Fine-tuning the Reference Algorithm for the MItosis DOmain Generalization (MIDOG) Challenge with Data Augmentation

Weiqin Zhao[1]

[1]Department of Statistics and Actuarial Science, Hong Kong University, Hong Kong

As described in The MItosis DOmain Generalization challenge, assessing the Mitotic Count has a known high degree of intra- and inter-rater variability. Computer-aided systems have proven to decrease this variability and reduce labelling time. These systems, however, are generally highly dependent on their training domain and show poor applicability to unseen domains. In histopathology, these domain shifts can result from various sources, including different slide scanning systems used to digitize histologic samples. The MItosis DOmain Generalization challenge focuses on this specific domain shift for the task of mitotic figure detection. This work presents a mitotic figure detection algorithm developed a fine-tuned model on the basis of the reference model provided by the challenge, based on domain adversarial training. On the preliminary test set, the algorithm scores an F_1 score of 0.7258.

Correspondence: ***wqzhao98@connect.hku.hk***

Introduction

The Mitotic Count (MC) [1] could be an important and effective metric when evaluating tumor proliferation. It requires pathologists identify mitotc figures from the the whole slide images, which certainly require experiences and great efforts [2]. Previous study shows taht deep learning models could be used for such task, helping pathologists to make faster and more accurate results [2]. These methods ignores the domain shift and they face great performance drop to whole slide images from unseen domains[3]. Several works make effort to solve this problem[4] [5] [3]. However, they all forget to consider the domain shift created by different scanners [6]. The MItosis DOmain Generalization (MIDOG) challenge [7], hosted as a satellite event of the 24^{th} International Conference at Medical Image Computing and Computer Assisted Intervention (MICCAI) 2021, addresses this topic in the form of assessing the MC on a multi-scanner dataset. This work presents my algorithm, which was fine-tuned from the reference model, for the MIDOG challenge. The RetinaNet-based architecture was fine-tuned with the 150 labelled samples from 3 different scanners and scored an F_1 score of 0.7258 on the preliminary test set.

Material and Methods

My algorithm was developed on the basis of the reference model and fine-tuned on the official training subset of the MIDOG dataset.I did not use any additional datasets. As described in the reference algorithm's short paper, the model architecture is based on a publicly available implementation of RetinaNet [8].

Dataset. As described in the MItosis DOmain Generalization (MIDOG) challenge [7], the MIDOG training subset consists of 200 Whole Slide Images (WSIs) from human breast cancer tissue samples stained with routine Hematoxylin & Eosin (H&E) dye. The samples were digitized with four slide scanning systems: the Hamamatsu XR NanoZoomer 2.0, the Hamamatsu S360, the Aperio ScanScope CS2 and the Leica GT450, resulting in 50 WSIs per scanner. For the slides of the first three scanners, a selected field of interest sized approximately $2\,\mathrm{mm}^2$ (equivalent to ten high power fields) was annotated for mitotic figures and hard negative look-alikes. These annotations were collected in a multi-expert blinded set-up. For the Leica GT450, no annotations were available. I only used the labelled samples in the dataset. Fig. 1 shows on example image from the dataset with annotation boxes.

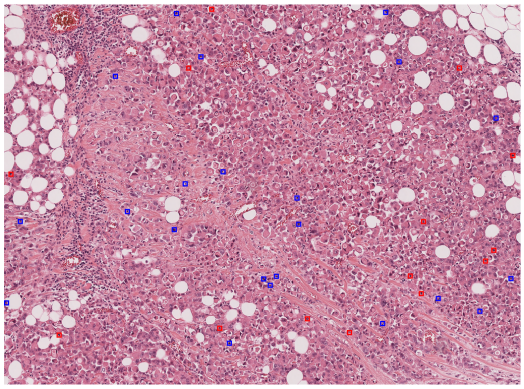

Fig. 1. One example from the dataset.

Fine-tuning with Data Augmentation and Random Patch Sampling. For model architecture of the reference model, the authors used a publicly available RetinaNet implementation [8] by adding a Gradient Reverse Layer (GRL) and a domain classifier. For the encoder of the reference model, they used a ResNet18 backbone pre-trained on ImageNet. For the domain discriminator they were inspired by the work of Pasqualino *et al*. [9] and likewise chose a se-

quence of three blocks consisting of a convolutional layer, batch normalization, ReLU activation and Dropout, followed by an adaptive average pooling and a fully connected layer. Fig. 2 schematically visualizes the modified RetinaNet architecture. In the training, I fine-tuned the model with randomly selected patches with data augmentation methods.

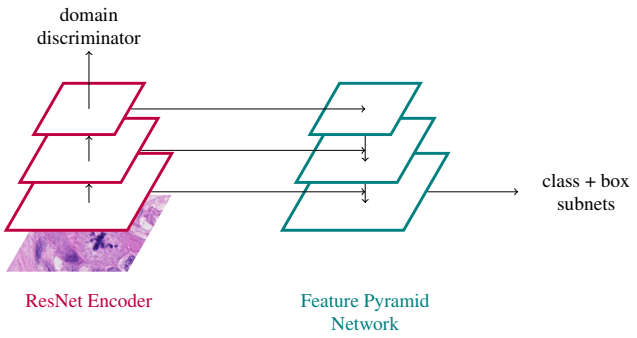

Fig. 2. Domain adversarial RetinaNet architecture.

Network Training. Based on the given reference model provided on the MIDOG Challenge. I used the 150 labelled training data on scanner Hammatsu, scanner S360 and Aperio CS to fine-tune the given reference model. In the training process, I used a patch size of of 512 $\times$ 512 pixels, and the batch size if set to be 25. Each sample in the batch was randomly selected from the 150 labelled training samples by the following rules. First, I randomly selected one sample from the training data, then I randomly chose one mitotic location hard negative location from the annotation of the chosen sample as the center of the selected patch. After adding a random offset ranged between 0 to 256 to the center coordinate, I finally sampled one training patch from the training data. Since 150 is a small number compared to the batch size, I randomly selected 1500 patched in total for each epoch during the training process. To improve the generality of the model, I also applied data augmentation methods, including random flip, random vertical flip, random rotate, random lightening and random contrast change. I also normalized all the training samples before the training according the mean and variance of the training data. I fine tune the reference model for 20 epochs with Adam optimizer with learning rate of 10^{-6}. For loss function, I adopted Focal loss for instance classification results and added the standard RetinaNet loss as the sum of bounding box regression loss. For training sample from Leica scanner, I simply abandoned them with the fine-tuning process, since they were not annotated.As the same in the reference model's short paper, Fig. 2 show the architecture I used for RetinaNet. For other training settings, this model was trained with Pytorch on one 2080Ti GPU.

Evaluation and Results

I use the leaderboard of MIDOG Challenge as evaluation methods. In the preliminary phase, my model scored a F1 score of 0.7258.

Discussion and Conclusion

In this work, I presented my algorithm for the MIDOG challenge, fine-tuning the reference model provided by the challenge with data augmentation. With a validation F_1 score of 0.7258, the performance was actually worse than the original baseline. This is probably mainly resulted in the domain shift between different scanners.

Bibliography

1. DJ Meuten, FM Moore, and JW George. Mitotic count and the field of view area: time to standardize, 2016.
2. Marc Aubreville, Christof A Bertram, Christian Marzahl, Corinne Gurtner, Martina Dettwiler, Anja Schmidt, Florian Bartenschlager, Sophie Merz, Marco Fragoso, Olivia Kershaw, et al. Deep learning algorithms out-perform veterinary pathologists in detecting the mitotically most active tumor region. *Sci Rep*, 10(16447):1–11, 2020.
3. Maxime W Lafarge, Josien PW Pluim, Koen AJ Eppenhof, Pim Moeskops, and Mitko Veta. Domain-adversarial neural networks to address the appearance variability of histopathology images. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 83–91. Springer, 2017.
4. Marc Macenko, Marc Niethammer, James S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1107–1110. IEEE, 2009.
5. David Tellez, Maschenka Balkenhol, Nico Karssemeijer, Geert Litjens, Jeroen van der Laak, and Francesco Ciompi. H and e stain augmentation improves generalization of convolutional networks for histopathological mitosis detection. In *Medical Imaging 2018: Digital Pathology*, volume 10581, page 105810Z. International Society for Optics and Photonics, 2018.
6. Marc Aubreville, Christof Bertram, Mitko Veta, Robert Klopfleisch, Nikolas Stathonikos, Katharina Breininger, Natalie ter Hoeve, Francesco Ciompi, and Andreas Maier. Quantifying the scanner-induced domain gap in mitosis detection. *arXiv preprint arXiv:2103.16515*, 2021.
7. Marc Aubreville, Christof Bertram, Mitko Veta, Robert Klopfleisch, Nikolas Stathonikos, Katharina Breininger, Natalie ter Hoeve, Francesco Ciompi, and Andreas Maier. Mitosis domain generalization challenge. Zenodo, doi: 10.5281/zenodo.4573978, 2021.
8. Christian Marzahl, Marc Aubreville, Christof A Bertram, Jason Stayt, Anne-Katherine Jasensky, Florian Bartenschlager, Marco Fragoso-Garcia, Ann K Barton, Svenja Elsemann, Samir Jabari, et al. Deep learning-based quantification of pulmonary hemosiderophages in cytology slides. *Scientific Reports*, 10(1):1–10, 2020.
9. Giovanni Pasqualino, Antonino Furnari, Giovanni Signorello, and Giovanni Maria Farinella. An unsupervised domain adaptation scheme for single-stage artwork recognition in cultural sites. *Image and Vision Computing*, 107:104098, 2021.