Semi-Supervised Hard-Negative Mining and Color Augmentation Strategies: an approach to the MIDOG 2022 Challenge

Maxime W. Lafarge and Viktor H. Koelzer

Department of Pathology and Molecular Pathology, University Hospital and University of Zurich, Zurich, Switzerland

Making histopathology image classifiers robust to a wide range of real-world variability is a challenging task. Here, we describe a candidate deep learning solution for the Mitosis Domain Generalization Challenge 2022 (MIDOG) to address the problem of generalization for mitosis detection in images of hematoxylin-eosin-stained histology slides under high variability (scanner, tissue type and species variability). Our approach consists in training a rotation-invariant deep learning model using aggressive data augmentation with a training set enriched with hard negative examples and automatically selected negative examples from the unlabeled part of the challenge dataset. Our candidate model ensemble achieved a F_1-score of .690 on the preliminary test set after automated evaluation on the challenge platform.

mitosis detection | domain generalization | MIDOG 2022
Correspondence: ***maxime.lafarge@usz.ch***

Introduction

To support the research community with the development of new mitosis detection algorithms that are robust to scanner variability, the MIDOG 2021 challenge (1) lead to an overview of efficient approaches towards solving this task. To further encourage the development of models that can generalize beyond inter-scanner variability, the MIDOG 2022 challenge was initiated (2): in this paper, we describe a candidate solution built upon the strategy we proposed for the MIDOG 2021 challenge (3).

Model architecture

The core of our approach consists on training a deep learning model based on the standard tile-based approach proposed by Cireşan et al. (4). We used a customized ResNet architecture with 70 layers and a single 2×2 max-pooling layer such that our model has a receptive field of 78×78 pixels. We replaced standard 2D convolution layers by P4-group convolution layers to guarantee invariance of our models to 90-degree rotations without requiring train-time or test-time rotation augmentations, and with a moderate computational overhead (5, 6). We adapted the overall architecture to facilitate dense application of the model to large input images.

Dataset Preparation

To train our models, we exclusively used the data provided for the track 1 of the *MIDOG 2022 Challenge*. We split images according to a 80-20 training-validation scheme such that labels and domains were stratified (training set: 7588 mitoses from 283 images; validation set: 1913 mitoses from 71 images).

Given the provided locations of mitotic figures in the images, we derived a set of positive patch examples centered on mitotic figures and a set of negative patch examples whose center is sufficiently distant from the positive examples.

Then, to further enrich the dataset with negative examples we implemented a stain unmixing algorithm derived from (7) to first separate the hematoxylin, eosin and residual components in the unlabeled images of the dataset, and then automatically extract image patches with high intensity in the estimated residual component to enrich the training set with extra negative examples (this procedure resulted in the selection of 72 negative examples after hard negative mining). This approach was motivated by the observation that true positive mitotic events reside in defined stain vectors and are separable from background events such as pigmentation and ink that form common impostors.

Furthermore, we relied on a hard negative mining protocol to remove easily classified negative examples from the training set based on the classification scores achieved by a model trained with a first version of the training dataset (4). We kept removing negative examples over multiple training rounds until the performances on the validation set stopped increasing.

Training Procedure

We trained our models by minimizing the cross-entropy loss with mini-batches of size 128 via stochastic gradient descent with momentum and applied a cosine cyclic learning rate scheduling. We used weight decay regularization and Batch Normalization throughout the network. Mini-batches were balanced between positive and negative examples and image patches were randomly transformed based on an augmentation protocol that consists of the following transformations: spatial shifting, image flipping, elastic deformations and color transformations (linear operations) in RGB space (*CA-RGB*) or Optical Density space (*CA-OD*). We saved the weights of the model that achieved the lowest validation loss and kept the associated models for evaluation.

Table 1. Comparison of F_1-scores of our models (with color augmentation in RGB-space (*CA-RGB*) or OD-space (*CA-OD*) and baselines of the MIDOG 2022 challenge on the four tumor types of the hidden preliminary test set of the challenge.

Model	*Internal Validation*	*Preliminary Test Overall*	*Preliminary Test Tumor 1*	*Preliminary Test Tumor 2*	*Preliminary Test Tumor 3*	*Preliminary Test Tumor 4*
ours (*CA-RGB*)	.784	.646	**.783**	.726	.548	.708
ours (*CA-OD*)	.787	.571	.757	**.775**	.428	.571
ours ensemble	–	.690	.758	.735	.653	.610
MIDOG 2022 Baseline 2	–	**.715**	.744	.732	**.692**	.711
MIDOG 2022 Baseline 1	–	.629	.753	.608	.585	**.743**

Inference Time

Once our models were trained, we produced prediction maps by applying them densely on test images, and then derived candidate detection locations as the set of local maxima in these maps. We then considered all candidate locations with a prediction score above a threshold value as final detection locations (this threshold was chosen as the one maximizing the F_1-score on the validation set).

For our final submission, we created a model ensemble by taking the agreement between the detection sets of two models trained under two variants of our augmentation protocol. The performances of our models is summarized in Table 1.

Discussion

We presented a candidate algorithm that achieves moderate performance while relying mostly on conventional methods (patch-based approach, ResNet architecture, hard negative mining, standard augmentation protocol). Furthermore, we investigated the use of a conventional stain unmixing method to automatically extract potential negative examples in unlabeled images. As this step did not decrease the performances on the validation set, we assumed this helped our models better generalizing and used the resulting enriched training set to train our final models.

Although the reported variations of our augmentation protocol produced similar results on the validation set, we observed that they produced heterogeneous performances on the preliminary test set (Table 1). This suggests that the generalization to this preliminary test set can be sensitive to variations of the augmentation protocol used for training. Indeed, the two reported variations of the augmentation protocol we used helped generalizing to different parts of the unseen variability of the preliminary test set (color transformations in RGB space enabled better generalization for Tumor 1 whereas color transformations in OD space enabled better generalization for Tumor 2).

We will aim at further investigating the effect of this type of variation of augmentation protocol in future work.

References

1. Marc Aubreville, Nikolas Stathonikos, Christof A Bertram, Robert Klopleisch, Natalie ter Hoeve, Francesco Ciompi, Frauke Wilm, Christian Marzahl, Taryn A Donovan, Andreas Maier, et al. Mitosis domain generalization in histopathology images–the midog challenge. *arXiv preprint arXiv:2204.03742*, 2022.
2. Marc Aubreville, Christof Bertram, Katharina Breininger, Samir Jabari, Nikolas Stathonikos, and Mitko Veta. Mitosis domain generalization challenge 2022. In *25th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2022)*, 2022. doi: 10.5281/zenodo.6362337.
3. Maxime W Lafarge and Viktor H Koelzer. Rotation invariance and extensive data augmentation: A strategy for the mitosis domain generalization (midog) challenge. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2021.
4. Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2013.
5. Erik J Bekkers, Maxime W Lafarge, Mitko Veta, Koen AJ Eppenhof, Josien PW Pluim, and Remco Duits. Roto-translation covariant convolutional networks for medical image analysis. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2018.
6. Bastiaan S Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant cnns for digital pathology. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2018.
7. Marc Macenko, Marc Niethammer, JS Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, 2009.