Two Stream Active Query Suggestion for Active Learning in Connectomics

Zudi Lin[1], Donglai Wei[1], Won-Dong Jang[1], Siyan Zhou[1], Xupeng Chen[2⋆], Xueying Wang[1], Richard Schalek[1], Daniel Berger[1], Brian Matejek[1], Lee Kamentsky[3⋆], Adi Peleg[4⋆], Daniel Haehn[5⋆], Thouis Jones[6⋆], Toufiq Parag[7⋆], Jeff Lichtman[1], and Hanspeter Pfister[1]

[1] Harvard University [2] New York University [3] MIT [4] Google
[5] University of Massachusetts Boston [6] Broad Institute [7] Comcast Research

Abstract. For large-scale vision tasks in biomedical images, the labeled data is often limited to train effective deep models. Active learning is a common solution, where a query suggestion method selects representative unlabeled samples for annotation, and the new labels are used to improve the base model. However, most query suggestion models optimize their learnable parameters only on the limited labeled data and consequently become less effective for the more challenging unlabeled data. To tackle this, we propose a *two-stream active* query suggestion approach. In addition to the supervised feature extractor, we introduce an unsupervised one optimized on all raw images to capture diverse image features, which can later be improved by fine-tuning on new labels. As a use case, we build an end-to-end active learning framework with our query suggestion method for 3D synapse detection and mitochondria segmentation in connectomics. With the framework, we curate, to our best knowledge, the largest connectomics dataset with dense synapses and mitochondria annotation. On this new dataset, our method outperforms previous state-of-the-art methods by 3.1% for synapse and 3.8% for mitochondria in terms of region-of-interest proposal accuracy. We also apply our method to image classification, where it outperforms previous approaches on CIFAR-10 under the same limited annotation budget. The project page is `https://zudi-lin.github.io/projects/#two_stream_active`.

Keywords: Active Learning, Connectomics, Object Detection, Semantic Segmentation, Image Classification

1 Introduction

Deep convolutional neural networks (CNNs) have advanced many areas in computer vision. Despite their success, CNNs need a large amount of labeled data to learn their parameters. However, for detection and segmentation tasks, dense

⋆ Works were done at Harvard University.

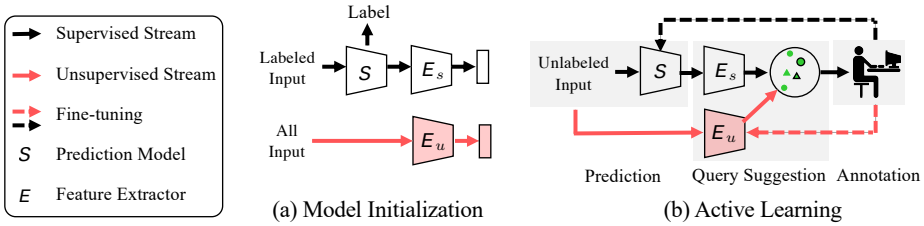

Fig. 1. Two-stream active query suggestion. Active learning methods transform unlabeled data into a feature space to suggest informative queries and improve the base model S. Previous methods optimize their feature extractor (E_s) only on the labeled data. We propose a second one (E_u) trained *unsupervisedly* on all data to capture diverse image features, which can later be updated by fine-tuning with new annotations.

annotations can be costly. Further, in the biomedical image domain, annotations need to be conducted by domain experts after years of training. Thus, under the limited annotation budget, it is critical to effectively select a subset of unlabeled data for annotation to train deep learning models.

Active learning is a common solution that iteratively improves the prediction model by suggesting informative queries for human annotation to increase labels. There are three main categories of query suggestion approaches that have been explored for CNNs: uncertainty-based [44,42,52], expected model change-based [53], and clustering-based methods [43]. However, all these methods use features extracted from CNNs that are trained on the labeled set (Fig. 1a, *!*). For example, core-set [45] uses the last feature space before the classification layer to find representative queries, and learning-loss [53] takes multiple features maps to estimate the loss of the model prediction. Therefore, these methods can be biased towards the feature distribution of the small labeled set. Notably, in many biomedical image applications, the labeled dataset is far from representative of the whole dataset due to its vast quantity and great diversity.

To address this challenge, we propose a *two-stream active clustering* method to improve query suggestion by introducing an additional *unsupervised* feature extractor to capture the image statistics of the whole dataset (Fig. 1a, *!*). During active learning, we combine features extracted by both the supervised and unsupervised streams from the unlabeled data (Fig. 1b). The unsupervised stream can better select representative samples based on image features even when the supervised model makes wrong predictions. Given new annotations, we can further finetune the unsupervised feature extractor to make the embedding space more discriminative. For the clustering module, we show that combining the features from both streams in a *hierarchical* manner achieves significantly better query suggestion performance than directly concatenating the feature vectors.

We test our method in the field of *connectomics*, where the goal is to reconstruct the wiring diagram of neurons to enable new insights into the workings of the brain [26,31]. Recent advances in electron microscopy (EM) allow researchers to collect brain images at nanometer resolution and petabyte scale [21,56]. One

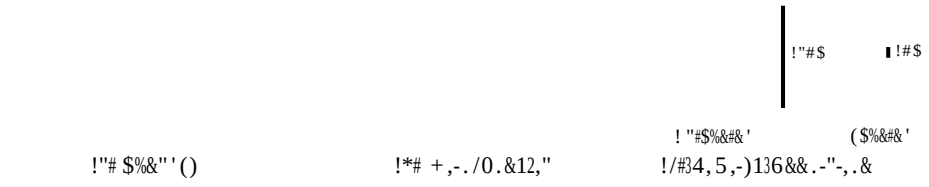

Fig. 2. Two essential vision tasks in connectomics: (a) object detection of synapses to quantify the neuronal connectivity strength and (b) semantic segmentation of mitochondria to estimate the neuronal activity level. (c) However, the terabyte-level test data can be 100 larger than the training data, making active learning necessary.

crucial task is to detect and segment biological structures like synapses and mitochondria for a deeper understanding of neural anatomy and activation patterns [2] (Fig. 2a-b). However, most labeled connectomics datasets [11,30] are only a few gigavoxels in size, hundreds of times smaller than the unlabeled volume needed for down-stream biological analysis (Fig. 2c).

With our *two-stream active clustering* as the key component, we build an end-to-end framework with a base model and an annotation workflow. Before active learning, our base model achieves state-of-the-art results on public benchmarks. Besides, our annotation workflow reduces interactive error correction time by 26%, as shown by a controlled user study. With this framework, we finished the dense annotation of synapse objects and mitochondria semantic mask for a (50 m)^3 EM image volume (300 gigavoxels) in the rat visual cortex, called EM-R50, which is over 100 larger than existing datasets. For the evaluation of active learning approaches on this connectomics dataset, our method improves the performance of previous state-of-the-art methods by 3.1% for synapses and 3.8% for mitochondria, respectively, in terms of the accuracy of the region-of-interest (ROI) proposals. We further perform ablation studies to examine the importance of different framework components and hyper-parameters. To demonstrate its broader impact, we also benchmark our method on natural image classification (CIFAR-10), which outperforms previous state-of-the-art methods by over 2% under a limited annotation budget 5% of the total training images.

Contributions. First, we introduce a novel active learning method that incorporates information from an unsupervised model to improve the effectiveness of query suggestions. Second, our method achieves state-of-the-art results for detection and segmentation tasks on connectomics datasets and image classification on CIFAR-10. Third, we release the code and a densely annotated connectomics dataset (100 bigger than current datasets) to facilitate future researches.

2 Related work

Synapse Detection and Mitochondria Segmentation. Synapse detection and mitochondria segmentation are two popular tasks in connectomics. Due to