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

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
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 become 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.
2 Methodology

Problem statement A classical supervised learning problem aims at finding a model that takes an object, \( x \in \mathbb{R}^D \), 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, \( X_K = \{x_1, \ldots, x_K\} \), 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:

\[
p(y|X_K) = (\theta(X_K))^y (1 - \theta(X_K))^{1-y},
\]

where \( \theta(X_K) \in [0, 1] \) is the probability of \( y = 1 \) given the bag of objects \( X_K \). Further, we consider a shared instance-level classifier (a neural network) with parameters \( \psi \), \( f_\psi(x_k) \), that returns a score for the \( k \)-th object, \( z_k = f_\psi(x_k) \) and \( z_k \in [0, 1] \). Then, the parameter \( \theta(X_K) \) is modeled using a permutation-invariant operator \( g : [0, 1]^K \rightarrow [0, 1] \), i.e., \( \theta(X_K) = g(f_\psi(x_1), \ldots, f_\psi(x_K)) \).

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 \([3]\):

\[
\theta(X_K) = 1 - \prod_{k=1}^{K} (1 - f_\psi(x_k)),
\]

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

\[
\theta(X_K) = \frac{\sum_{k=1}^{K} v_k}{1 + \sum_{k=1}^{K} v_k},
\]

where \( v_k = \frac{f_\psi(x_k)}{1 - f_\psi(x_k)} \).

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

\[
\theta(X_K) = \frac{1}{r} \ln \left( \frac{1}{K} \sum_{k=1}^{K} \exp \left( rf_\psi(x_k) \right) \right). 
\]

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

\[
\mathcal{L}(\psi) = -\frac{1}{N} \sum_n \ln p(y_n|X_{K,n}, \psi). 
\]
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.

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>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPMIL [5]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.86</td>
</tr>
<tr>
<td>RGPMIL [5]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.90</td>
</tr>
<tr>
<td>DeepNOR-MIL</td>
<td>0.879</td>
<td>0.828</td>
<td>0.923</td>
<td>0.873</td>
<td>0.88</td>
</tr>
<tr>
<td>DeepISR-MIL</td>
<td>0.828</td>
<td>0.808</td>
<td>0.808</td>
<td>0.808</td>
<td>0.90</td>
</tr>
<tr>
<td>DeepLSE-MIL (r = 10)</td>
<td>0.621</td>
<td>0.833</td>
<td>0.192</td>
<td>0.312</td>
<td>0.88</td>
</tr>
</tbody>
</table>

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