Cost-Effective Active Learning for Melanoma 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

One of the major problems in medical diagnosis is the subjectivity of the specialist's decisions. More concretely, in the fields of medical imaging interpretation, the experience of the specialist can greatly determine the outcome of the final diagnosis. Manual methods of visualization can sometimes be very tedious, time-consuming and subject to errors on part of the interpreter. This has led the growing of intelligent image-based diagnostics as a support, being one of the most current research topics nowadays. 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, deep convolutional neural networks (CNNs) contain large amounts of trainable parameters and require huge amounts of useful and labeled data for the training to converge while avoiding over-fitting. This may be a heavy handicap in the medical imaging field, where the annotation costs of highly qualified professionals could make unfeasible the generation of large labeled datasets. Active Learning (AL) is an established way to reduce this labeling workload in order to select, in an iterative way, a subset of informative examples from a large collections of unlabeled images. The chosen candidates are labeled and subsequently added to the training set. There exist multiple methods choose which samples are to be annotated by human experts. One approach is based on the uncertainty of the prediction. In the case of image segmentation, the solution proposed in this paper estimates this uncertainty based on the stability of the pixel-wise predictions when a drop out is applied on a deep neural network.

The main contributions of this work are the following:

- The design and training of a framework for medical imaging semantic segmentation using Convolutional Neuronal Networks (CNN) and the Cost-Effective Active Learning method.
- The development of image information interpreters for medical imaging based on *Monte Carlo Dropout* for the analysis of the intrinsic network distribution.

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

2 Related work

2.1 Cost-Effective Active Learning (CEAL) algorithm

An active learning is an algorithm able to interactively query the human annotator (or some other information source) new labeled instances from a pool of unlabeled data. Candidates to be labeled can be chosen with several methods based on informativeness and uncertainty of the data. Opposite to common Active Learning approaches that only consider the most informative and representative samples, a Cost-Effective methodology [11] proposes to automatically select and pseudo-annotate unlabeled samples. A CEAL progressively feeds the samples from the unlabeled dataset into the CNN, and selects two kinds of samples for fine-tuning according to the CNN's output. The sampling selection procedure is also known as complementary sample selection. One type of samples are a minority samples with low prediction confidence, which correspond to the most informative/uncertain samples whose prediction score is low. The selected samples are added into the labeled set after an active user (oracle) labels them. Another type of samples are the majority samples with high prediction score, because their prediction is highly confident. For these certain kind of samples, the proposed CEAL automatically assigns pseudo-labels with no human cost. These two types of samples are complementary to each other for representing different confidence levels of the current model on the unlabeled dataset.

2.2 CNN's for Image Segmentation: U-Net architecture

Our active learning approach is applied over a popular convolutional neural network (CNN) in the medical domain: the U-Net. The U-Net [10] is a convolutional neural network architecture designed to solve biomedical image segmentation tasks. It was successfully used for winning the ISBI cell tracking challenge [1] in 2015. The network combines a convolutional network architecture with a deconvolutional architecture to obtain the semantic segmentation [9][8]. The convolutional network is composed of a repetitive pattern of two 3 x 3 convolutions operations, followed by a ReLU layer and a downsampling process through a 2 x 2 maxpooling operation with stride 2. On the other hand, the deconvolutional architecture includes a upsampling operation of the feature map obtained during the contracting path, followed by a 2 x 2 deconvolution that fractions the feature map channels into 2. A posteriori concatenation of the resulting feature map and the obtained during the contracting path is needed, followed by a 3 x 3 convolutions and a ReLU layer. The entire network is 23 convolutional layers deep, where the last layer is used to map each component feature vector related to the number of classes.

3 Proposed methodology

3.1 Image uncertainty estimation

The active learning criteria used to perform the complementary sample selection is based on the intrinsic distribution of the unlabeled data, ranking the unlabeled pool D^U based on their influence on the model. Being D_x^U, D_y^U the unlabeled data and their labels respectively, we are interested in finding the posterior network distribution $p(W|D_x^U, D_y^U)$. In general, this posterior distribution is not tractable, therefore we need to approximate the distribution of these weights using variational inference [4]. This technique allows us to learn the distribution over the network's weights, $q(W)$, by minimizing the *Kullback-Leibler (KL)* divergence between this approximating distribution and the full posterior one.

In [7] it is explored the possibility to use *Monte Carlo Dropout* to approximate $q(W)$ term. The dropout works by randomly deactivating network activations following a *Bernoulli* distribution with a basis probability p_d per layer. With all the above, the introduction of *Dropout* during the training prevents overfitting [3], while at test time it allows introducing 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 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 *Dropout* steps T and the *Dropout* probability p_d. High p_d means high variation of network weights, making difficult a consistent result with a finite number of step predictions. As was shown in [7] an optimal p_d value is 0.5 and a maximum precision would be obtained when $T \to \infty$.

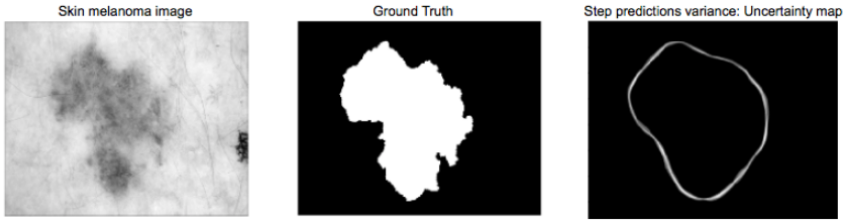


Figure 1: Pixel-wise uncertainty map using 10 step predictions.

In order to integrate the method to the CEAL complementary sample selection, the pixel-wise uncertainty must be transformed into a numerical score to estimate the prediction confidence. We propose summing up all the pixel values from the uncertainty map, obtaining this way higher scores for those most doubtful segmentations.

However, a simple pixel-wise addition does not cope with the variability of predictions across the contours. The further from the contour, the greater should be the contribution to the overall uncertainty of the prediction. For this reason, we introduce an additional weighting step based on the distance map over the predicted segmentation, which corresponds to computing the euclidean distance of each pixel to the closest pixel of the contour [6]. This distance map is multiplied with the uncertainty map, so that the values of the further pixels from the predicted contour are boosted. Intuitively, this processing maker thicker the thickest contours in the uncertainty maps, and thinner the thinnest ones.

3.2 Complementary sample selection

Once the uncertainty score is defined, we can visualize its relation with the accuracy of the predictions. Figure 2 (left) shows the correlation between both, using as predictor the best model found in the next Section 4. The plot allows distinguishing four types of samples:

1. Undetected melanomas: although they have low uncertainty, they are also the most informative candidates to be manually annotated because they correspond to missing detections but with high certainty.
2. Highly uncertain samples: secondary candidates to be annotated by the oracle.
3. Certain samples: the most common ones, best candidates to be selected as a pseudo-labels.
4. Uncertain and wrong predictions: the most undesirable case whose population aims at be reduce by iterating the active learning algorithm.

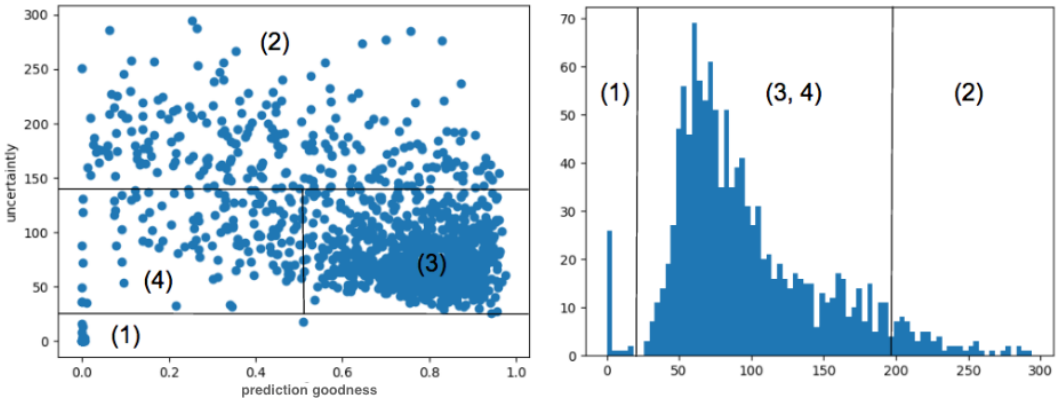


Figure 2: Right: Regions uncertainty representation. Left: histogram projection.

Notice that in the real case scenario, the accuracy of the predictions cannot be computed because the ground truth of the analyzed samples is not available. In a real case only the uncertainty values can be estimated. A histogram of these uncertainty values is depitced in 2 (right), which corresponds to

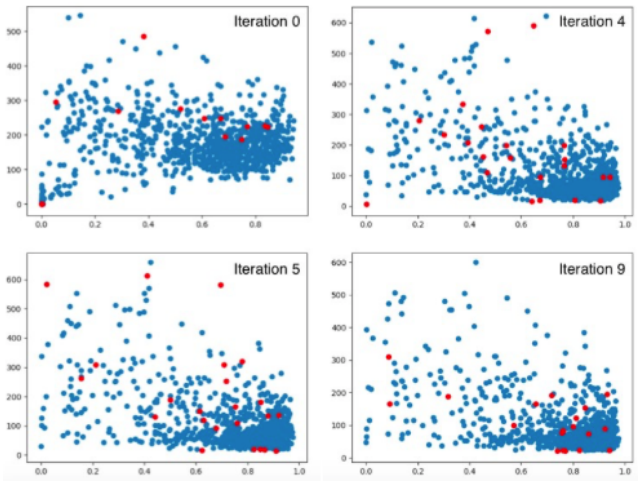
a projection of 2 (left) over the uncertainty axis. In this case, samples between the third and fourth regions become shuffled and cannot be discriminated. With all the above, the complementary data selection method will plan a strategy to select sequentially the correct proportion of samples for each region in each iteration, to improve gradually the network performance, without falling into over-fitting.

4 Results and Future work

All the tests were done with the ISIC 2017 Challenge dataset for Skin Lesion Analysis towards melanoma detection [5], splitting the training set into labeled and unlabeled amount of data to simulate the Active Learning problem with large amounts of unlabeled data at the beginning.

The dataset contains $2,000$ 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). The dataset was modified for this work, transforming the original images to gray scale and modifying their aspect ratio to adapt to the CNN input requirements. Before starting the interactive learning process, the training sets are initialized based on the Cost-Effective Active Learning methodology. First, it is randomly selected the labeled amount D^L from the dataset and then the other labels are deleted initializing the unlabeled set D^U.

One powerful solution at the beginning is the data augmentation [2], a widely used technique in matching learning to generate more training instances, performing basic transformations with the initial dataset. This methods help the initial training to achieve highest accuracy in less training epochs, in order to start the complementary sample selection with the best possible generalization ability, preventing further over-fitting issues. Based on regions representation we modeled an optimum parametrization for the complementary data selection defined in Figure 3, it is also shown the regions evolution during training process, being the certain samples better distributed to be chosen to be pseudo-annotated without interferences and hence the amount chosen per iteration could be progressively increased in order to take more advantage of them in the last iterations.



Initial database balancing	
Initial Set	Samples
Labeled Set	600
Unlabeled Set	1000
Test Set	400

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

Figure 3: Left: Regions evolution. Red samples showing the manual annotations. Right: Cost-Effective Active Learning approach and parametrization.

The results reached a Dice Coefficient [10] of 74 % after 9 interactive iterations with 2 training epochs per run. However, results indicate that after these 9 iterations there are still samples in the central region (4) 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] Ieee international symposium on biomedical imaging. cell tracking challenge. [online]. 2015. `http://biomedicalimaging.org/2015/hardi-reconstruction-challenge/`.

[2] I. Sutskever A. Krizhevsky and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *In Advances in neural information processing systems*, pages 1097–1105, 2012.

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

[4] A. Graves. Practical variational inference for neural networks. In *In Advances in Neural Information Processing Systems*, pages 2348–2356, 2011.

[5] David Gutman, Noel CF Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1605.01397*, 2016.

[6] D. Kirkpatrick M. Werman H. Breu, J. Gil. Linear time euclidean distance transform algorithms. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 17, pages 529–533, 1995.

[7] 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.

[8] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *CoRR*, abs/1210.5644, 2012.

[9] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. *CoRR*, abs/1505.04366, 2015.

[10] 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.

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