Computerized Medical Imaging and Graphics 70 (2018) 43–52

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Deep learning nuclei detection: A simple approach can deliver state-of-the-art results

Henning Höfener[a,*], André Homeyer[a], Nick Weiss[a], Jesper Molin[b], Claes F. Lundström[b,c], Horst K. Hahn[a,d]

[a] Fraunhofer MEVIS, Am Fallturm 1, 28359, Bremen, Germany
[b] Sectra AB, Teknikringen 20, 58330, Linköping, Sweden
[c] Center for Medical Image Science and Visualization, Linköping University, 58183, Linköping, Sweden
[d] Jacobs University, Campus Ring 1, 28759, Bremen, Germany

ARTICLE INFO

Article history:
Received 8 February 2018
Received in revised form 13 July 2018
Accepted 23 August 2018

Keywords:
Nuclei detection
Deep learning
PMap
Histology
Image analysis

ABSTRACT

Background: Deep convolutional neural networks have become a widespread tool for the detection of nuclei in histopathology images. Many implementations share a basic approach that includes generation of an intermediate map indicating the presence of a nucleus center, which we refer to as PMap. Nevertheless, these implementations often still differ in several parameters, resulting in different detection qualities.
Methods: We identified several essential parameters and configured the basic PMap approach using combinations of them. We thoroughly evaluated and compared various configurations on multiple datasets with respect to detection quality, efficiency and training effort.
Results: Post-processing of the PMap was found to have the largest impact on detection quality. Also, two different network architectures were identified that improve either detection quality or runtime performance. The best-performing configuration yields f1-measures of 0.816 on H&E stained images of colorectal adenocarcinomas and 0.819 on Ki-67 stained images of breast tumor tissue. On average, it was fully trained in less than 15,000 iterations and processed 4.15 megapixels per second at prediction time.
Conclusions: The basic PMap approach is greatly affected by certain parameters. Our evaluation provides guidance on their impact and best settings. When configured properly, this simple and efficient approach can yield equal detection quality as more complex and time-consuming state-of-the-art approaches.

1. Introduction

The quantification of cell nuclei in histological images is essential for many pathological assessments, including the determination of various biomarkers. Prominent examples in cancer diagnosis are the Ki-67 index or the progesterone and estrogen receptor status. Detecting nuclei also enables the quantification of tumor immune infiltrates, which have been shown to be of strong prognostic importance (Mahmoud et al., 2011), and are commonly assessed in immunotherapy trials (Denkert et al., 2016).

Such assessments are usually performed by visual estimation, which is labor- and time-intensive and can lead to high inter- and intra-observer variability (Andrion et al., 1995). The ongoing digitalization in pathology allows for automated analysis methods to support pathologists at such tasks and to increase the reliability of quantitative assessments.

However, the automatic detection of cell nuclei is challenging. The appearance of nuclei varies considerably with staining and tissue preparation conditions, as well as with different nuclear types and pathologies.

The first attempts to automate nuclei detection date back to the mid-1950s (Meijering, 2012), starting with static rule-based approaches from simple intensity thresholds to using intensity-derived features. Those approaches suffered from not being able to capture the complexity of the input data sufficiently well. The next generation of methods addressed the aforementioned variability by using hand-crafted features and applying machine-learning to build more complex and flexible rule sets (Arteta et al., 2012;

* Corresponding author.
E-mail addresses: henning.hoefener@mevis.fraunhofer.de (H. Höfener), andre.homeyer@mevis.fraunhofer.de (A. Homeyer), nick.weiss@mevis.fraunhofer.de (N. Weiss), Jesper.Molin@sectra.com (J. Molin), claes.lundstrom@liu.se (C.F. Lundström), horst.hahn@mevis.fraunhofer.de (H.K. Hahn).

https://doi.org/10.1016/j.compmedimag.2018.08.010

Kårsnäs et al., 2011; Vink et al., 2013). Recent developments mainly employ convolutional neural networks (CNN) (Jacobs et al., 2017; Janowczyk and Madabhushi, 2016; Sirinukunwattana et al., 2016; Wang et al., 2016; Xie et al., 2016, 2015a,b; Xing et al., 2016), as those tend to yield significantly better results. The first major breakthrough was reported by Cireşan et al. (2013), who were able to detect mitotic nuclei with an f1-measure of 0.782, while the closest competitors achieved 0.718.

Most deep learning–based nuclei detection methods employ CNNs to predict a value for each input image pixel. That value represents the proximity to a nucleus center or the probability of being close to one. Together, the values of all input image pixels constitute a map, which we refer to as PMap. Nuclei positions are afterwards determined by finding local maxima in the PMap. Predicting the PMap can be interpreted as either a classification or a regression problem. The classification problem is to distinguish *nucleus center* and *background* positions and to populate the PMap with the each position's probability to belong to the *nucleus center* class, whereas the regression problem is to map a position to a continuous value, which is dependent on the distance to the nearest nucleus center. This basic PMap approach will be described in more detail in Section 2.1.

The basic PMap approach is controlled by several parameters. Examples are the post-processing of the PMap before finding local maxima, the use of data augmentation or the use of dropout.

1.1. Related work

There are different variants of the basic PMap approach proposed in the literature, using CNN classification or regression, even if that term is not used.

As described above, Cireşan et al. (2013) have used CNN classification for the detection of mitoses. Their approach uses a 12 and a 10 layer deep network and achieves processing speeds between 0.01 and 0.03 megapixels per second at prediction time. CNN classification with the 8 layer deep AlexNet (Krizhevsky, 2010) has been used by Janowczyk and Madabhushi (2016) to detect lymphocytes in breast cancer images. They have achieved an f1-measure of 0.900. The 7 layer deep LeNet (LeCun et al., 1998) classification network has been applied by Wang et al. (2016) to detect nuclei for a subsequent cell subtype classification. For the detection, they have reported an f1-measure of 0.822. Khoshdeli et al. (2017) have used a 5 layer deep CNN classification for the detection of nuclei in Hematoxylin and Eosin (H&E) stained images of various tissue types. They have proposed to preprocess the input images by extracting the Hematoxylin channel using color deconvolution and applying a Laplacian of Gaussian filter. The result is then fed into to the network. An f1-measure of 0.722 has been reported. Jacobs et al. (2017) have used a 14 layer deep regression network to detect nuclei in H&E stained prostate cancer biopsies for a subsequent nucleus type classification. The authors have evaluated transfer learning for the application with limited training data. They have trained on colon images and have fine-tuned their model with the prostate images. They have reported f1-measures between 0.849 and 0.864, depending on the amount of training data for the fine-tuning, as well as a processing speed of 2.2 megapixels per second.

Some approaches leave out the extraction of local maxima from the PMap. Xie et al. (2016) have estimated the nuclei count in an image region by integrating the PMap over that region. They have applied a 9 layer deep network. Xing et al. (2016) have applied a threshold to the PMap and have used the connected regions as initialization for nuclei segmentation. The generation of the PMap has been performed with 0.008 megapixels per second.

In other publications, the basic PMap approach has been used as a baseline algorithm to compare the proposed methods with. Xie et al. (2015a) have mapped each pixel of the input image to a 2D-vector pointing to the nearest nucleus center, using an 8 layer deep network. At prediction time, the positions, where the vectors point to, are accumulated to form a PMap. On Ki-67 stained neuroendocrine tumor (NET) images they have reported an f1-measure of 0.815 and a processing speed of 0.007 megapixels per second. They have compared their method with the basic PMap approach using CNN classification, which has yielded an f1-measure of 0.784. In another publication, the same authors have used a 7 layer deep network to predict a small region of the PMap at once instead of a single pixel value (Xie et al., 2015b). As before, they have accumulated the predictions to generate the PMap. They have evaluated the approach on H&E stained breast tumor images, Ki-67 stained NET images and phase contrast images of HeLa cervical cancer cells. F1-measures of 0.913, 0.906 and 0.957 have been reported, respectively. Processing speed has been 0.01 megapixels per second. A comparison with both CNN classification and regression according to the basic PMap approach has been conducted, but no quantitative measures have been given. Sirinukunwattana et al. (2016) have proposed a similar approach of predicting a region of the PMap. In contrast to (Xie et al., 2015b), the first 6 layers of their network are followed by a parameter estimation layer and a spatially constrained regression layer. They have reported an f1-measure of 0.802 on H&E stained images of colorectal adenocarcinomas and processing speed of 0.02 megapixels per second. They have compared their method with the basic PMap approach using CNN regression, for which an f1-measure of 0.692 has been reported. Xu et al. (2016) have used multiple stacked auto-encoders to learn feature representations of the input images in an unsupervised manner. The features are then fed into a softmax classifier, which classifies each input patch as either nuclear or non-nuclear. The softmax classifier has been trained supervisedly and the authors have reported an f1-measure of 0.845 on H&E stained breast cancer images. They compare their method with the basic PMap approach using CNN classification. There, an f1-measure of 0.820 and processing speed of 0.04 megapixels per second have been reported.

We want to stress here that the reported f1-measures should not be compared directly. Most approaches have been evaluated using different datasets with different nuclear types, varying quality and tissue complexity. Additionally, the hardware used to perform the experiments, especially the usage of GPUs, has a great influence on the processing speed. Although only comparable to a very limited extent, we listed processing speeds if available for completeness. Only few of the approaches above explicitly focused on processing time, although speed is critical when aiming at applying these methods in clinical routine.

In summary, for nuclei detection using deep learning, the basic PMap approach is widely used in the literature. Even if not termed basic PMap approach, numerous publications describe such methods either as the proposed or as alternative solutions for nuclei detection tasks. However, there are some parameters of these methods that differ from case to case. Most of the papers above only present a single configuration of them. There is no systematic evaluation of the influence and importance of the individual parameters.

The main contribution of this work is a systematic listing, evaluation and comparison of these parameters. We assess the impact of the individual parameters with respect to detection quality, efficiency and training effort. By doing so, we give guidance on which parameters to focus on when optimizing nuclei detection with the basic PMap approach. The second contribution is to combine those parameter settings, which perform best in our experiments and to evaluate this configuration. We show that the basic PMap approach delivers state-of-the-art results when parameterized well.