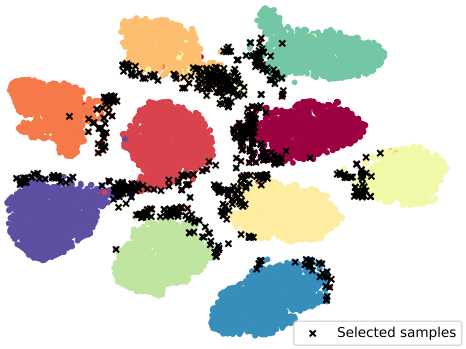

Figure 8: The tSNE embeddings of the CIFAR10 dataset and the selection behavior of our method. Selected samples are shown in black and remaining unlabeled samples in colored.

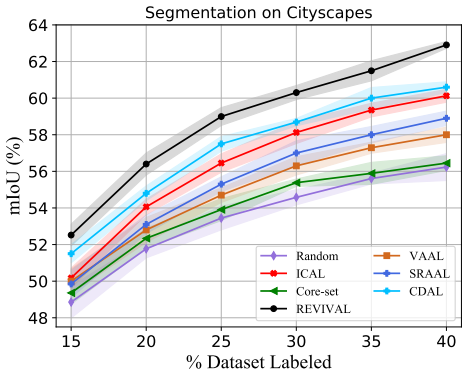

Figure 9: Performances for segmentation on the Cityscapes.

5.5. Visualization Analysis

In Figure 8, we visualize our method's sample selection behavior via tSNE embedding [45]. We compute embeddings for all samples using extracted features learned by labeled samples and visualize the samples selected by our method. This visualization suggests that REVIVAL can select samples, which are near the boundary of clusters and also have the largest uncertainty (overlapping regions of two clusters).

5.6. Analysis of Algorithm Generality

To verify the generality of our algorithm, we also conduct experiments on the semantic segmentation task. For semantic segmentation, we consider pixel-level annotation for AL, and the image-level uncertainty is defined as the sum of the uncertainty of all the pixels in the image. Besides, the task model adopted here is a supervised model. Figure 9 demonstrates the performance. We can observe that REVIVAL significantly outperforms current state-of-the-art methods across different labeled data ratios. In terms of required annotations for each approach, REVIVAL needs 25% training data to assist the task model to reach the mIoU of 58.9%, while the other methods need 10% more labeled data. The results show that our algorithm has good generality and can be extended to another task.

6. Conclusion

This paper proposed a SSL-AL framework (REVIVAL) where AL and SSL achieve mutual enhancement. The SSL-AL framework unleashes the power of SSL task models. In the algorithm, we converted the relation between intra-class similarity and inter-class distinguishability into a graph structure and performed label propagation based on the graph. The propagated label information helped AL find unlabeled samples near the boundary of clusters, which further improved label propagation. The experiment results showed that REVIVAL significantly beat the state-of-the-art approaches on classification and segmentation tasks.

Acknowledgments This work is supported in part by National Key Research and Development Program of China (2018AAA0101900), Zhejiang NSF (LR21F020004), Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies, Alibaba Group through Alibaba Innovative Research Program & Alibaba Research Fellowship Program, Zhejiang University iFLYTEK Joint Research Center, Chinese Knowledge Center of Engineering Science and Technology (CKCEST).

References

[1] Sharat Agarwal, Himanshu Arora, Saket Anand, and Chetan Arora. Contextual diversity for active learning. In *European Conference on Computer Vision*, pages 137–153. Springer, 2020.

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

[3] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[4] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *Advances in Neural Information Processing Systems*, pages 5049–5059, 2019.

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[6] 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.

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

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

[9] 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.

[10] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. *arXiv preprint arXiv:1703.02910*, 2017.

[11] Mingfei Gao, Zizhao Zhang, Guo Yu, Sercan Ö Arık, Larry S Davis, and Tomas Pfister. Consistency-based semi-supervised active learning: Towards minimizing labeling cost. In *European Conference on Computer Vision*, pages 510–526. Springer, 2020.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[13] Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

[14] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, Teddy Furon, and Ondrej Chum. Efficient diffusion on region manifolds: Recovering small objects with compact cnn representations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2077–2086, 2017.

[15] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *arXiv preprint arXiv:1702.08734*, 2017.

[16] Ajay J Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2372–2379. IEEE, 2009.

[17] 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.

[18] Mina Karzand and Robert D Nowak. Active learning in the overparameterized and interpolating regime. *arXiv*, pages arXiv–1905, 2019.

[19] Kwanyoung Kim, Dongwon Park, Kwang In Kim, and Se Young Chun. Task-aware variational adversarial active learning. *arXiv preprint arXiv:2002.04709*, 2020.

[20] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[21] Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. In *Advances in Neural Information Processing Systems*, pages 7026–7037, 2019.

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

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

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

[25] Juncheng Li, Siliang Tang, Fei Wu, and Yueting Zhuang. Walking with mind: Mental imagery enhanced embodied qa. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1211–1219, 2019.

[26] Juncheng Li, Siliang Tang, Linchao Zhu, Haochen Shi, Xuanwen Huang, Fei Wu, Yi Yang, and Yueting Zhuang. Adaptive hierarchical graph reasoning with semantic coherence for video-and-language inference. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[27] Juncheng Li, Xin Wang, Siliang Tang, Haizhou Shi, Fei Wu, Yueting Zhuang, and William Yang Wang. Unsupervised reinforcement learning of transferable meta-skills for embodied navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12123–12132, 2020.

[28] Wenjie Luo, Alex Schwing, and Raquel Urtasun. Latent structured active learning. In *Advances in Neural Information Processing Systems*, pages 728–736, 2013.

[29] Christoph Mayer and Radu Timofte. Adversarial sampling for active learning. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3071–3079, 2020.

[30] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE*

transactions on pattern analysis and machine intelligence, 41(8):1979–1993, 2018.

[31] Avital Oliver, Augustus Odena, Colin Raffel, Ekin D Cubuk, and Ian J Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. *arXiv preprint arXiv:1804.09170*, 2018.

[32] Sujoy Paul, Jawadul H Bappy, and Amit K Roy-Chowdhury. Non-uniform subset selection for active learning in structured data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6846–6855, 2017.

[33] Robert Pinsler, Jonathan Gordon, Eric Nalisnick, and José Miguel Hernández-Lobato. Bayesian batch active learning as sparse subset approximation. In *Advances in Neural Information Processing Systems*, pages 6359–6370, 2019.

[34] Phill Kyu Rhee, Enkhbayar Erdenee, Shin Dong Kyun, Minhaz Uddin Ahmed, and Songguo Jin. Active and semi-supervised learning for object detection with imperfect data. *Cognitive Systems Research*, 45:109–123, 2017.

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

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

[37] Burr Settles and Mark Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 1070–1079, 2008.

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

[39] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[40] Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational adversarial active learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5972–5981, 2019.

[41] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685*, 2020.

[42] Kihyuk Sohn, Zizhao Zhang, Chun-Liang Li, Han Zhang, Chen-Yu Lee, and Tomas Pfister. A simple semi-supervised learning framework for object detection. *arXiv preprint arXiv:2005.04757*, 2020.

[43] Shuang Song, David Berthelot, and Afshin Rostamizadeh. Combining mixmatch and active learning for better accuracy with fewer labels. *arXiv preprint arXiv:1912.00594*, 2019.

[44] Toan Tran, Thanh-Toan Do, Ian Reid, and Gustavo Carneiro. Bayesian generative active deep learning. *arXiv preprint arXiv:1904.11643*, 2019.

[45] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[46] Vikas Verma, Alex Lamb, Juho Kannala, Yoshua Bengio, and David Lopez-Paz. Interpolation consistency training for semi-supervised learning. *arXiv preprint arXiv:1903.03825*, 2019.

[47] Kai Wei, Rishabh Iyer, and Jeff Bilmes. Submodularity in data subset selection and active learning. In *International Conference on Machine Learning*, pages 1954–1963, 2015.

[48] Yuexin Wu, Yichong Xu, Aarti Singh, Yiming Yang, and Artur Dubrawski. Active learning for graph neural networks via node feature propagation. *arXiv preprint arXiv:1910.07567*, 2019.

[49] Yifan Yan, Sheng-Jun Huang, Shaoyi Chen, Meng Liao, and Jin Xu. Active learning with query generation for cost-effective text classification. In *AAAI*, pages 6583–6590, 2020.

[50] Donggeun Yoo and In So Kweon. Learning loss for active learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 93–102, 2019.

[51] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 472–480, 2017.

[52] Beichen Zhang, Liang Li, Shijie Yang, Shuhui Wang, Zheng-Jun Zha, and Qingming Huang. State-relabeling adversarial active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8756–8765, 2020.

[53] Shengyu Zhang, Ziqi Tan, Zhou Zhao, Jin Yu, Kun Kuang, Tan Jiang, Jingren Zhou, Hongxia Yang, and Fei Wu. Comprehensive information integration modeling framework for video titling. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2744–2754, 2020.

[54] Wenqiao Zhang, Haochen Shi, Siliang Tang, Jun Xiao, Qiang Yu, and Yueting Zhuang. Consensus graph representation learning for better grounded image captioning. In *Proc 35 AAAI Conf on Artificial Intelligence*, 2021.

[55] Wenqiao Zhang, Siliang Tang, Yanpeng Cao, Jun Xiao, Shiliang Pu, Fei Wu, and Yueting Zhuang. Photo stream question answer. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3966–3975, 2020.

[56] Wenqiao Zhang, Xin Eric Wang, Siliang Tang, Haizhou Shi, Haochen Shi, Jun Xiao, Yueting Zhuang, and William Yang Wang. Relational graph learning for grounded video description generation. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3807–3828, 2020.

[57] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[58] Zongwei Zhou, Jae Shin, Lei Zhang, Suryakanth Gurudu, Michael Gotway, and Jianming Liang. Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7340–7351, 2017.

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