Systematic investigation of Basic Data Augmentation Strategies on Histopathology Images

Jonas Annuscheit, Benjamin Voigt, Oliver Fischer, Patrick Baumann, Sebastian Lohmann, Christian Krumnow and Christian Herta

University of Applied Sciences (HTW) Berlin, Centrum für Biomedizinische Bild- und Informationsverarbeitung (CBMI), Ostendstraße 25, 12459 Berlin, Germany
{first name}.{last name}@htw-berlin.de

Abstract. Recent years have witnessed the rapid progress of deep neural networks. However, in supervised learning, the success of the models hinges on a large amount of training data. Therefore, data augmentation techniques were developed to increase the effective size of the training data. Using such techniques is especially important for domains where the amount of available data is limited. In digital pathology, data augmentation is therefore often applied to improve the performance of classifications. This work systematically investigates single data augmentation techniques on different datasets using multiple network architectures. Furthermore, it proposes guidelines on using data augmentation when training deep neural networks on histopathological data.

Keywords: Convolutional Neural Network, Data Augmentation, Digital Pathology

1 Introduction

The prediction quality of supervised learning models relies on the available data's quantity, quality, and heterogeneity. In training a deep neural network, these factors are essential to create a robust and generalizing model. Different transformation techniques can be utilized on the available data to synthesize new samples if a dataset lacks some of these aspects. Such techniques are summarized under the term data augmentation.

Nowadays, there are a variety of different augmentation methods to synthesize new data. These range from classic image manipulation approaches to more contemporary methods like training with adversarial examples [1] or generating entirely new datasets using generative adversarial networks [2]. Several studies investigated the effect of such data transformations on traditional machine learning datasets and proved their benefit [3, 4]. We applied and reviewed some of these transformations to the domain of histopathological datasets.

In recent years, the medical field of pathology has been subject to digital change. Part of this change is to aid the traditional diagnostics, i.e., inspecting extracted tissue sections under a light microscope with computer algorithms [5, 6]. A promising option is to let machine learning or deep learning support pathologists' diagnostic work. Therefore, numerous research studies attempt to answer specific pathological questions using neural networks. Since these questions are usually image classification problems, the approaches use the supervised learning regime, utilizing convolutional neural networks (CNN) [7–9].

Although there are examples of publicly available digitized tissue samples [10], there is a lack of well-curated datasets useful for the supervised learning approach. In addition, highlighting the need for data augmentation methods in this domain, most public datasets

are relatively small. Collecting suitable images for a given medical problem is challenging due to the non-uniformity of manifestations and the need to consider patient rights. Labeling these images requires the highly specialized expertise of a pathologist, adding to an already busy workload.

In this work, we build a pipeline to systematically investigate basic data augmentation techniques on different classification datasets and network architectures. For this purpose, we selected two public histopathological datasets for different medical problems: *classification of mitosis candidates* and *tissue type classification*. We trained three contemporary CNN architectures for all of these data sets, examining different types of augmentation methods. This paper describes the experimental setup to measure the influence of a single data augmentation technique on the model's performance. In addition, it proposes guidelines for using data transformations in the supervised learning setting regarding different types of histopathological data. Finally, it discusses under which circumstances data augmentation has a reliable benefit for a model's training process.

2 Related Work

Due to its regularization effect, data augmentation is a popular method used in deep learning pipelines to reduce overfitting and increase the robustness of a model, especially concerning an image classification problem. In fact, the method is so established that several tools exist to make standard techniques more accessible [11–13] or even automate the augmentation process [14, 15].

Several studies examined the actual influence of different data augmentation techniques and showed its beneficial effects in the context of natural images [3, 16]. A widely used taxonomy to divide the common techniques into categories is *basic image manipulations*, e.g., geometric transformations, cropping, occlusion, noise injection, filtering, color transformations, and *deep learning approaches*, e.g., adversarial training, style transfer, synthetic image generation via generative adversarial networks [3, 17].

Unlike in the natural image domain, where datasets can provide millions of images, far fewer qualitatively annotated samples are available in the field of histopathology. Hence, data augmentation has established itself as an integral part of the training pipelines in this area as well. Interestingly, it is almost exclusively the use of newer augmentation techniques from the deep learning approaches that have been broadly reviewed thus far. Generative adversarial networks (GANs) were investigated to solve the stain normalization problem using style transfer methods. Color differences and disturbances are a considerable challenge through various tissue staining protocols and the varying digitization processes. Style transfer can homogenize the color distribution in a data set and thus the distributional shift in a dataset [18–20]. Some studies even explored the transfer of staining protocols utilizing GANs; e.g., Mercan et al. trained a model that converts images obtained from H&E stained tissue into virtual PHH3 staining [21]. In addition, GANs are used to synthesize completely artificial samples to enrich small data sets [22–24].

Concerning *basic image manipulation*, many approaches use several techniques to augment their datasets but do not evaluate the influence of augmentations; see, e.g. [25, 26]. Primarily, basic manipulation techniques are used intensively in conjunction with semi-supervised learning methods, which are becoming increasingly popular in this domain [27–29]. Color transformations, in particular, are one of the most widely used techniques due to the nature of histopathological images [26]. Tellez et al.[30] and Karimi et al.[31]

examined the stain normalization problem more closely and developed custom augmentation techniques for it.

However, in-depth studies which comprehensively evaluate the effect of basic image manipulation techniques can only be found for radiology images in the medical domain [17, 32–34].

3 Method

We have developed a pipeline to measure the effect of data augmentation techniques in supervised learning on histopathology datasets.[1] We can configure experiments as a triple $(d, m, t) \in D \times M \times T$ where D is the set of possible datasets, M denotes the set of considered deep neural network architectures, and T corresponds to the set of different data augmentations. A specific transformation t manipulates online the batches drawn from the dataset d to create transformed batches, that are used to train a deep neural network m. Keeping the pair (d, m) constant, the influence of t on the trained model can be measured by comparing its parameters and performance. This section describes the sets D, M, T in more detail and contains information about the training and evaluation protocol used.

3.1 Datasets D

Set D consists of the publicly available datasets MIDOG[2] and BACH[3]. Both sets were pre-processed to fit a classification problem.

The task of the BACH dataset is to distinguish between four different tissue types. It is a tiny dataset with images of 2048x1536 pixels and 100 samples per class, i.e., 400 samples overall. Therefore, we cropped patches from the original images using a 512x512 window with a 256-pixel step to increase the dataset size. In addition, we discarded patches not containing any H&E-stained tissue during the process by removing tiles with less than 3% tissue. Finally, we split the dataset into three subsets using random sampling for training (4801 samples), validation (1655 samples), and test (1647 samples). Patches with overlapping pixels in the subsets were removed. The classes of the dataset are nearly balanced.

The MItosis DOmain Generalization Challenge [35] published a dataset of human breast cancer tissue samples. However, we note that up to this point, only the training set is publicly available, consisting of 1721 mitotic figures and 2714 non-mitotic examples. The samples were acquired using three different whole slide image scanners and annotated by trained pathologists with a multi-expert blind annotation pipeline. We pre-processed the dataset by cropping a 250x250 patch around each annotation center. We sampled three distinct subsets keeping the class balance: training (2219), validation (1071), and test (1145).

3.2 Models M

The model set M consists of the networks VGG[36], Inception[37] and Densenet[38]. These networks form a cross-section over the development of CNN architecture and, therefore,

[1] The source code is available via Github: https://github.com/CBMI-HTW/Data-Augmentation-Histology

[2] https://imi.thi.de/midog/the-dataset/

[3] https://iciar2018-challenge.grand-challenge.org/Dataset/

have distinct structural elements. We intend to investigate whether these structures react differently to data augmentation. We use a pre-trained PyTorch model with reinitialized classification head, i.e. *vgg11_bn*, *inception_v3* and *densenet121*, as baseline for each network type. These models require as input square images of a model-dependent size n_m, i.e. the input images have shape $n_m \times n_m$.

3.3 Transformations T

The transformations in T fall in the categories: color-based, geometric-based, filter-based transformations, and erasing. All transformations are realized by using the implementation of the torchvision library.

In our setup, transformations are applied with a certain probability p, with $p = 1$ if not stated differently. For most of these transformations t an additional parameter s controls the strength of the transformation on the input x, i.e. the output of the transformation is $t_s(x)$. For each corresponding transformation, s is sampled from a certain interval, where the size of this interval is a hyperparameter in our setting.

The hyperparameters of the geometric-based transformations are determined by n_m and the maximal distortion without getting blacked borders. For all other transformations we perform a hyperparameter optimization on the BACH dataset and VGG model to identify the best parameter ranges. We trained for each configuration of parameters a minimum of 7 models and choose the setting with the highest mean validation accuracy as the best performing one. These models were only used to determine the hyperparamters and not for the final test results.

Color-Based Transformation As color-based transformations we use the standard transformation for brightness, contrast, gamma value, hue angle and saturation. The strength s is sampled randomly from the interval $[s_0 - s_1, s_0 + s_1]$ where, the center s_0 is defined by $t_{s_0} = \text{identity}$. Here, we sample s indirectly by drawing $r \in [0, 1]$ according to the beta distribution $\text{Beta}(\alpha = 8, \beta = 8)$ and computing $s = s_0 + (2r - 1)s_1$. The half-width s_1 was determined in a hyperparameter optimization. The parameters for the hyperparameter optimization are summarized in Tab. 1, where the intervals for choosing s_1 where found iteratively by hand over multiple trials, ensuring that the chosen value does not lie on the boundary.

Geometric-Based Transformation As geometric transformations, we investigate flips, rotations, random cropping to size n_m, shearing and scaling. For scaling, rotation, and shearing, we sample s uniformly from an interval $[s_0 - s_1, s_0 + s_1]$ as summarized in Tab. 1, where we consider two scaling scenarios. In the case of the flip transformation, we apply horizontal and/or vertical flips to the input, each with a probability of 0.5. Both scaling transformations, as well as shearing, is done with $p = 0.9$.

Filter- and Erasing-Based Transformation As filter-based transformations, we study Gaussian blurring as well as a sharpness adaption.

For Gaussian blurring, we pick uniformly an odd kernel size between 3 and 15 and use a minimum and maximum sigma (these are direct inputs to the trochvision implementation) of 0.001 and 0.5, respectively and set $p = 0.5$. We did a hyperparameter search for the maximal kernel-size in $\{7, 11, 15\}$, the maximum sigma value in $[0.5, 8]$ and p in $\{0.5, 0.75\}$.

Table 1. Parameters used for the color- and geometric-based transformations. The center s_0 is fixed where the interval half-width s_1 was either optimized with a hyperparameter optimization within the given intervals or chosen as stated.

transformation	s_0	s_1	choice of s_1	transformation	s_0	s_1	choice of s_1
brightness	1.0	0.0175	$[0.005, 0.3]$	rotation	0	180°	fixed geometrically
contrast	1.0	0.1	$[0.025, 0.6]$	scale I	1	0.15	chosen by hand
hue	0.0	0.00625	$[0.00025, 0.6]$	scale II	1	0.29	fixed geometrically
saturation	1.0	0.025	$[0.01, 0.6]$	shear	0	22°	fixed geometrically
gamma	1.0	0.05	$[0.00625, 0.6]$				

The sharpness adaption depends on a parameter $s \geq 0$ where for $s < 1$, the image is blurred and for $s > 1$ sharpened and $s = 1$ corresponds to the identity. To define s we sample $r \in [0, 1]$ according to Beta($\alpha = 8, \beta = 8$) and set $s = 2r$ if $r \leq 0.5$ and $s = 4.5(2r - 1) + 1$ if $r > 0.5$. The factor 4.5 determines the maximal sharpening factor and was chosen from a hyperparameter search in the interval $[1.0, 8.0]$.

For erasing-based transformations, we select 3 potentially overlapping erasing rectangles with a size ranging from $0.01n_m$ to $0.2n_m$ and aspect ratio ranging from 0.5 to 2. The trochvision implementation selects the corresponding parameters uniformly within these ranges. Each rectangle is then applied with $p = 0.75$. The applied rectangles are then filled with zeros (erasing (black)) or with white noise (erasing (random)). All three parameters, the number of rectangles, the maximal size and the appliance probability were hyperparameter optimized for the erasing (black) scenario over the sets $\{1, 3\}$, $[0.025n_m, 0.4n_m]$ and $\{0.5, 0.75\}$, respectively.

3.4 Training Protocol

For the data augmentation experiments we calculated all triplet combinations (d, m, t). We implemented equal training settings to ensure maximum comparability with the baseline models, i.e., optimizer, scheduler, learning rates, weight initialization, and fixed data loading. The only difference in the pipeline was using the investigated transformation t in the default transformation sequence.

To implement one training cycle, we followed Chollet's recommendations [39] on fine-tuning. First, we trained reinitialized layers for a warm-up period before updating all network parameters' to prevent the negative effect of a possible large error signal on previously learned features. We implemented this by training 10 epochs with a 1-cycle learning rate scheduler [40], a LAMB optimizer [41] and a L2 loss regularization. Then, followed by another training over 50 epochs, all weights are updated in the network using the same optimizer and scheduler policies. We always selected the last model of the training process for the evaluation to ensure that the data was seen equally.

When loading a sample x, we apply a fixed transformation sequence. We perform this sequence to avoid artificial padded black borders in arbitrary rotation transformations. However, to maintain comparability, the sequence is always used. (1) Resize the x to $\sqrt{2} * n_m$ with n_m being the network input size. (2) Apply the investigated transformation t to the resized sample with probability p. (3) Center crop the transformed sample with the size n_m. (4) Normalize the final input by the mean and standard deviation of each color channel. Exceptions to this sequence are erasing-based transformations, which are applied at the end of the sequence to not corrupt the normalization process.

We performed a hyperparameter optimization for the training settings of each baseline model and dataset (d, m) using grid search. These final settings are reported in table 2 and used for the training of each triplet (d, m, t).

Table 2. Final parameter used for the training of each triplet (d, m, t) where the parameters are optimized for the baseline model of every pair $(d, m) \in D \times M$.

	BACH			MIDOG		
Parameter	VGG	Inception	Densenet	VGG	Inception	Densenet
Epochs (total)	60	60	60	60	60	60
Epochs (warm-up)	10	10	10	10	10	10
Learning Rates	0.0016	0.0016	0.0016	0.0064	0.0032	0.0032
Weight Decay	0.01	0.01	0.01	0.01	0.01	0.01
Batch Size	64	64	64	64	64	64

3.5 Evaluation Protocol

To account for the randomness during training given due to weight initialization of the classification head, dropout layer and random application of augmentations, that results in a spread of experiment outcomes, we conducted 10 experiment runs per configuration (d, m, t). The 10 experiments only differ by a random seed given at the start of training. This allows gathering a statistic, which makes a more accurate statement about the effectiveness of an augmentation and also enables an expressive comparison between them. As metric, we use the accuracy value on the plain test set for testing the model.

4 Results and Discussion

We summarize our findings for each triplet in a boxplot.[4] We interpret an augmentation technique as effective if its median is above the mean of the baseline model (continuous horizontal line) and their overlap in the interquartile range (IQR = Q_3 - Q_1) is minimal. We show the results for each dataset in Figure 1 and Figure 2 grouped by the augmentation strategies and the network architecture. For a clean plot appearance, a few extreme outliers were discarded from the visualization.

For the BACH dataset, we observe an expected behavior of the baseline models. With the rising complexity of the network architecture, the accuracy increases slightly. However, comparing the baseline with models trained using data augmentations shows that the geometric transformations stand out. Since the histopathological data is rotation invariant and the importance of morphological structures, we anticipated such results. Furthermore, erasing-based augmentations also provide a beneficial effect. On the other hand, filter-based and color-based transformations do not seem to have a positive influence, nor do they harm the model's performance for the chosen hyperparameters. Especially regarding the color-based augmentations methods, that finding was surprising since we presumed the manipulation of the color space to be a critical factor in the context of histopathological data. Our hypothesis that the color distortions caused by

[4] Interactive versions of the charts and more details of the results can be viewed at https://cbmi-htw.github.io/Data-Augmentation-Histology-Website/

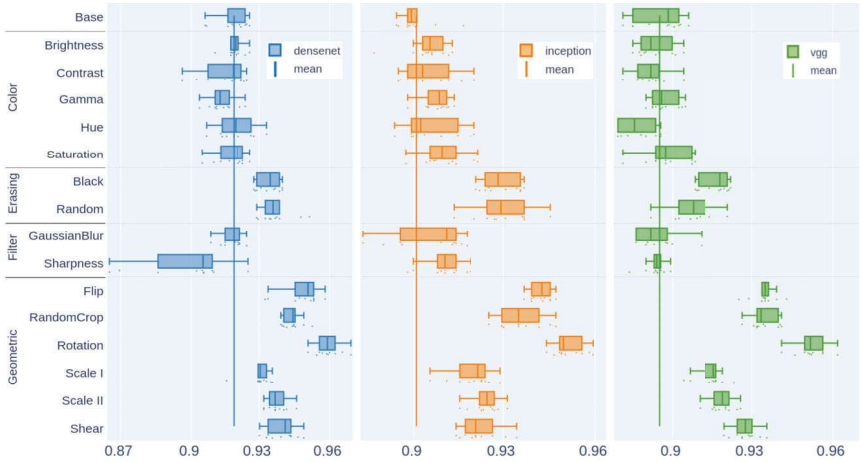

Fig. 1. Results for the BACH dataset. We show the individual mean accuracies on the test set as well as resulting boxplots for the 10 runs for the baseline configuration and the individual transformations for all three network architectures.

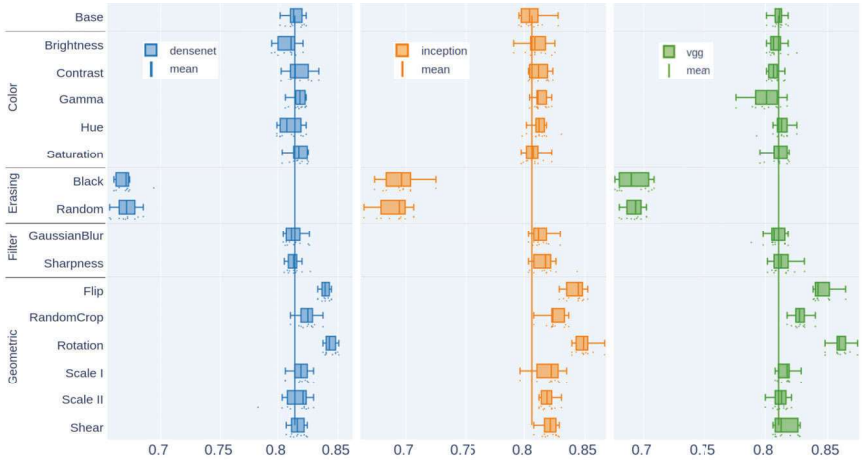

Fig. 2. Results for the MIDOG dataset. We show the individual mean accuracies on the test set as well as resulting boxplots for the 10 runs for the baseline configuration and the individual transformations for all three network architectures.

the hyper-optimized transformations were too close to the identity could not be verified. Additional experiments with artificial parameter values for the transformations to change the input drastically lead to similar results. The results for the MIDOG dataset are essentially alike. Geometric-based augmentations raise the performance, whereas color-based and filter-based transformations have no significant effect. Indeed, the scatter and the overlap with the baseline is so tremendous that no positive effect can be ascribed to these transformations. Erasing-based augmentation harms the model performance significantly. We assume that the augmentation occasionally covers the cell nucleus, which is essential for distinguishing between mitosis and non-mitosis. However, we did not investigate this further and leave this for future work.

Additionally, we have to mention that it took much computational effort to get the results reliable and robust due to intensive hyperparameter optimization. We observed a strong sensitivity towards the hyperparameters of the learning process. Tiny changes in the training settings, e.g., learning rate or batch size, let the benefits of data augmentation vanish in noise. Therefore, we advise using augmentation techniques in combination with well-optimized training hyperparameters to profit from the method. We even suggest tuning the transformation parameters for an optimal result.

5 Conclusion

We examined basic data augmentation techniques frequently used in deep learning classifier training pipelines on two histopathological data. Overall geometric-based techniques increase the model performance on such datasets. However, surprisingly, color-based augmentations do not have the expected impact and are costly due to the required parameter optimization. Next to supervised learning settings, we expect this work to improve contemporary semi-supervised learning methods, e.g., contrastive learning, and assume such methods will considerably impact the training of deep learning models in the histopathological domain.

Acknowledgments
The authors thank Dr. Tim-Rasmus Kiehl, M.D. (Charité, Institute of Pathology) for helpful comments and acknowledge the financial support by the Federal Ministry of Education and Research of Germany (BMBF) in the project deep.HEALTH (13FH770IX6).

References

1. Miyato, T., Maeda, S., Koyama, M., Nakae, K., Ishii, S.: Distributional smoothing by virtual adversarial examples. In: 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings. (2016)
2. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
3. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of Big Data **6**(1) (2019) 1–48
4. O'Gara, S., McGuinness, K.: Comparing data augmentation strategies for deep image classification. In: Irish Machine Vision and Image Processing Conference (IMVIP). (2019)
5. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: Challenges and opportunities. Medical image analysis **33** (2016) 170–175
6. Niazi, M.K.K., Parwani, A.V., Gurcan, M.N.: Digital pathology and artificial intelligence. The lancet oncology **20**(5) (2019) e253–e261

7. Ertosun, M.G., Rubin, D.L.: Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks. In: AMIA Annual Symposium Proceedings. Volume 2015., American Medical Informatics Association (2015) 1899
8. Li, C., Wang, X., Liu, W., Latecki, L.J.: Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks. Medical image analysis **45** (2018) 121–133
9. Rouhi, R., Jafari, M., Kasaei, S., Keshavarzian, P.: Benign and malignant breast tumors classification based on region growing and cnn segmentation. Expert Systems with Applications **42**(3) (2015) 990–1002
10. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: The cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary oncology **19**(1A) (2015) A68
11. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Information **11**(2) (2020)
12. Casado-García, A., Domínguez, C., García-Domínguez, M., Heras, J., Inés, A., Mata, E., Pascual, V.: Clodsa: a tool for augmentation in classification, localization, detection, semantic segmentation and instance segmentation tasks. BMC bioinformatics **20**(1) (June 2019) 323
13. Bloice, M.D., Roth, P.M., Holzinger, A.: Biomedical image augmentation using augmentor. Bioinformatics **35**(21) (2019) 4522–4524
14. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation strategies from data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019) 113–123
15. Lim, S., Kim, I., Kim, T., Kim, C., Kim, S.: Fast autoaugment. Advances in Neural Information Processing Systems **32** (2019) 6665–6675
16. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621 (2017)
17. Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., Haworth, A.: A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology (2021)
18. Cho, H., Lim, S., Choi, G., Min, H.: Neural stain-style transfer learning using gan for histopathological images. arXiv preprint arXiv:1710.08543 (2017)
19. Zanjani, F.G., Zinger, S., Bejnordi, B.E., van der Laak, J.A., de With, P.H.: Stain normalization of histopathology images using generative adversarial networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE (2018) 573–577
20. Mahapatra, D., Bozorgtabar, B., Thiran, J.P., Shao, L.: Structure preserving stain normalization of histopathology images using self supervised semantic guidance. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2020) 309–319
21. Mercan, C., Mooij, G., Tellez, D., Lotz, J., Weiss, N., van Gerven, M., Ciompi, F.: Virtual staining for mitosis detection in breast histopathology. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE (2020) 1770–1774
22. Xue, Y., Zhou, Q., Ye, J., Long, L.R., Antani, S., Cornwell, C., Xue, Z., Huang, X.: Synthetic augmentation and feature-based filtering for improved cervical histopathology image classification. In: International conference on medical image computing and computer-assisted intervention, Springer (2019) 387–396
23. Wei, J., Suriawinata, A., Vaickus, L., Ren, B., Liu, X., Wei, J., Hassanpour, S.: Generative image translation for data augmentation in colorectal histopathology images. Proceedings of machine learning research **116** (2019) 10
24. Quiros, A.C., Murray-Smith, R., Yuan, K.: Pathologygan: Learning deep representations of cancer tissue. In: Proceedings of the Third Conference on Medical Imaging with Deep Learning. (2020)
25. Roy, K., Banik, D., Bhattacharjee, D., Nasipuri, M.: Patch-based system for classification of breast histology images using deep learning. Computerized Medical Imaging and Graphics **71** (2019) 90–103

26. Lafarge, M.W., Pluim, J.P., Eppenhof, K.A., Moeskops, P., Veta, M.: 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. Springer (2017) 83–91
27. Ciga, O., Martel, A.L., Xu, T.: Self supervised contrastive learning for digital histopathology. arXiv preprint arXiv:2011.13971 (2020)
28. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., et al.: Big self-supervised models advance medical image classification. arXiv preprint arXiv:2101.05224 (2021)
29. Liu, Q., Louis, P.C., Lu, Y., Jha, A., Zhao, M., Deng, R., Yao, T., Roland, J.T., Yang, H., Zhao, S., et al.: Simtriplet: Simple triplet representation learning with a single gpu. arXiv preprint arXiv:2103.05585 (2021)
30. Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.M., Ciompi, F., van der Laak, J.: Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical image analysis **58** (2019) 101544
31. Karimi, D., Nir, G., Fazli, L., Black, P.C., Goldenberg, L., Salcudean, S.E.: Deep learning-based gleason grading of prostate cancer from histopathology images—role of multiscale decision aggregation and data augmentation. IEEE journal of biomedical and health informatics **24**(5) (2019) 1413–1426
32. Fabian, Z., Heckel, R., Soltanolkotabi, M.: Data augmentation for deep learning based accelerated mri reconstruction with limited data. In: International Conference on Machine Learning, PMLR (2021) 3057–3067
33. Tran, N.T., Tran, V.H., Nguyen, N.B., Nguyen, T.K., Cheung, N.M.: On data augmentation for gan training. IEEE Transactions on Image Processing **30** (2021) 1882–1897
34. Nalepa, J., Marcinkiewicz, M., Kawulok, M.: Data augmentation for brain-tumor segmentation: A review. Front Comput Neurosci **13** (12 2019) 83
35. Aubreville, M., Bertram, C., Veta, M., Klopfleisch, R., Stathonikos, N., Breininger, K., ter Hoeve, N., Ciompi, F., Maier, A.: Mitosis domain generalization challenge (March 2021)
36. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In Bengio, Y., LeCun, Y., eds.: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. (2015)
37. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 2818–2826
38. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2017) 4700–4708
39. Chollet, F.: Deep Learning with Python. Manning (November 2017)
40. 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. Volume 11006., International Society for Optics and Photonics (2019) 1100612
41. You, Y., Li, J., Reddi, S., Hseu, J., Kumar, S., Bhojanapalli, S., Song, X., Demmel, J., Keutzer, K., Hsieh, C.J.: Large batch optimization for deep learning: Training bert in 76 minutes. In: International Conference on Learning Representations. (2020)