Cluster-Based Active Learning

Fábio Perez[*]
University of Campinas, Brazil
fabiop@dca.fee.unicamp.br

Rémi Lebret **Karl Aberer**
EPFL, Lausanne, Switzerland
{remi.lebret,karl.aberer}@epfl.ch

Abstract

In this work, we introduce Cluster-Based Active Learning, a novel framework that employs clustering to boost active learning by reducing the number of human interactions required to train deep neural networks. Our experiments show that the proposed framework requires 82% and 87% less human interactions for CIFAR-10 and EuroSAT datasets respectively when compared with the fully-supervised training, while maintaining similar performance on the test set.

1 Introduction

Deep learning techniques represent the state-of-the-art in many machine learning problems in diverse fields such as computer vision, speech recognition, and natural language processing. However, training deep learning models usually require considerably large amounts of labeled data, which can be expensive to obtain.

Supervised learning is the most effective way of training machine learning models, but it requires the annotation of every sample in the training dataset. To reduce the need for annotated labeled data, machine learning models can work under other supervision schemes. Unsupervised learning methods use unlabeled data during training. Semi-supervised learning methods can learn from both labeled and unlabeled data. Weakly-supervised learning methods use data samples that may contain annotation noise.

Active learning is a subclass of semi-supervised learning methods in which an expert iteratively annotates data in order to maximize the model knowledge while minimizing the number of interactions. A commonly used query strategy is the pool-based sampling, in which the active learning framework selects the most informative samples (i.e., samples that should most increase the model knowledge) from a pool of unlabeled data based on some criteria.

We propose Cluster-Based Active Learning, a novel framework that employs clustering to boost active learning by reducing the number of human interactions required to annotate datasets. Our method combines semi-supervised learning and weakly-supervised learning. Instead of annotating single images, experts may annotate clusters that have class consistency (i.e., the vast majority of samples belong to the same class), greatly reducing the required number of human interactions to train a model.

2 Related Work

Despite being used in machine learning for a long time, it was only recently that active learning was applied to deep learning architectures. Most active learning methods for convolutional neural networks (CNNs) are based on uncertainty measurements. However, it is hard to measure uncertainty in deep learning models [7].

[*]The author was an intern at EPFL during the development of this work.

Third workshop on Bayesian Deep Learning (NeurIPS 2018), Montréal, Canada.

Wang et al. [11] apply active learning to CNNs for image classification. To select the most uncertain samples, they use the output of the softmax layer after inputting an unlabeled image to the network. One of the criteria used for selecting samples is the maximum entropy of the softmax layer.

Gal et al. [3] use Monte Carlo dropout, introduced in Gal and Ghahramani [2], to measure Bayesian uncertainty in CNNs for selecting samples for active learning. They show that collecting multiple outputs from networks with dropout during inference results in better uncertainty measures than on a single forward pass. However, this method is slightly slower since it is necessary to perform multiple forward passes instead of a single one.

Sener and Savarese [10] argue that the uncertainty based methods for active learning in CNNs are ineffective since they query a set of images instead of a single image per iteration, causing the selection of correlated images. They propose a geometric approach as an alternative, in which a subset – called core-set – is selected in a way that the trained model has similar performance to when the full dataset is used. Our method is related to this work in a way that it also uses geometric information (clustering), but our goal is not to select the most informative samples for the annotation, but to maximize the number of data points that will be annotated.

Lakshminarayanan et al. [9] use ensembles as a way of measuring uncertainty in CNNs, by using predictions produced by multiple models. Beluch et al. [1] show that ensembles outperform other methods for active learning for image classification. One problem with the method is that it requires training several models instead of one, requiring more computational resources or time.

3 Cluster-Based Active Learning

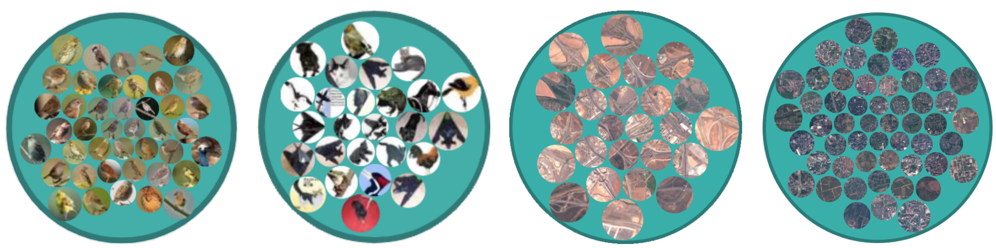

Figure 1: examples of clusters that would be presented to experts for annotation. Left to right: CIFAR-10 cluster that would be labeled as *bird*; CIFAR-10 cluster that would not be annotated; EuroSAT cluster that would be labeled as *highway*; EuroSAT cluster that would not be annotated. This figure is best viewed on screen.

The cluster-based active learning framework consists of adding the clustering and cluster annotation steps into the common pool-based active learning framework. During the annotation step, it can ask the expert to annotate clusters, single samples based on some acquisition criteria (e.g., the most uncertain samples), or both. When asking for both, the order of annotation matters; that is, experts can first annotate the most uncertain samples and then clusters, or vice-versa.

The cluster annotation step is performed by presenting a visualization of samples that belong to a cluster to the expert, who can then decide to annotate the cluster with a label if the majority of samples are from the same class. Otherwise, the expert does not annotate the cluster. Experts may be able to quickly skim through images inside the cluster and decide if they want to annotate it. Since the work and time required for annotating a cluster or an individual image are similar, we consider each of them as one human interaction. Figure 1 shows real examples of clusters that would be presented to experts for annotation.

4 Experiments

We applied cluster-based active learning for image classification with two different datasets: CIFAR-10 (50,000 training images, 10,000 test images, 10 classes) [8], and EuroSAT (27,000 images, 80/20 splits, 10 classes) [6]. We chose CIFAR-10 because it is widely used in image classification research and EuroSAT as it is from a more specific domain and has more real-life applicability. CIFAR-10 classes are very related to the ImageNet dataset (used for pretraining the models), but EuroSAT categories have a very specific domain and are not directly related with ImageNet classes.

We fine-tuned a ResNet-18 [5] network for 5 epochs for CIFAR-10, and 8 epochs for EuroSAT on every iteration. The network was pretrained on ImageNet with weights provided by PyTorch[2]. Images were resized to 224×224 before fed to the network. We used SGD as the optimizer, with learning rate of $1\mathrm{e}{-3}$ for CIFAR-10, and $1\mathrm{e}{-4}$ for EuroSAT, weight decay $5\mathrm{e}{-4}$, and momentum 0.9. Networks were reset to pretrained ImageNet weights for every iteration. We extracted features from the average pooling layer (512 dimensions for ResNet-18) and used then in k-means clustering with 40 iterations (implemented by the Faiss library[3]), which did not impact training time. Experiments were repeated five times for statistical significance.

To simulate experts annotating clusters, we automatically assigned a label to a cluster only if the modal class in the cluster corresponds to at least 80% of the samples in it. Otherwise, the cluster is not annotated.

We experimented with five different scenarios: **a) random:** randomly select samples for annotation; **b) uncertain-only:** use maximum entropy of softmax for selecting images for annotation; **c) cluster-only:** only annotate clusters; **d) uncertain+cluster:** first annotate uncertain samples, then annotate clusters; **e) cluster+uncertain:** first annotate clusters, then annotate uncertain samples.

We consider the annotation of one single image or one cluster as one human interaction. The numbers of clusters were chosen in a way that each cluster had between 50 and 100 samples on average. We used 1000 and 400 clusters per iteration for CIFAR-10 and EuroSAT, respectively, for the cluster-only scenario. Similarly, for the uncertainty-only scenario, we annotated the most 1000 and 400 uncertain samples per iteration, for CIFAR-10 and EuroSAT, respectively. For the uncertain+cluster and cluster+uncertain scenarios, we annotated 500 clusters and 500 samples, and 200 clusters and 200 samples, for CIFAR-10 and EuroSAT, respectively.

5 Results

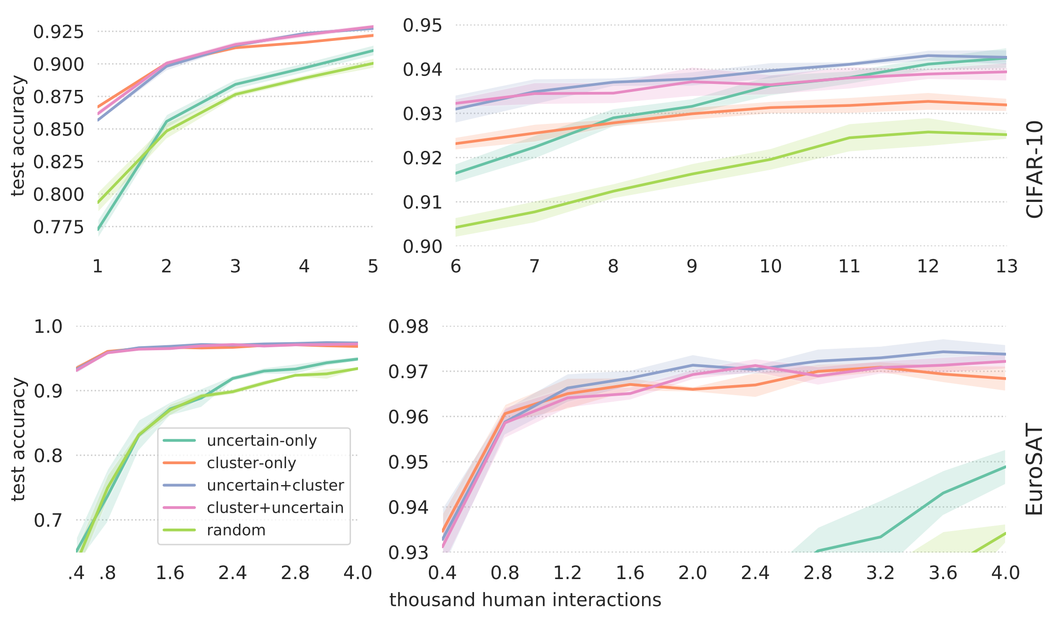

Figure 2: test-accuracy per thousand human interactions. Each view (left and right) shows different limits in both axes for better visualization. Shade areas represent standard deviations.

[2]`https://pytorch.org/`

[3]`https://github.com/facebookresearch/faiss`

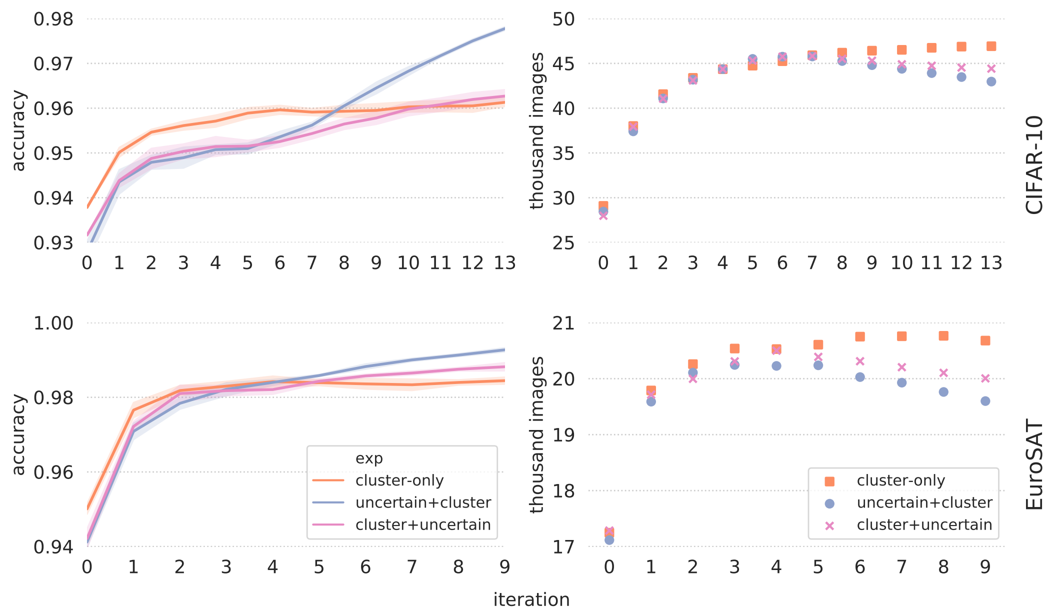

Figure 3: accuracy of samples annotated by clustering (left), and total images annotated by clustering (right). Shade areas represent standard deviations.

Figure 2 shows that the test accuracy for cluster-based scenarios grows much faster than scenarios without clustering. However, the cluster-only scenario stagnates after some iterations, demonstrating the importance of also annotating samples based on uncertainty selection. This is confirmed with scenarios that use both clustering and uncertainty selection.

We can see that uncertain+cluster slightly surpasses cluster+uncertain in terms of test accuracy. Figure 3 shows that the accuracy of cluster-annotated samples gets higher than in other cluster-based scenarios, indicating that uncertainty selection acts as a cluster cleaner.

Figure 3 also indicates that the number of samples annotated by clustering increases every iteration. For scenarios with clustering and uncertainty selection, the number of cluster-annotated samples keeps increasing until it reaches a peak, and then decreases. This happens because the number of images left in the unlabeled pool decreases as they are being annotated during uncertainty selection.

Under the same conditions, test accuracies for supervised training are 0.950 ± 0.003 for CIFAR-10, and 0.973 ± 0.001 for EuroSAT. With cluster-based active learning, CIFAR-10 has similar results from fully-supervised (50,000 images) with 9,000 (18%) human interactions, and EuroSAT with 2,800 (13%) human interactions (against 21,600 in supervised learning).

6 Conclusion

We introduced the cluster-based active learning framework and demonstrated that it can reduce the number of human interactions needed to train a CNN for image classification. Furthermore, the framework can still be improved to achieve better results: training techniques that are more robust to label noise; better feature extraction and clustering methods; better training conditions (e.g., different architectures, data augmentation, better hyperparameters etc.).

6.1 Future Work

We believe that the framework can be applied to other domains, such as natural language processing. However, there are two main difficulties when applying it to other domains: knowledge transfer, which works well in computer vision, but is still developing in other areas; and how to present clusters for annotation, which can be trivial for images, but much harder for other domains such as audio and text. In a future work, we want to apply the framework to the problem of text classification.

References

[1] William H. Beluch, Tim Genewein, Andreas Nürnberger, and Jan M. Köhler. The power of ensembles for active learning in image classification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[2] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 1050–1059. JMLR.org, 2016.

[3] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 1183–1192. PMLR, 2017.

[4] Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, 2017.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90.

[6] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *CoRR*, abs/1709.00029, 2017.

[7] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In Guyon et al. [4], pages 5580–5590.

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

[9] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In Guyon et al. [4], pages 6405–6416.

[10] O. Sener and S. Savarese. Active Learning for Convolutional Neural Networks: A Core-Set Approach. *ArXiv e-prints*, August 2017.

[11] Keze Wang, Dongyu Zhang, Ya Li, Ruimao Zhang, and Liang Lin. Cost-effective active learning for deep image classification. *IEEE Trans. Circuits Syst. Video Techn.*, 27(12):2591–2600, 2017. doi: 10.1109/TCSVT.2016.2589879.