Deep Learning with Order-invariant Operator for Multi-instance Histopathology Classification

Tomczak, J.M.; Ilse, M.; Welling, M.

DOI
10.48550/arXiv.1712.00310

Publication date
2017

Document Version
Author accepted manuscript

Deep Learning with Permutation-invariant Operator for Multi-instance Histopathology Classification

Jakub M. Tomczak
University of Amsterdam

Maximilian Ilse
University of Amsterdam

Max Welling
University of Amsterdam

Abstract

The computer-aided analysis of medical scans is a longstanding goal in the medical imaging field. Currently, deep learning has became a dominant methodology for supporting pathologists and radiologist. Deep learning algorithms have been successfully applied to digital pathology and radiology, nevertheless, there are still practical issues that prevent these tools to be widely used in practice. The main obstacles are low number of available cases and large size of images (a.k.a. the *small n, large p* problem in machine learning), and a very limited access to annotation at a pixel level that can lead to severe overfitting and large computational requirements. We propose to handle these issues by introducing a framework that processes a medical image as a collection of small patches using a single, shared neural network. The final diagnosis is provided by combining scores of individual patches using a permutation-invariant operator (combination). In machine learning community such approach is called a multi-instance learning (MIL).

1 Introduction

Deep learning has become a leading tool for analyzing medical images, and digital pathology as its major application area [8]. Main practical issues in current deep learning methods for medical imaging are low number of recorded cases, large size of images (slides) and low availability of a diagnosis with a pixel level annotation (a.k.a. *weakly labeled data*). These problems lead to severe overfitting, impractical computations, *e.g.*, training using images larger than 250×250 pixels requires already a considerably large amount of computational resources, and difficulties in information flow from single label for large images. We propose to handle these issues by introducing a framework that processes a medical image as a collection of small patches using a single, shared neural network. The final diagnosis is provided by combining scores of individual patches. In machine learning community such approach is called a *multi-instance learning* (MIL) [9].

Related work There are different approaches to MIL with various combining operators [1, 6, 10, 11, 14] but these methods were mainly used for already pre-processed data. Recently, there is an increase of interest in applying MIL to medical imaging and, especially, to histopathology. One of first such methods used SVM and Boosting to cluster and classify colon cancer images [13]. Recently, a single neural network with a MIL-pooling layer was used to classify and segment microscopy images with populations of cells [7]. A method that is closely related to our approach utilized a neural network to process small patches in the first stage of training and the Expectation Maximization algorithm to determine latent labels of the patches in the second stage [4]. However, our model is trained end-to-end by backpropagation.

Workshop on "Medical Imaging meets NIPS" at 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

2 Methodology

Problem statement A classical supervised learning problem aims at finding a model that takes an object, $\mathbf{x} \in \mathbb{R}^D$, and predicts a value of a target variable, $y \in \{0, 1\}$. In the multi-instance learning problem, however, there is a bag of objects, $\mathcal{X}_K = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$, that exhibit neither dependency nor ordering among each other. There is also a single label associated with this bag. We assume that K could vary for different bags. We do not have access to individual labels of the objects within the bag, *i.e.*, we assume $y_1, \ldots, y_K$ are unknown, but we know that the label of the bag is 1 if at least one object is 1, *i.e.*, $y = 1 \iff \exists_k : y_k = 1$. This statement is equivalent to the logic OR operator and could be further re-formulated as the maximum operator: $y = \max_k\{y_k\}$. The max-operator is **permutation-invariant** that is an important property since objects within a bag are independent.

Training a bag-level classifier requires a permutation-invariant combination of individual labels y_k that are given by an instance-level (shared) classifier. In this paper, we propose to train a model using the likelihood approach. We take the Bernoulli distribution for the bag label:

$$p(y|\mathcal{X}_K) = \big(\theta(\mathcal{X}_K)\big)^y \, \big(1 - \theta(\mathcal{X}_K)\big)^{1-y}, \tag{1}$$

where $\theta(\mathcal{X}_K) \in [0, 1]$ is the probability of $y = 1$ given the bag of objects $\mathcal{X}_K$. Further, we consider a shared instance-level classifier (a neural network) with parameters ψ, $f_\psi(\mathbf{x}_k)$, that returns a score for the k-th object, $z_k = f_\psi(\mathbf{x}_k)$ and $z_k \in [0, 1]$. Then, the parameter $\theta(\mathcal{X}_K)$ is modeled using a permutation-invariant operator $g : [0, 1]^K \to [0, 1]$, *i.e.*, $\theta(\mathcal{X}_K) = g\big(f_\psi(\mathbf{x}_1), \ldots, f_\psi(\mathbf{x}_K)\big)$.

Permutation-invariant operators Obviously, we can choose the max-operator as g but it is not necessarily well-suited for training neural networks using the backpropagation. Alternatively, we consider the following differentiable operators:

(i) Noisy-Or (NOR) operator [3]:

$$\theta(\mathcal{X}_K) = 1 - \prod_{k=1}^{K} \big(1 - f_\psi(\mathbf{x}_k)\big),$$

(ii) Integrated Segmentation and Recognition (ISR) operator [6]:

$$\theta(\mathcal{X}_K) = \frac{\sum_{k=1}^{K} v_k}{1 + \sum_{k=1}^{K} v_k},$$

where $v_k = \frac{f_\psi(\mathbf{x}_k)}{1 - f_\psi(\mathbf{x}_k)}$,

(iii) Log-sum-exp (LSE) operator [10] with $r > 0$:

$$\theta(\mathcal{X}_K) = \frac{1}{r} \ln \frac{1}{K} \sum_{k=1}^{K} \exp\big(r f_\psi(\mathbf{x}_k)\big).$$

Training Once the operator is chosen, we train the model by minimazing the negative log-likelihood using (1):

$$\mathcal{L}(\psi) = -\frac{1}{N} \sum_{n} \ln p(y_n|\mathcal{X}_{K,n}, \psi). \tag{2}$$

3 Workflow

In our framework the input is a slide or a patch from a needle biopsy stained with Hematoxylin & Eosin (H&E). Further, we divide the input into small patches (*e.g.*, 96×96 pixels). Each small patch is processed by a shared neural network $f_\psi(\mathbf{x}_k)$, which consists of several convolutional layers and fully-connected layers with dropout, and it returns a score of each small patch, z_k. A larger score determines a Region of Interest (ROI) that could be later presented to a human doctor. Eventually, an application of a permutation-invariant operator provides the probability of a diagnosis, *e.g.*, benign or malignant tumor. The proposed framework is presented in Figure 1.

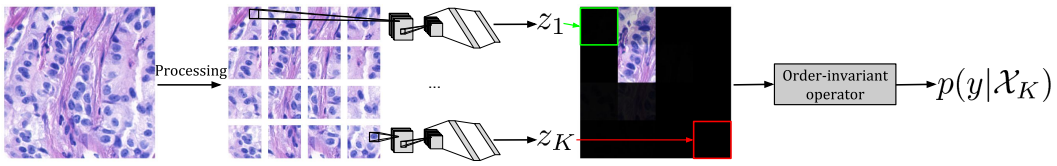

Figure 1: A schematic representation of the proposed workflow.

4 Experiments

Data In the experiments we used a dataset that consists of 58 H&E stained histopathology image excerpts (896×768 pixels) taken from 32 benign and 26 malignant breast cancer patients [2]. Due to a limited size of the dataset, a 4-fold cross-validation is used as in [5]. For images in the training set, we select eight 768×768 overlapping subimages. However, for images in the test set we select a single 768×768 subimage from the center of the image. During training, we use 10% of the training set for validation and monitoring a training progress. Subsequently, each subimage is divided into patches of 96×96 pixels. A patch is discarded if more than 75% of the pixels are white.

Data augmentation In every training iteration we perform data augmentation to prevent overfitting. We randomly adjust the amount of H&E by decomposing the RGB color of the tissue into the H&E color space [12], followed by multiplying the magnitude of H&E for a pixel by two i.i.d. Gaussian random variables with expectation equal to one. We randomly rotate and mirror every patch. Lastly, we blur the patch using a Gaussian blur filter with a randomly chosen blur radius. See Figure 2 for examples of data augmentation transformations.

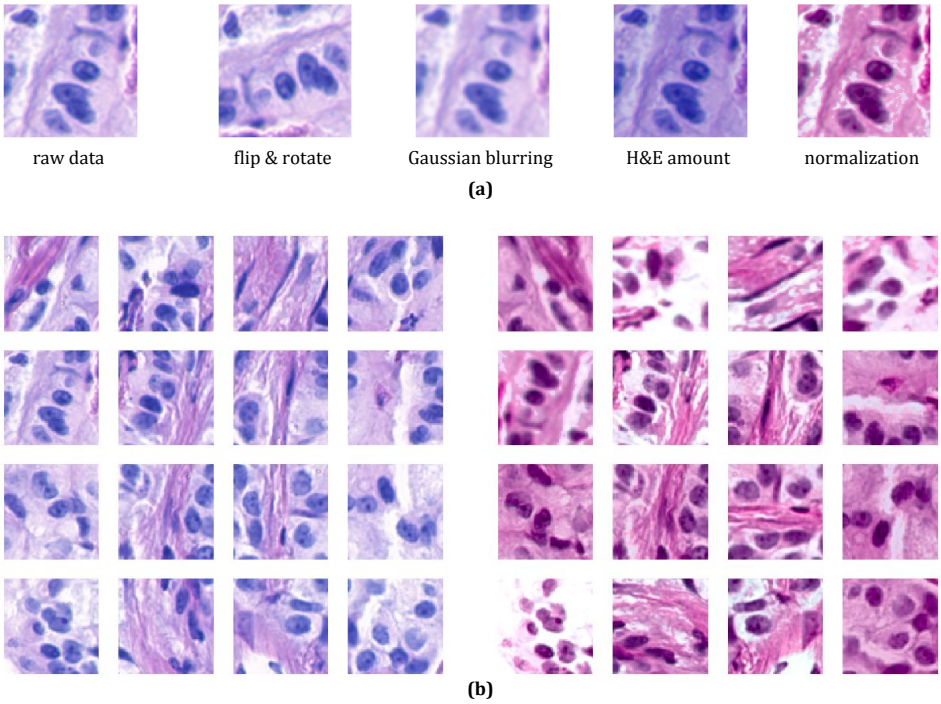

Figure 2: **(a)** Four different data augmentation transformations applied to a raw data. **(b)** An image divided into patches (on the left) processed by data augmentation transformations (on the right).

Results and discussion We compared our approach (DEEP{NOR,ISR,LSE}-MIL) with the Gaussian process multi-instance learning (GPMIL) and its relational extension RGPMIL [5]. Results are given in Table 1.

First, we notice that the proposed approach achieved similar performance to Gaussian process-based methods in terms of AUC. Second, the LSE operator failed to obtain high accuracy and F-score but it still resulted in high AUC. Comparing all operators, we believe that Noisy-or is the most promising but in order to obtain even better results a kind of regularization is required. A possible extension of the presented work would be an application of the Bayesian learning similarly to [11]. However, we leave investigating these issues for further research.

Table 1: Results of the 4-cross-validation on the breast cancer data.

METHOD	ACCURACY	PRECISION	RECALL	F-SCORE	AUC
GPMIL [5]	N/A	N/A	N/A	N/A	0.86
RGPMIL [5]	N/A	N/A	N/A	N/A	**0.90**
DEEPNOR-MIL	**0.879**	0.828	**0.923**	**0.873**	0.88
DEEPISR-MIL	0.828	0.808	0.808	0.808	**0.90**
DEEPLSE-MIL ($r = 10$)	0.621	**0.833**	0.192	0.312	0.88

Acknowledgments

Jakub M. Tomczak was funded by the European Commission within the Marie Skłodowska-Curie Individual Fellowship (Grant No. 702666, "Deep Learning and Bayesian Inference for Medical Imaging"). Maximilian Ilse was funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Grant "DLMedIa: Deep Learning for Medical Image Analysis").

References

[1] G. Chen and S. N. Srihari. A noisy-or discriminative restricted boltzmann machine for recognizing handwriting style development. In *Int. Conf. on Frontiers in Handwriting Recognition*, pages 714–719, 2014.

[2] E. D. Gelasca, J. Byun, B. Obara, and B. Manjunath. Evaluation and benchmark for biological image segmentation. In *IEEE Int. Conf. on Image Processing*, pages 1816–1819, 2008.

[3] Y. Halpern and D. Sontag. Unsupervised learning of noisy-or bayesian networks. *arXiv preprint arXiv:1309.6834*, 2013.

[4] 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 *CVPR*, pages 2424–2433, 2016.

[5] M. Kandemir, C. Zhang, and F. A. Hamprecht. Empowering multiple instance histopathology cancer diagnosis by cell graphs. In *MICCAI*, pages 228–235, 2014.

[6] J. D. Keeler, D. E. Rumelhart, and W. K. Leow. Integrated segmentation and recognition of hand-printed numerals. In *NIPS*, pages 557–563, 1991.

[7] O. Z. Kraus, J. L. Ba, and B. J. Frey. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics*, 32(12):i52–i59, 2016.

[8] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *arXiv preprint arXiv:1702.05747*, 2017.

[9] O. Maron and T. Lozano-Pérez. A framework for multiple-instance learning. In *NIPS*, pages 570–576, 1998.

[10] J. Ramon and L. De Raedt. Multi instance neural networks. In *ICML Workshop on Attribute-value and Relational Learning*, pages 53–60, 2000.

[11] V. C. Raykar, B. Krishnapuram, J. Bi, M. Dundar, and R. B. Rao. Bayesian multiple instance learning: automatic feature selection and inductive transfer. In *ICML*, pages 808–815, 2008.

[12] A. C. Ruifrok and D. A. Johnston. Quantification of histochemical staining by color deconvolution. *Analytical and Quantitative Cytology and Histology*, 23(4):291–299, 2001.

[13] Y. Xu, J.-Y. Zhu, E. Chang, and Z. Tu. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In *CVPR*, pages 964–971, 2012.

[14] C. Zhang, J. C. Platt, and P. A. Viola. Multiple instance boosting for object detection. In *NIPS*, pages 1417–1424, 2006.