

# Annotation Cost-Sensitive Deep Active Learning with Limited Data (Student Abstract)

Renaud Bernatchez<sup>1,2</sup>, Audrey Durand<sup>4,5,6</sup>, Flavie Lavoie-Cardinal<sup>1,3</sup>

<sup>1</sup>CERVO Brain Research Center, 2601 de la Canardière, Québec (QC), Canada, G1J 2G3

<sup>2</sup>Département de biochimie, de microbiologie et de bio-informatique, <sup>3</sup>Département de psychiatrie et de neurosciences,

<sup>4</sup>Département d'informatique et de génie logiciel, <sup>5</sup>Département de génie électrique et de génie informatique, Université Laval, Québec (QC), Canada, <sup>6</sup>Canada CIFAR AI Chair, Mila, Canada

renaud.bernatchez.1@ulaval.ca

## Abstract

Deep learning is a promising avenue to automate tedious analysis tasks in biomedical imaging. However, its application in such a context is limited by the large amount of labeled data required to train deep learning models. While active learning may be used to reduce the amount of labeling data, many approaches do not consider the cost of annotating, which is often significant in a biomedical imaging setting. In this work we show how annotation cost can be considered and learned during active learning on a classification task on the MNIST dataset.

## Introduction

The training of supervised deep learning models requires vast amount of labeled training data (Krizhevsky, Sutskever, and Hinton 2012). This can be very challenging in research fields such as biomedical imaging, where both the acquisition and the labeling of data are costly (Ronneberger, Fischer, and Brox 2015). Active learning (AL) aims to alleviate this issue by reducing the required quantity of data to label by selecting the most informative sample within an unlabeled dataset. However, many AL approaches do not consider the cost of annotating data and assume a uniform labeling cost for all data points (Settles 2011), which may result in selecting samples that would be very costly to acquire in practice. In this work, we tackle the AL problem with non-uniform sample costs. We first investigate the performance of a commonly used AL approach, Monte Carlo (MC) Dropout (Gal, Islam, and Ghahramani 2017), on a synthetic classification task crafted using the MNIST dataset (Deng 2012). We then highlight the possible trade-offs between annotation cost and accuracy and show how an unknown annotation cost can be learned during AL using previously labeled data, allowing to trade-off cost and accuracy on-the-fly.

## Methods

We consider uncertainty-based AL, where examples to label are selected according to the uncertainty of the model on its classifications, as a proxy for the information gained. The

common *query by committee* strategy estimates the uncertainty of a model using an ensemble of such models. This strategy can be implemented by using MC Dropout to generate different outputs (Gal, Islam, and Ghahramani 2017) followed by an acquisition function to aggregate the prediction probabilities of those outputs. Similar to Gal, Islam, and Ghahramani (2017), we consider the following acquisition functions: Bayesian Active Learning by Disagreement (BALD), Entropy, and Standard Deviation (STD). In order to emulate the context of biomedical imaging where large images are acquired and we must decide which parts an expert should annotate, we consider the following task using the MNIST dataset. On each AL iteration  $t = 1 \dots T$ , we present the AL agent with a set  $\mathcal{S}_t$  of  $|\mathcal{S}_t| = N$  images, from which the agent must select  $n < N$  images to label. Images from previous iterations are never considered again, i.e.  $\mathcal{S}_t \cap \mathcal{S}_{t'} = \emptyset$  for  $t \neq t'$ . This is similar to a biomedical image acquisition setting in which an image is acquired at every iteration, divided into  $N$  crops, and the AL agent must decide which of the  $n$  crops to label. We adapt the convolutional neural network (CNN) from Gal, Islam, and Ghahramani (2017) as the predictive model for this task. We simulate the annotation cost of an image using two components: i) a base acquisition cost that depends on the MNIST class, and ii) an annotation cost calculated from the perimeter of the binarized digit within each image. The uncertainty  $I$  and the cost  $C$  are combined using a ratio inspired by Mackowiak et al. (2018):

$$v = I^\alpha / C^\beta, \quad (1)$$

where  $\alpha$  and  $\beta$  control the relative importance of uncertainty and cost. Note that for  $\beta = 0$ , this falls back to the typical AL problem where the acquisition cost is not considered. The selected examples in our cost-sensitive AL are those giving the most information per unit of cost. We tackle two cost-sensitive AL settings: i) with known annotation costs and ii) with learned annotation costs. In the latter, a CNN is used to learn annotation costs at every iteration, using the cost of previously labeled images as training data.

## Experiments and Results

An initial training dataset of 20 images (2 of each class) is randomly generated, emulating the way an expert would

	Repeat					Mean
	1	2	3	4	5	
BALD	0.67	0.75	1.04	0.59	0.62	0.73 ± 0.16
Entropy	0.71	0.85	0.87	0.65	0.65	0.75 ± 0.09
STD	<b>0.57</b>	<b>0.53</b>	<b>0.79</b>	<b>0.53</b>	<b>0.59</b>	<b>0.60 ± 0.09</b>

Table 1: Proportion of required images compared to random selection to reach a classification accuracy of 90% with the different AL acquisition functions.

likely generate the initial dataset (Gal, Islam, and Ghahramani 2017). Then, on each iteration  $t = 1 \dots 48$ , the AL agent is presented with  $N = 1000$  images from which they must select  $n = 10$  images to label. This experiment is repeated 5 times for each considered approach and the results are averaged. As a baseline, we first evaluate the different acquisition functions without considering cost on the classification task. Results show that AL allows to reach a higher accuracy than random selection for the same amount of training data. We find that STD performs best, consistently requiring the least training images to reach an accuracy of 90% (Table 1). For the following, we will therefore use the STD acquisition function.

We next conduct experiments which consider the annotation cost in the image selection phase. At first, the annotation costs are assumed to be known prior to labeling. Annotation cost and uncertainty are combined using Eq. 1, where  $\beta \in [0, 0.5]$  and  $\alpha = 1$ , to highlight the possible trade-offs between cost and model performance. In these experiments,  $T = 98$  to better observe the behavior of cost-sensitive AL over more iterations. Importantly, a base acquisition cost was assigned to each class in order to make it more costly to select more informative images, leading to a situation where the trade-off between cost and accuracy is more difficult. Figure 1 shows possible trade-offs between cost and accuracy depending on  $\beta$ . While the accuracy decreases with a high  $\beta$  values, it is possible to reach an accuracy equal or superior to that of random selection with a considerably lower cost. For  $\beta < 0.2$ , the total annotation cost is higher than that of random selection for the same number of training images. As the number of training images increases, the trade-off between cost and accuracy becomes less important. This is expected as performance plateaus when enough training data is used.

In a last experiment, we consider the setting where the labeling cost is not known prior to labeling. A CNN is therefore used for learning the annotation cost of unlabeled data using the known cost of previously acquired (labeled) data. We consider  $\beta = 0.3$  since this value allows to reach a lower annotation cost than random selection for a minimal sacrifice in accuracy. We observe that annotation costs can be learned simultaneously with the classification task while reducing the annotation cost compared to random and uniform cost selection (Figure 1). However, the predicted cost is often underestimated, leading to a slightly lower cost efficiency for the same  $\beta$  value. When data acquisition and annotation costs are unknown, this estimation still allows a better performance than assuming a uniform cost ( $\beta = 0$ ),

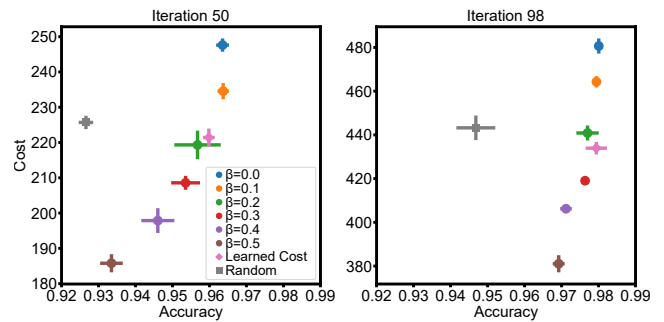


Figure 1: Trade-off between annotation cost and classification accuracy halfway through the AL process (iteration 50) and at the end of the process (iteration 98) for different values of  $\beta$  and random selection. For learned cost,  $\beta = 0.3$ . Error bars are the standard deviation of 3 repetitions.

and is therefore a viable method to consider annotation cost in AL.

## Conclusion

In this work, we show how annotation cost can be considered by AL agents and how it can be learned during the AL process. While these preliminary results are limited to the synthetic task crafted from the MNIST dataset, we plan on deploying these methods on a real-world problem in biophotonics, where the complexity of the data, its limited availability, and the difficulty of labeling images are considerable hurdles to deep learning applications. While many AL methods have been proposed in the literature, they are often evaluated without considering the limitations of a real image acquisition cost and annotation process. Ultimately, we hope that such AL approaches will increase the accessibility of deep learning-based analysis to generate adequate datasets in application fields where annotation and acquisition costs are not negligible.

## References

- Deng, L. 2012. The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6): 141–142.
- Gal, Y.; Islam, R.; and Ghahramani, Z. 2017. Deep Bayesian Active Learning with Image Data. In *Proceedings of the 34th ICML*, volume 70 of *Proceedings of MLR*, 1183–1192.
- Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.
- Mackowiak, R.; Lenz, P.; Ghori, O.; Diego, F.; Lange, O.; and Rother, C. 2018. CEREALS - Cost-Effective REgion-based Active Learning for Semantic Segmentation. In *Proc. Brit. Mach. Vis. Conf.*, 1–21.
- Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *MICCAI*, 234–241.
- Settles, B. 2011. From Theories to Queries: Active Learning in Practice. In *Active Learning and Experimental Design workshop*, 1–18. JMLR Proceedings.