ABSTRACT

Mitotic figure detection is a challenging task in digital pathology that has a direct impact on therapeutic decisions. While automated methods often achieve acceptable results under laboratory conditions, they frequently fail in the clinical deployment phase. This problem can be mainly attributed to a phenomenon called domain shift. An important source of a domain shift is introduced by different microscopes and their camera systems, which noticeably change the colour representation of digitized images. In this method description, we present our submitted algorithm for the Mitosis Domain Generalization Challenge [1], which employs a RetinaNet [5] trained with strong data augmentation and achieves an F1 score of 0.7138 on the preliminary test set.

Keywords Mitosis detection · Domain generalization · Digital pathology

1 Methods

Motivated by recent data-centric approaches we use a RetinaNet [5] trained with strong data augmentation to enforce prediction consistency.

1.1 Dataset

We use the publicly available Mitosis Domain Generalization Challenge (MIDOG) dataset [1]. The data consists of 200 Whole Slide Images (WSIs) from hematoxylin and eosin (HE) stained breast cancer cases. Furthermore, the dataset can be divided into subsets of 50 images, which were acquired and digitized with four different scanners (Aperio ScanScope CS2, Hamamatsu S360, Hamamatsu XR NanoZoomer 2.0, Leica GT450). For three scanners annotations for mitotic figures and hard negatives (imposters) are provided. The disclosed preliminary and final test sets contain samples of two known scanners and two unknown ones.

1.2 Model

Our object detection algorithm consists of a RetinaNet [5] with an EfficientNet B0 [11] backbone. The backbone is initialized with state of the art ImageNet weights, which were trained using RandAugment [2] and Noisy Student [13]. We did not change the feature pyramid and used all five pyramid levels. The network’s heads consist of four layers with a channel size of 128. Anchor ratios are set to one while the differential evolution search algorithm introduced by [14] is employed to determine three anchor scales (0.781, 1.435, 1.578).

1.3 Domain generalization through augmentation

Our main method to approach domain generalization is data augmentation. Data-driven approaches such as RandAugment [2] have been proven to increase model robustness and have been used in recent state of the art models.

Figure 1: Used augmentations with different strengths.

Inspired by Trivial Augment [7] a very simple random augmentation strategy is used, where a single augmentation is applied to each image. The augmentations are drawn uniformly from a set of color, noise and special transformations while the augmentation strength is random to some defined degree. The pool of augmentations consists of color jitter, HE [12], fancy PCA, hue, saturation, equalize, random contrast, auto-contrast, contrast limited adaptive histogram equalization (CLAHE), solarize, solarize-add, sharpness, Gaussian blur, posterize, cutout, ISO noise,
JPEG compression artefacts, pixel-wise channel shuffle and Gaussian noise. In addition, every image is randomly flipped and RGB channels are randomly shuffled.

1.4 Training and evaluation

For experimentation, we divide the dataset into five folds with three training, one validation and one test split for each scanner (test splits are added to the train set for submissions). During the training phase, we uniformly sample the images of the train set and randomly select a mitotic figure or an imposter annotation. A patch with a size of 448 pixels is randomly cropped around the selected annotation similar to [6]. The RetinaNet is trained for 100 pseudo epochs with a batch size of 16 using the super-convergence scheme [3]. Adam optimizer with a maximum learning rate of 1e-4 is used. The best models are selected based on the lowest validation loss. After the training phase, we combine the training and validation set and optimize the model’s confidence threshold with respect to the best F1 score. During inference, incoming WSIs are tiled into overlapping patches of 448 pixels. All models are trained and tested using an Nvidia GeForce RTX 3060 with 12GB GPU RAM.

2 Results

For the final submission, we only use labelled data to train a single RetinaNet with the proposed data augmentation strategy. This method achieves an F1 score of 0.7138 on the preliminary test set of the MIDOG challenge.

3 Discussion

Overall, we are able to generalize better across multiple scanner domains with strong data augmentation. The magnitude at which such simple transformations improve generalization at no cost of inference speed is higher than expected. Even models trained with only one scanner reach similar results on our test split, showing only a small performance drop. In the following, we will lay out unsuccessful attempts to improve the quality further. One major issue was the model selection based on the validation loss. The models were not capable of overfitting the data, assumingly due to the sampling and the strong data augmentation, models ended up in an equilibrium mode where performance improvements were wiggling between the different scanners back and forth. Because of that, the representation shift metric proposed by Stacke et al. [10] was tested. It was applied to the three convolutional layers, which flow into the feature pyramid, but was found to not help the model selection process. Another strategy was a dual-stage attempt with a verification net proposed by Li et al. [4]. The network was trained on the predicted patches of the first stage using the same augmentation and in addition a Gradient Reversal Layer [2] to remove even the last bits of scanner dependent information. Unfortunately, this resulted in a performance drop of 12.1% on the preliminary dataset. Finally, the choice of using an EfficientNet originated from the attempt to incorporate the unlabeled data using a self-supervised Student-Teacher learning procedure based on the STAC framework [9]. While increasing the performance on our test split, this resulted in a small performance drop of 1% on the preliminary dataset. One problem was that producing pseudo labels with a high confidence threshold resulted in very few labelled samples while self-training reportedly needs a huge amount of pseudo labelled data to make use of it. A second problem arises with false positive pseudo labels. We used a labelled scanner to check the number of wrong labels incorporated in the pseudo labels and found that for mitotic figures pseudo labels were mainly correct while hard negatives actually included a lot of mitotic figures. This probably led to more confusion than having a positive effect.

References

[1] Aubreville, M., Bertram, C., Veta, M., Klopfleisch, R., Stathonikos, N., Breininger, K., ter Hoeve, N., Ciompi, F., Maier, A.: Mitosis domain generalization challenge (Mar 2021). https://doi.org/10.5281/zenodo.4573978

[2] Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.: RandAugment: Practical Automated Data Augmentation with a Reduced Search Space. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 18613–18624. Curran Associates, Inc. (2020)

[3] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-Adversarial Training of Neural Networks. Journal of Machine Learning Research 17(59), 1–35 (2016)

[4] Li, C., Wang, X., Liu, W., Latecki, L.J.: Deep-Mitosis: Mitosis detection via deep detection, verification and segmentation networks. Medical Image Analysis 45, 121–133 (Apr 2018). https://doi.org/10.1016/j.media.2017.12.002

[5] Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal Loss for Dense Object Detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988 (2017)

[6] Marzahl, C., Aubreville, M., Bertram, C.A., Stuyt, J., Jasensky, A.K., Bartenschlager, F., Fragoso-Garcia, M., Barton, A.K., Elsemann, S., Jabari, S., Krauth, J., Madhu, P., Voigt, J., Hill, J., Klopfleisch, R., Maier, A.: Deep Learning-Based Quantification of Pulmonary Hemosiderophages in Cytology Slides. Scientific Reports 10(1), 9795 (Dec 2020). https://doi.org/10.1038/s41598-020-65958-2

[7] Müller, S.G., Hutter, F.: TrivialAugment: Tuning-Free Yet State-of-the-Art Data Augmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 774–782 (2021)
[8] Smith, L.N., Topin, N.: Super-convergence: very fast training of neural networks using large learning rates. In: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications. vol. 11006, pp. 369–386. SPIE (May 2019). https://doi.org/10.1117/12.2520589

[9] Sohn, K., Zhang, Z., Li, C.L., Zhang, H., Lee, C.Y., Pfister, T.: A Simple Semi-Supervised Learning Framework for Object Detection. [arXiv:2005.04757] [cs] (Dec 2020). arXiv: 2005.04757

[10] Stacke, K., Eilertsen, G., Unger, J., Lundstrom, C.: Measuring Domain Shift for Deep Learning in Histopathology. IEEE journal of biomedical and health informatics 25(2), 325–336 (Feb 2021). https://doi.org/10.1109/JBHI.2020.3032060

[11] Tan, M., Le, Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Proceedings of the 36th International Conference on Machine Learning. pp. 6105–6114. PMLR (May 2019). iSSN: 2640-3498

[12] Tellez, D., Balkenhol, M., Otte-Holler, I., van de Loo, R., Vogels, R., Bult, P., Wauters, C., Vreuls, W., Mol, S., Karssmeijer, N., Litjens, G., van der Laak, J., Ciompi, F.: Whole-Slide Mitosis Detection in H&E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks. IEEE Transactions on Medical Imaging 37(9), 2126–2136 (Sep 2018). https://doi.org/10.1109/TMI.2018.2820199

[13] Xie, Q., Luong, M.T., Hovy, E., Le, Q.V.: Self-Training With Noisy Student Improves ImageNet Classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10687–10698 (2020)

[14] Zlocha M., Dou Q., Glocker B.: Improving RetinaNet for CT Lesion Detection with Dense Masks from Weak RECIST Labels. In: Shen D. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science, vol 11769, pp 402-410. Springer, Cham. https://doi.org/10.1007/978-3-030-32226-7_45

Acknowledgments
This work was supported by the Bavarian Ministry of Economic Affairs, Regional Development and Energy through the Center for Analytics – Data – Applications (ADA-Center) within "BAYERN DIGITAL II" and by the BMBF (16FMD01K, 16FMD02 and 16FMD03).