Active Deep Learning for Medical Imaging Segmentation

Marc Gorriz
Xavier Giro-i-Nieto
Universitat Politecnica de Catalunya
Barcelona, Catalonia/Spain
xavier.giro@upc.edu

Axel Carlier
Emmanuel Faure
University of Toulouse
Toulouse, France
axel.carlier@enseeiht.fr

Abstract

We propose a novel Active Learning framework capable to train effectively a convolutional neural network for semantic segmentation of medical imaging, with a limited amount of training labeled data. Our contribution is a practical Cost-Effective Active Learning approach using Dropout at test time as Monte Carlo sampling to model the pixel-wise uncertainty and to analyze the image information to improve the training performance.

1 Motivation

The emergence of deep learning paradigm working through neural networks followed by the recent advances in computational power have enabled the development of new intelligent diagnostics based on computer vision. These diagnostics are capable to analyze images, performing accurate segmentations, in order to detect the lesion areas and to make final decisions about the patient's health as the best of clinical eyes. Nevertheless, the use of deep convolutional neural networks (DCNN) may be a handicap due to the human and logistic costs to get large labeled datasets to train the models avoiding over-fitting. Active Learning (AL) is an established way to reduce this labeling workload in order to select in an iterative way, the most informative examples from a subset of unlabeled instances, to be manual labeled. It has been shown that using this methodology can train deep networks faster and with fewer training samples than traditional semi-supervised learning methods.

2 Proposed methodology

Different from common Active Learning approaches, a Cost-Effective methodology [4] proposes to automatically select and pseudo-annotate good unlabeled samples using the output network predictions. Feeding all the unlabeled samples into the CNN, there are selected two kinds of samples: the minority with low prediction confidence, called most informative/uncertain samples, to be manually labeled by the user (oracle); and the majority with high prediction confidence, to be automatically labeled with no human cost. As one can see that, these two kinds of samples are complementary to each other for representing different confidence levels of the current model on the unlabeled dataset. This procedure is often known as complementary sample selection.

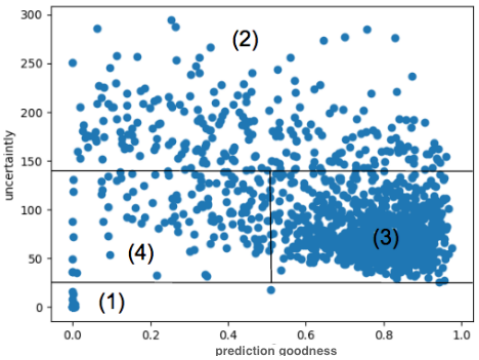
The uncertainty for the model to predict a given input image could be a good criterion to rank all the unlabeled data. In [2] it is explored the possibility to use *Monte Carlo Dropout* to approximate the network weights distribution and hence study their behavior for specific inputs. It works by randomly deactivating network activations during the training in order to prevent overfitting [1], but being applied on test time will allow us to compute the pixel-wise sample uncertainty. Being I_x a image pixel, we can estimate the uncertainty of its predicted label $\widetilde{I_y}$ computing the variance

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

of T different predictions on the same pixel by the effect of *Dropout* through the network weights. The precision of pixel-wise uncertainty maps will depend on the number of step predictions and the *Dropout* probability p_d. Hight p_d means hight variation of network weights making difficult a consistent result with a finite number of step predictions. As was shown in [2] the ideal p_d value is 0.5 and a maximum precision will be obtained when $T \to \infty$. In order to integrate the method to the CEAL complementary sample selection, we need to transform the pixel-wise uncertainty to a numerical score in order to make a based-ranking. The easiest way is achieved by adding all the pixel values from the uncertainty map, getting highest scores at the most doubtful samples. Depending on the nature of the database could be needed a normalization to avoid a size-correlation.

Once all the overall uncertain values are computed, we can visualize the correlation between the predictions goodness and their overall uncertain, by using the database ground truth. In Figure 1 we can distinguish four regions according to the sample nature: (1) No-detected samples, although they have low uncertainty are the most informative candidates to be manually annotated first. (2) Hight uncertain samples, other possible candidates to be annotated by the oracle. (3) Certain samples, the most common ones, perfect candidates to be selected as a pseudo-labels. (4) Random amount of samples that will difficult the complementary sample selection, located at the central region.

Nevertheless, in the real world an hypothetic ground truth for the unlabeled samples can not be available to generate that representation and a histogram could be used to obtain a lossy approximation by projecting all the information to the uncertainty axis, despite being a poor criterion for discerning between overlapped samples in the third and fourth regions. It can be demonstrated that selecting randomly K samples from all the regions to be manually annotated at the initial CEAL iterations, could be a good practice to reduce the concentration of these bad instances that could interfere at the pseudo-labeling assignation procedure. With all the above, the complementary data selection approach will plan a strategy to select sequentially the correct proportion of samples for each region in each iteration, to improve gradually the network performance, without falling into over-fitting.



Sample selection per iteration	
User annotations	Samples
No-detections	10
Most uncertain	10
Random	15
Pseudo annotations	Progressive

Figure 1: Left: Regions uncertainty representation. Right: CEAL approach.

3 Results and Future work

The dataset used is publicly available and contains 2000 RGB dermoscopy images manually annotated by medical experts, by manual tracing the lesion boundaries in the form of a binary mask (International Skin Imaging Collaboration). Before starting the interactive learning process, the training sets are initialized, randomly selecting 200 samples for the test 600 samples for the labeled amount D^L to initialize the network weights and deleting the other labels to initialize the unlabeled set D^U. We trained the U-Net [3]: a convolutional neural network architecture specifically designed to solve Biomedical Image Segmentation problems, using the CEAL approach defined in Figure 1, obtaining a Dice Coefficient of 74 % after 9 active iterations with 2 training epochs per time.

The results are satisfactory but not as good as desired, still being samples in the central region that interfere with the system improvement. Nonetheless, it remains an open door for future works to continue researching new and more adapted solutions to achieve better starting points for the pseudo-labels, with the desire to see their potential power in this field.

References

[1] Y. Gal and Z. Ghahramani. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. 2015.

[2] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. volume abs/1511.02680, 2015.

[3] P. Fischer O. Ronneberger and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[4] Keze Wang, Dongyu Zhang, Ya Li, Ruimao Zhang, and Liang Lin. Cost-effective active learning for deep image classification. volume abs/1701.03551, 2017.