Classification of visibility in multi-stain microscopy images

Jonathan Ganz[1] JONATHAN.GANZ@THI.DE
[1] *Technische Hochschule Ingolstadt, Ingolstadt, Germany*

Christof A. Bertram *Institute of Pathology, University of Veterinary Medicine Vienna, Vienna, Austria*

Robert Klopfleisch *Institute of Veterinary Pathology, Freie Universität Berlin, Berlin, Germany*

Samir Jabari *Institue of Neuropathology, University Hospital Erlangen, Erlangen, Germany*

Katharina Breininger
Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Marc Aubreville[1]

Editors: Under Review for MIDL 2022

Abstract

Annotating mitotic figures (MF) in hematoxylin and eosin (H&E) stained slides is an error-prone task that can lead to low inter-rater concordance. Immunohistochemical staining against phospho-histone H3 (PHH3) can lead to higher concordance but, at the same time, to generally higher mitotic figure counts. By annotating MF in PHH3-stained specimen and transferring them to an H&E- re-stained version of the same slide, the high specificity of PHH3 can be used to create high-quality data sets for H&E images. Since considerably more MF can be recognized only in PHH3, this in turn leads to the introduction of label noise. To overcome this problem, we present an attention-based dual-stain classifier which is designed to discriminate MF based on their visibility in H&E. Additionally, we present a data augmentation approach that focuses especially on presenting a large variability of cell pairs to the attention network. The combination of the two methods leads to a weighted accuracy of 0.740 in discriminating H&E-identifiable from non-identifiable MF. Therefore, by automatically discriminating the visibility of MF in H&E slides, PHH3-guided annotation can be used to generate a more reliable ground truth for MF in H&E.

1. Introduction

Mitotic figures (MF) depict cells undergoing cell division and play an important role in the grading schemes of different tumor types. Their density correlates with cell proliferation, which is a key predictor of biological tumor behavior (Louis et al., 2016). In routine pathology, these cells are counted manually on sections stained with hematoxylin and eosin (H&E) by a pathologist within a region of defined size, covering the mitotically most active tumor region. This is a laborious task with low inter-rater and intra-rater concordance (Tellez et al., 2018). The usage of immunohistochemical staining against phospho-histone

H3 (PHH3) for counting MF increases the concordance but leads to generally higher mitotic counts (Duregon et al., 2015). Particularly the former is important to create high quality data sets. Since H&E is the standard stain in routine pathology, any detector intended for broad clinical use should be able to operate on this stain. Still, one could utilize PHH3 stains to generate high-quality annotations for matching H&E-stained images (Tellez et al., 2018). However, it is problematic that several cells are positive for PHH3 but cannot be identified as MF in H&E, which can be attributed to, e.g., the angle at which the cells are cut, staining of early mitotic phases by PHH3 without identifiable morphological changes to the nucleus in H&E, or the clearer distinction between MF and mitosis-like cells in PHH3. This introduces label noise into the data set. To overcome this problem, we present an attention-based dual stain classifier that is designed to discriminate H&E-identifiable from H&E-non-identifiable MF. Additionally, we present a novel data augmentation procedure, which is designed to improve the generalization of the attention mechanism by showing a large variability of cell-pairs to it.

2. Datasets

Our dataset consists of 40 regions of interest (ROI) from 40 cases of canine cutaneous mast cell tumors retrospectively collected from the Department of Veterinary Pathology, Freie Universität Berlin, Germany, each with a size of 2.37 mm^2. First the specimens were stained with H&E and digitized. Afterward they were destained, restained against PHH3 and again digitized. The resulting scans are registered, and MF found in the PHH3-stained slides are localized in their H&E stained counterparts. All ROIs were fully annotated for MF using both stains, as well as for their visibility in H&E and PHH3. We refer to MF that are visible in H&E as H&E-identifiable, and to ones that are only visible in PHH3 as H&E-non-identifiable. Overall, the dataset contains 2,783 MF of which 2,172 (78.05%) are H&E-identifiable and 611 (21.95%) are H&E-non-identifiable.

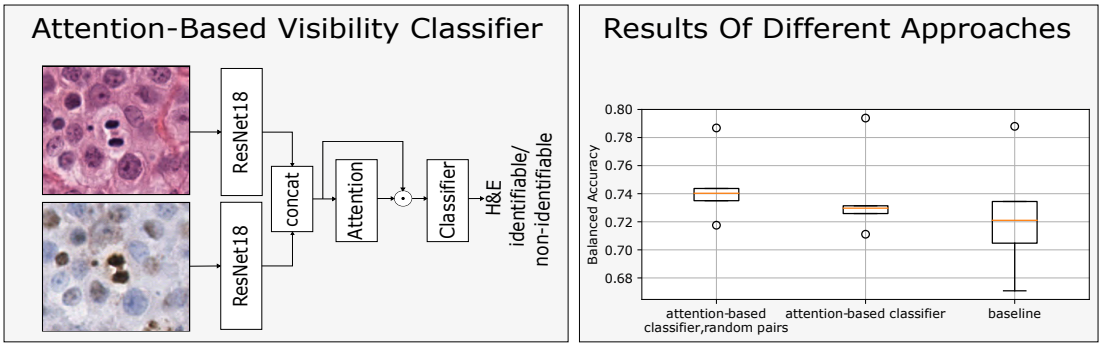

Figure 1: **Left**: Architecture of our attention-based visibility classifier. For feature encoding, two ResNet18 are used, each followed by an average pooling layer and a layer normalization. The attention mechanism is composed of two linear layers, with a hyperbolic tangent activation in between. The classifier only consists of one linear layer. **Right**: Results of the validation on five randomly selected data splits.

3. Methods

The architecture of our attention-based visibility classifier is depicted in Figure 1. The attention mechanism works as follows: The outputs of two ResNet18 feature extractors are concatenated to a 2×512 vector and passed to the attention mechanism. This returns a 1×2 weight vector. The matrix multiplication of both vectors results in a 1×512 weighted feature vector, which is passed to the classification network. Our baseline differs from the attention-based model only in the absence of the attention mechanism. Instead, both vectors are added before the resulting vector is passed to the classifier. To present a large variability to the attention mechanism during training, we developed a novel augmentation strategy. Instead of showing only corresponding H&E/PHH3 pairs during network training, random positive and negative pairs were sampled from the training data at the beginning of each epoch. We denote this strategy as *random pairs approach*. All feature extractors were pre-trained on ImageNet. The models were trained for 30 epochs using focal loss and Adam optimizer with a learning rate of 10^{-4}. To ensure comparability, all models were trained on the same five randomly selected training, validation and test splits. Splits were performed on case level. Model selection was performed on the balanced accuracy metric on the validation set. For evaluation, we compute the balanced accuracy since we have a non-negligible imbalance between H&E positive and negative samples in our dataset. To get an estimate of the quality of the labels in our dataset, we re-presented 100 randomly selected MF from our dataset for re-labeling to our annotator in a single-blinded experiment.

4. Results and Discussion

We find a satisfactory accuracy in all three approaches, as can be seen in Figure 1. The attention-based model in combination with the random pairs approach gives the best results with a median balanced accuracy of 0.740. Without the random pairs approach, the performance drop to 0.730, while the baseline yields a value of 0.721. In comparison, our expert achieved an balanced accuracy score of 0.850 when reproducing his own labels. The results indicate that the described approaches are capable of discriminating MF based on their visibility and that the used attention mechanism, as well as the described sampling approach, give an advantage over the described baseline. We note that the relabeling experiment suggests that we have a non-negligible ambiguity in our visibility labels, which affects our results. Future work should focus on the integration of our pipeline as a data-cleaning step in a PHH3-based annotation pipeline for H&E stained images.

References

Eleonora Duregon et al. *Neuro-Oncology*, 17(5):663–669, 2015. ISSN 15235866. doi: 10.1093/neuonc/nov002.

David N Louis et al. *Acta Neuropathologica*, 131(6):803–820, 2016. ISSN 1432-0533. doi: 10.1007/s00401-016-1545-1.

David Tellez et al. *IEEE Transactions on Medical Imaging*, 37(9):2126–2136, 2018. ISSN 1558254X. doi: 10.1109/TMI.2018.2820199.