Monte-Carlo Sampling applied to Multiple Instance Learning for Whole Slide Image Classification

Marc Combalia
Signal Theory and Communications
Universitat Politècnica de Catalunya
Barcelona, Spain
marc.combalia@alu-etsetb.upc.edu

Verónica Vilaplana
Signal Theory and Communications
Universitat Politècnica de Catalunya
Barcelona, Spain
veronica.vilaplana@upc.edu

Abstract

In this paper we propose a patch sampling strategy based on sequential Monte-Carlo methods for Whole Slide Image classification in the context of Multiple Instance Learning and show its capability to achieve high generalization performance on the differentiation between sun exposed and not sun exposed pieces of skin tissue.

1 Introduction

Deep learning has been widely used in the context of image classification with great success. However, neural networks can not be directly applied to very high resolution images, such as Whole Slide Tissue images, due to the high computational cost involved. A common solution consists in dividing the image into patches and using patch-level annotations to train a supervised classifier. However, these patch-level annotations are not usually available, specially when working with medical datasets. On the contrary, image-level annotations are much easier to obtain so practitioners have used Multiple Instance Learning to train patch-level classifiers in a weakly supervised manner using them [5] [2] [4] [3].

Nevertheless, not much attention has been paid to how the image is transformed into a collection of patches. Patches are usually sampled using a regular grid (with or without overlapping) [2] and fed directly to the MIL neural network. In this work we propose a novel patch sampling strategy which concentrates the effort on the most discriminative regions of the image for a given instant in the training process, permitting faster convergence and higher generalization performance. We compare this approach to the conventional grid sampling and uniform sampling techniques.

2 Method

2.1 Multiple Instance Learning

Multiple Instance Learning (MIL) is a kind of weakly supervised learning algorithm which uses as input a set of N bags, where each bag n contains M_n instances. Considering the standard Multiple Instance assumption (SMI), a bag is negative if all its instances are negative. On the other hand, a bag is positive, if at least one instance in the bag is positive [6]. In the case of high resolution image classification, an image (bag) n is divided into M_n patches (instances), and these patches are processed as pertaining to the whole image. If an image is positive (label of image $Y_n = 1$), it will contain at least one positive patch (label of patch $y_{nm} = 1$ at least for one m). On the contrary, if the image is negative (label of image $Y_n = 0$), all its patches will be negative (label of patch $y_{nm} = 0$ for all m). Then, the most discriminative instance in a bag will be the one with the higher output: $Y_n = max_m(y_{nm})$ [5].

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

This formulation can be implemented in the context of neural networks (for the SMI assumption) as a Max Pooling layer. Other aggregating functions have been proposed in the literature [5] [3], but we will use the max aggregating function to illustrate the variations in accuracy between the proposed strategies.

2.2 Patch sampling

Patches need to be sampled from the images prior to training the classifier. In this section we discuss various patch sampling strategies.

Grid sampling: The extraction of patches is performed in a grid-like manner; the image is divided into a regular grid of patches, with or without overlapping.

Uniform sampling: Patches are extracted from the images using a uniform distribution to select their centroids at every epoch. This approach is similar to the grid sampling strategy in the sense that it will sample all the image uniformly, but this time the sampling is performed in a stochastic manner at every batch.

Monte-Carlo sampling: We use a Monte Carlo method to estimate the output probability map, and use this estimation to extract the patch with higher output probability, which will be the most informative for the training process. When a new image is fed into the network, we perform an estimation of the output probability map using a variation of a sequential Monte-Carlo method:

1. Initialization: We initialize n centroids using a uniform distribution.
2. Evaluation: a patch is sampled from every centroid and forwarded though the network. The output produced by the patch is used to represent the centroid.
3. Normalization: we expand the range of all the centroids' representation values to cover values from 0 to 1. The centroids whose value is closer to 1 will be the ones which have obtained the highest output from the neural network.
4. Re-sampling: We stochastically eliminate the centroids whose value is closer to 0 using a random uniform distribution and relocate them on top of the highest scoring centroids.
5. Displacement: we displace the new centroids using a random 2D Gaussian distribution.
6. Back to step 2 for k iterations

The proposed method relocates centroids which have not been relevant for classification into more discriminative regions in the image, that is, around the centroids with higher activations.

After estimating the output probability, we can use the patches corresponding to the highest scoring centroids to train the neural network. It is important to perform this process at every batch, since the discriminative regions in the WSI will vary as the network learns during the training process.

3 Experiments and results

The results have been obtained on a problem of classification between sun-exposed and not sun-exposed pieces of skin (8000 / 2000 Whole Slide images for train / validation) extracted from the GTEx database [1].

We train three identical networks using the proposed sampling strategies. The neural network is a shallow ResNet (8 layers) with an input size of 50 x 50 pixels trained with Adam optimization (learning rate $1e-3$). We use $n = 500$ centroids in both the Monte-Carlo sampling scheme and in the uniform sampling scheme, and $k = 1$ iterations in the Monte-Carlo sampling scheme. We evaluate the neural networks using grid-sampling to divide the image into patches and aggregate patch scores using the max operator to obtain image-wise predictions. The results for image-wise accuracy are shown in table 1.

When using the grid sampling strategy, the neural network over-fits on the training split. The neural network sees exactly the same patches at every epoch, failing to generalize on the validation split. The uniform sampling method overcomes this issue as the neural network will stochastically see all the possible patches from the image. However, both methods waste a lot of sampling power in regions of the image which are not relevant for classification. Instead, the Monte-Carlo sampling strategy

Table 1: Training and validation accuracies for the various sampling strategies

Sampling technique	Train acc	Validation acc
Grid sampling	**0.967**	0.826
Uniform sampling	0.929	0.920
Monte-Carlo sampling	0.946	**0.942**

is able to concentrate the sampling effort on the discriminative regions at every step of the training phase obtaining a higher validation accuracy compared with the other two sampling strategies.

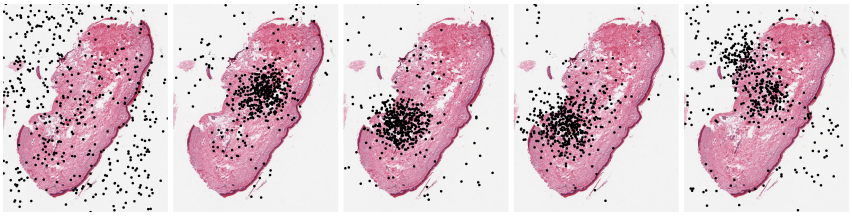

Figure 1: Black points corresponds to the centroids' positions at epochs 2, 6, 8, 37 and 72 in the training process of the neural network.

On figure 1, we can see how the proposed sampling strategy adapts to the neural network. At early stages, when the network hasn't learned the distribution of the data, the centroids follow a uniform distribution. However, as the classifier keeps training, the sampling strategy proceeds to extract patches from the most discriminative regions at every batch.

4 Conclusions

We have proposed a sampling strategy based in sequential Monte-Carlo methods which makes the neural network focus on the most discriminative regions in the image during the training, obtaining higher generalization performance on the problem of sun exposure classification on Whole Slide Tissue images.We expect this sampling strategy to work best on datasets which have very localized information. We plan to test it on other artificial and histological datasets.

Acknowledgments

This work has been partially supported by the project MALEGRA TEC2016-75976-R financed by the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund.

References

[1] L. J. Carithers, K. Ardlie, M. Barcus, P. A. Branton, A. Britton, S. A. Buia, C. C. Compton, D. S. DeLuca, J. Peter-Demchok, E. T. Gelfand, et al. A novel approach to high-quality postmortem tissue procurement: the gtex project. *Biopreservation and biobanking*, 13(5):311–319, 2015.

[2] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2424–2433, 2016.

[3] M. Ilse, J. M. Tomczak, and M. Welling. Attention-based deep multiple instance learning. *CoRR*, abs/1802.04712, 2018.

[4] C. Mercan, E. Mercan, S. Aksoy, L. G. Shapiro, D. L. Weaver, and J. G. Elmore. Multi-instance multi-label learning for whole slide breast histopathology. In *Medical Imaging 2016: Digital Pathology*, volume 9791, page 979108. International Society for Optics and Photonics, 2016.

[5] J. M. Tomczak, M. Ilse, and M. Welling. Deep Learning with Permutation-invariant Operator for Multi-instance Histopathology Classification. *ArXiv e-prints*, Dec. 2017.

[6] X. Wang, Y. Yan, P. Tang, X. Bai, and W. Liu. Revisiting Multiple Instance Neural Networks. *ArXiv e-prints*, Oct. 2016.