DeepAf: One-Shot Spatiospectral Auto-Focus Model for Digital Pathology

Yousef Yeganeh[1,2], Maximilian Frantzen[1], Michael Lee[3], Kun Hsing-Yu[4,5,6,7], Nassir Navab[1,2], and Azade Farshad[1,2]

[1]Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[2]Munich Center for Machine Learning (MCML)
[3]Southern Taiwan University of Science and Technology, Taiwan
[4]Dep. of Biomedical Informatics and [5] Pathology, Brigham and Women's Hospital, Boston, MA, USA
[6]Harvard Data Science Initiative and [7]Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, MA, USA
y.yeganeh@tum.de

Abstract. While Whole Slide Imaging (WSI) scanners remain the gold standard for digitizing pathology samples, their high cost limits accessibility in many healthcare settings. Other low-cost solutions also face critical limitations: automated microscopes struggle with consistent focus across varying tissue morphology, traditional auto-focus methods require time-consuming focal stacks, and existing deep-learning approaches either need multiple input images or lack generalization capability across tissue types and staining protocols. We introduce a novel automated microscopic system powered by DeepAf, a novel auto-focus framework that uniquely combines spatial and spectral features through a hybrid architecture for single-shot focus prediction. The proposed network automatically regresses the distance to the optimal focal point using the extracted spatiospectral features and adjusts the control parameters for optimal image outcomes. Our system transforms conventional microscopes into efficient slide scanners, reducing focusing time by 80% compared to stack-based methods while achieving focus accuracy of 0.18 μm on same-lab samples—matching the performance of dual-image methods (0.19 μm) with half the input requirements. DeepAf demonstrates robust cross-lab generalization with only 0.72% false focus predictions and 90% of predictions within the depth of field. Through an extensive clinical study of 536 brain tissue samples, our system achieves 0.90 AUC in cancer classification at 4× magnification, a significant achievement at lower magnification than typical 20× WSI scans. This results in a comprehensive hardware-software design enabling accessible, real-time digital pathology in resource-constrained settings while maintaining diagnostic accuracy.

Keywords: Digital Pathology · Auto-focusing · Spatiospectral

1 Introduction

Through the digitization of tissue samples in pathology, machine learning models trained on these images have transformed our ability to detect and classify diseases [1]. However, the fundamental challenge of capturing high-quality microscopic images at speed remains unsolved. Although tissue samples are sliced into micrometer-thin sections, due to the fine focal length of microscopic lenses, they retain complex 3D morphological structures that create continuously varying optimal focal planes during the scanning process [2]. Capturing images in the optimal focal plane is critical to achieving high-quality and detailed images, and neglecting this leads to image quality degradation that can severely impact diagnostic accuracy and increase healthcare costs through repeated scans or potential misdiagnoses [3].

Auto-focus methods can be broadly classified into two categories: traditional and learning-based. Contrast-based traditional methods assess focus through gradient-based algorithms [4,5,6,7] —all requiring time-intensive capture of complete focal stacks. Later, optimization-based approaches [8] approximated the Brenner gradient with Lorentzian functions using sparse focal positions, while others [9,10] employed Gaussian models. Yet these methods falter in noisy environments due to local maxima issues. Phase-based methods leveraged dual-pixel sensors for disparity-based depth computation [11], but struggled with the fundamental complexity of depth-disparity modeling [12]. Deep learning approaches have transformed auto-focusing in pathology by leveraging CNNs. Early methods focused on single-domain feature extraction. Wei et al. [13] pioneered CNN-based focus prediction for time-lapse cell microscopy, while Jiang et al. [14] introduced autocorrelation in their residual architecture. Recent approaches [15,16] have adapted CNN architectures from computer vision, such as MobileNet. These learning-based methods [17] significantly outperform traditional approaches in speed and noise robustness, though they often struggle to generalize across tissue types and staining protocols. Recent developments on improving generalization propose a sample-invariant CNN scoring function [18], Kernel Distillation using paired samples for training but single-shot inference [19], or tackling focused image reconstruction under incoherent lighting conditions [20].

Automating conventional microscopes for pathology remains a significant challenge, with most existing work focusing on narrow, specialized applications. Chow et al. [21] enabled multi-region mosaic imaging for multi-photon microscopes, while [22] developed automation for micro-injection procedures. Li et al. [23] implemented a three-motor system for parasite detection, but relied on proprietary software, limiting reproducibility and customization. Collins et al. [24] contributed to accessibility through 3D-printed frameworks for motor integration, yet left the crucial auto-focus challenge unaddressed.

This work proposes DeepAf (**Deep Autofocus**)[1], a novel deep learning-based auto-focus framework based on the learnable features from Spatial and Spectral encoders in a regression model to predict the optimal focal distance without additional stacking or multiple views. We show that a single out-of-focus image contains enough information to infer the optimal focus due to the relationship between defocus and frequency domain characteristics. Defocusing creates unique signatures in the cut-off frequency and spectral distribution, which correlate

[1] **Project Page**: `https://deepautofocus.github.io/`

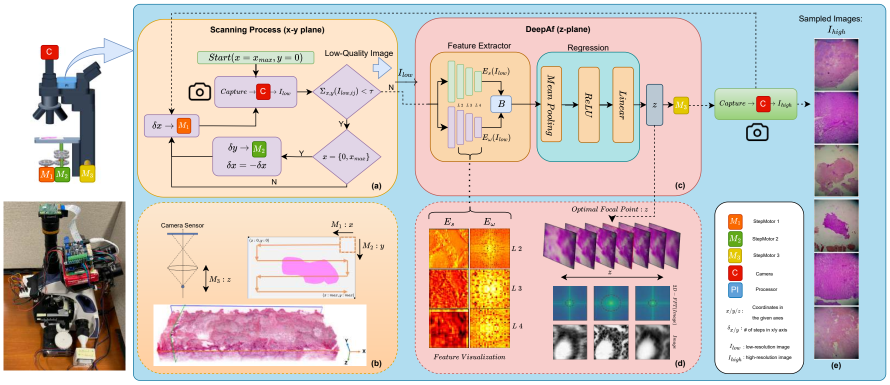

Fig. 1: **Left:** Implemented prototype and the schematic of the microscopic setup. Step motors M_1, M_2, M_3 control slide movement using processor PI along x, y, z respectively (a), with M_3 controlling the focus position (b). At each step, a low-resolution image I_{low} is captured by the camera (C), and the non-empty images are fed to the *DeepAf* network to adjust the focus (c). Finally, the high resolution images I_{high} are obtained with the correct focus (d).

with the defocus distance. Based on the observations that the cut-off frequency of a Fourier-transformed image indicates its distance to the optimal focal plane [14], we developed an automated microscopy system with a hybrid architecture inspired by Y-Net [25]. Our model simultaneously learns features from both spatial and frequency domains, enabling better generalization across different tissue types. This approach requires only a single input image to predict the optimal focus point, significantly reducing focusing time compared to traditional methods. We developed an efficient motorized microscopic design based on a conventional manual microscope [26]. Figure 1 depicts the auto-scanning setup of the microscope. Our framework incorporates DeepAf to efficiently capture high-quality images from histopathology slides. For clinical validation, we used the framework to create a dataset of brain tissue samples and demonstrated its effectiveness through automated cancer classification. This validation shows the potential of our approach for real-world clinical applications. In summary, our key contributions are: (1) DeepAf: A novel single-shot auto-focus approach effectively utilizing features from the spatial and spectral encoder in a regression model, demonstrating superior generalization across diverse tissue types and staining protocols in a single-shot without additional views or resampling, (2) An open-source, automated microscopy system that transforms conventional microscopes into efficient slide scanners, maintaining optimal focus throughout the imaging process, (3) A comprehensive dataset of brain tissue samples captured using our system, showcasing its ability to consistently produce high-quality digital pathology images, (4) Experimental validation through automated cancer classification, demonstrating the system's practical utility in clinical diagnostics.

2 Method

2.1 Microscopic System

The digitization pipeline consists of four sequential stages: (1) Slide Scanning: Following DICOM standards, scanning begins from the bottom right corner and follows a predefined trajectory (Figure 1-b). The system acquires low-resolution (1280×720) images for initial tissue detection, optimizing computational efficiency. (2) Tissue Detection: Simple thresholding to skip empty spaces is widely used in histopathology image analysis, such as Otsu thresholding [27]. Similarly, we apply thresholding in the HSV color space to filter empty regions:

$$\frac{1}{N}\sum_{i=1}^{N} V_i > \tau$$

where V_i represents pixel values in the Value channel of HSV and τ is the threshold value. (3) Auto-focus: Our spatiospectral network performs single-shot focus prediction, efficiently determining the optimal focal plane for the detected tissue region. The model's compact size enables real-time inference on the Raspberry Pi CPU. (4) High-Resolution Capture: The system captures the final image at 4056×3040 resolution at the predicted focal position $\hat{z}$. The scanning step size in the x-y plane critically impacts system performance. Small steps increase overlap and scanning time, while large steps risk missing tissue regions. Optimal step size selection balances coverage completeness with scanning efficiency.

Our automated microscopy system is built around a SWIFT-380t microscope [26]. The system consists of three key components: (1) A motorized stage with three stepper motors: M1 and M2 control x-y slide positioning, and M3 adjusts focus in the z-direction with 0.002mm precision. Custom 3D-printed gears translate motor motion to the microscope bed. (2) A dual-mode camera mounted directly above the objective lens eliminates parallax correction requirements. The camera operates at either low-resolution (I_{low}) for rapid tissue detection or high-resolution (I_{high}) for final image capture. (3) A Raspberry Pi control system with dual motor control HAT (Hardware Attached Top) coordinates motor movements and image acquisition based on auto-focus predictions.

2.2 DeepAf

Let $\mathcal{M} = (X, Y, Z)$ define the microscope coordinate system where Z represents the focus direction. For an input image $I \in \mathbb{R}^{H\times W\times C}$, the system comprises: A focus prediction function $f : \mathbb{R}^{H\times W\times C} \rightarrow \mathbb{R}$ defined as: $f(I) = R(B(E_s(I), E_\omega(I)))$, where E_s, E_ω are spatial and spectral encoders respectively, B is a bottleneck layer, and R is the regression head. A motorized control system with precision $\delta_z$mm adjusts the focal plane according to $f(I)$. The system captures images at position $(x, y) \in X \times Y$ with focus position $\hat{z} = f(I)$ to maintain optimal focus throughout scanning (*cf.* Figure 1-a). The dual encoder architecture (Figure 1-c) consists of: 1) Spatial Encoder: $E_s : \mathbb{R}^{H\times W\times C} \rightarrow$

$\mathbb{R}^{h\times w\times d_s}$ following U-Net architecture [28] extracting spatial features via four hierarchical convolutional layers. 2) Spectral Encoder: $E_\omega : \mathbb{R}^{H\times W\times C} \rightarrow \mathbb{R}^{h\times w\times d_\omega}$ comprising four sequential FFC blocks [29] that extract frequency domain features. The concatenated features $[E_s(I), E_\omega(I)] \in \mathbb{R}^{h\times w\times (d_s+d_\omega)}$ are processed by bottleneck B, followed by a regression head R consisting of 2D average pooling and a linear layer to predict the optimal focus position $\hat{z}$ along the Z axis of the microscope coordinate system. Unlike previous approaches such as [14], our architecture learns to extract both spatial and spectral features directly from the input image, eliminating preprocessing overhead. The modular design enables independent evaluation of spatial and spectral contributions to focus prediction. We train the network using smooth $\mathcal{L}1$ loss to handle the extensive range of focal values while maintaining gradient stability [30].

As demonstrated in Figure 1-d, in the 2D FFT image, the power spectrum analysis reveals distinct patterns between in-focus and out-of-focus images. Out-of-focus images significantly attenuate low-frequency components near the spectrum's center, while in-focus images display enhanced low-frequency amplification and higher cut-off frequencies, as also observed in [14]. This consistent relationship between focus quality and frequency distribution suggests that spectral features could provide robust indicators for auto-focus systems, potentially offering better generalization across different tissue types and imaging conditions.

3 Experiments

Datasets To evaluate our method, we employ the Incoherent dataset by Jiang *et al.* [14], which comprises two distinct test sets. The first test set contains tissue samples prepared by the same lab as the training data, while the second one includes specimens from a different lab. The two test sets exhibit significant differences in their color distribution, which serves as a good indicator of the generalization capabilities of the models. All images were captured at magnification level 20× with a depth of field (DoF) of 1μ. For both datasets, during training, all images are divided into tiles of size 224×224 pixels. Furthermore, for the case study of our microscopic system, we train the auto-focus network on 406 brain tissue samples from the Brigham and Women's Hospital in Boston and the University of Pennsylvania. For each sample, we create a focal stack of 1000 slices such that 500 images are below and above the optimal focal plane. Here, we capture the tissue at magnification level 4× with a DoF of 60μ to ensure faster scanning times in the subsequent case study.

Training and Evaluation During training, we optimize the model on each individual patch with a total of 130K patches. For hyperparameter optimization, we further split the data into 80% training and 20% validation. The test data consists of 700 patches. During testing, the median of all patches of one image serves as the final prediction for evaluation. We train all models with a batch size of 32, a learning rate of 8e-4, a weight decay of 0.006, 100 epochs, and the Adam optimizer. Moreover, all models presented in this work have been trained with data augmentation, namely, channel-wise normalization, random erasing,

Table 1: Comparison of SOTA auto-focus methods to ours on incoherent dataset [14] test set.

Method	Params.	# of input images	FE ↓	
			Same protocol	Diff. protocol
Dastidar *et al.* [15]	3.5M	2	$0.19_{\pm 0.18}$	$\mathbf{0.25}_{\pm \mathbf{0.26}}$
Jiang *et al.* [14]	10.8M	1	$0.46_{\pm 0.34}$	$0.53_{\pm 0.59}$
Chen *et al.* [31]	4.2M	1	$0.21_{\pm 0.21}$	$0.44_{\pm 0.50}$
DeepAf Spatial (Ours)	4.7M	1	$\mathbf{0.18}_{\pm \mathbf{0.17}}$	$0.39_{\pm 0.50}$
DeepAf Spatiospectral (Ours)	4.2M	1	$\mathbf{0.18}_{\pm \mathbf{0.17}}$	$\underline{0.32_{\pm 0.36}}$

Gaussian blur, random perspective, random auto contrast, and color jittering. We report the model performance as focus error (FE) computed by the mean absolute error between the predicted and optimal focal distance and its standard deviation.

3.1 Autofocus Results

Comparison to SOTA Table 1 shows our model's performance compared to previous auto-focus methods. All reported results from previous work are taken from the original publications. On the same protocol data, our spatiospectral model achieves the best overall focus error while using significantly fewer parameters compared to [14] and taking only one input image instead of two as in [15]. On the different protocol data, the spatiospectral network performs significantly better than [14] and only exhibits a slightly bigger focus error than reported by [15], outperforming the single-shot SOTA auto-focus model [31] by Chen *et al*. It is noteworthy to mention that our method only takes one input image while [15] takes two input images to predict the distance to the optimal focal plane. The choice of only taking one input image is motivated by higher inference times, which is critical to allow for efficient scanning of histopathology slides in a clinical setting. Our method shows promising generalization capabilities compared to previous single-shot auto-focus methods [14,31].

Ablation Study Table 2 shows the effect of different components in our network on the open dataset (20× magnification) [14] and our own data (4× magnification). Here, FD, DoF, and FE denote the predictions with false directions, depth of field, and focus error, respectively. For the public dataset, we can see that the spatiospectral and fully-spatial models achieve the same performance on the same protocol data. However, the spatiospectral network outperforms the fully-spatial one on the different protocol data, exhibiting a better generalization in different lab environments. Moreover, the fully-spectral encoder shows the highest error rate for both datasets. On the one hand, these observations indicate that including features from the spectral domain can enhance the generalization performance of the auto-focus. On the other hand, the network seems to learn essential features from the spatial domain, as evidenced by the good performance

Table 2: **Ablation Study.** Focus Error comparison between different encoders on the Incoherent dataset (same and different protocols) [14] and our curated dataset. Spatial: Two spatial encoders, Spectral: Two spectral encoders, Spatiospectral: One spatial and one spectral encoder.

		FD ↓		DoF ↑		FE ↓		
Network	Params.	Same 20× Magn.	Diff. 20× Magn.	Same 20× Magn.	Diff. 20× Magn.	Same 20× Magn.	Diff. 20× Magn.	Ours 4× Magn.
Spatial	4.7M	0.86%	**1.22%**	89.24%	71.34%	**0.18**±**0.17**	0.39±0.50	**5.40**±**6.06**
Spectral	3.6M	1.29%	3.12%	73.03%	53.81%	0.29±0.26	0.46±0.39	9.24±9.33
Spatiospectral	4.2M	**0.72%**	1.60%	**89.81%**	**73.78%**	**0.18**±**0.17**	**0.32**±**0.36**	6.67±7.51

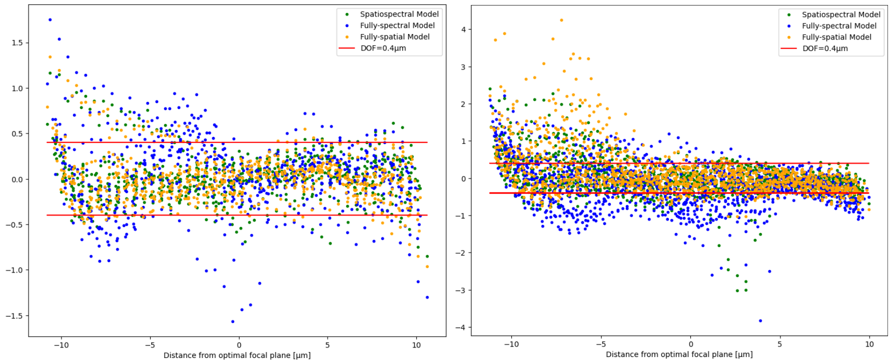

Fig. 2: **Focal Distance from Optimal Focal Point.** Left: data from the same protocol, Right: data from the different protocol.

of the fully-spatial model for the same protocol data and the overall bad performance of the fully-spectral model. In the case of our data, the fully spatial model slightly outperforms the spatiospectral network. Since we only have data from one lab for this dataset, we cannot check for the generalization capabilities of the individual models. We also see that the spatiospectral model only predicts 0.72% of the cases in the wrong focus direction (above or below the optimal focal plane), indicating that it can solve the focus ambiguity problem with just one input image. Moreover, approximately 90% of all predictions lie inside the DoF. From a practical point of view, this implies that at least 90% of the captured images based on this auto-focus model appear visibly sharp to the human eye. We also show a visualization of the focal error distribution given different distances from the optimal focal point using the data from the same and different protocols in Figure 2 for different models. As can be seen, while the spatiospectral model generally handles the different distances from the same protocol data well, the fully spatial model can generalize better to data with a different protocol. We assume that this can be due to the larger changes in the frequency domain between different data distributions.

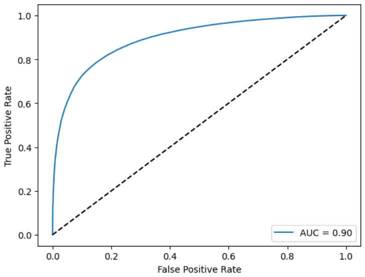

Fig. 3: **Classification Performance.** Receiver operating curve for the binary brain cancer classifier with an AUC score of 0.9.

3.2 Case study: Brain Tissue Slides Scanning and Classification

Brain Tissue Dataset With the proposed scanning strategy and utilizing the spatiospectral auto-focus model, we scanned 536 histopathology slides containing brain tissue from Brigham and Women's Hospital in Boston. The tissue samples comprise four different cancer subtypes, namely high-grade glioma, low-grade glioma, inflammatory, and normal tissue. As proof of concept, we used the 4× magnification objective lens of the microscope to have a bigger FoV and, thus, faster scanning times. The average scanning time for each slide is approximately 400 seconds, depending on the amount of tissue present. Figure 1-e illustrates some qualitative results of the acquired images.

Brain Cancer Classification We use the captured images to train a binary classification model. We define the high and low-grade glioma classes as the "cancer" class, while the inflammatory and normal samples are defined as the "normal" class. As seen in Figure 1-e, due to the high magnification rate, the images exhibit large radial distortions at the edges. Thus, each image is center-cropped to a size of 2000×3000 pixels. During training, each tile and its corresponding label are considered individually, while at inference, the global mean of all tiles belonging to one tissue sample is computed as the final prediction.

Results The binary classifier achieves an AUC score of 0.90 and an F1 score of 0.83 in the 5-fold cross-validation setting. Figure 3 shows the corresponding ROC curve. It is noteworthy to mention that these results were achieved with just a magnification level of 4×. We do not compare this result with previous works since they rely on WSIs, which are scanned at a magnification of 20×, capturing much more detail of the tissue. Nevertheless, this result indicates the high quality of the generated images of our microscopic system and their relevance and validity for automated cancer diagnosis in a clinical setting.

4 Conclusion

In this work, we developed an automated robotic microscopic system for scanning histopathology glass slides while maintaining the optimal focus position during the process. A key component of this system is our deep auto-focus model, which shows superior generalization performance by only taking one input image. In a large study using 536 brain tissue samples, we successfully tested our proposed microscopic system by training a brain cancer classifier on the generated images. The results of this case study show our system's potential to automate diagnostic tasks in pathology and support pathologists in their work. We believe that the developed microscopic design could potentially pave the way for the democratization of high-precision diagnosis in resource-constrained settings while maintaining diagnostic quality comparable to traditional high-end equipment.

Acknowledgments We thank Alexander Bruce for his assistance with slide scanning at the Digital Imaging Facility, Department of Pathology, Brigham and Women's Hospital, and Kristina C. Grieco, Raquel Arias-Camison, Nina Thakur, Dominique Ballinger, and Alexa Craig for their assistance in collecting the pathology slides. We thank Mariam Kapanadze and Faith McDonald for their administrative support.

Disclosure of Interests. This research was partially supported by the Munich Center of Machine Learning, BaCaTeC, and the EU BigPicture project. K.-H.Y. is supported in part by the National Institute of General Medical Sciences grant R35GM142879, the National Heart, Lung, and Blood Institute grant R01HL174679, the Department of Defense Peer Reviewed Cancer Research Program Career Development Award HT9425-231-0523, the Research Scholar Grant RSG-24-1253761-01-ESED (grant DOI: `https://doi.org/10.53354/ACS.RSG-24-1253761-01-ESED.pc.gr.193749`) from the American Cancer Society, and the Harvard Medical School Dean's Innovation Award.

References

1. B. Bejnordi, M. Veta, P. van Diest, . Van Ginneken, L. Latonen, P. Ruusuvuori, K. Liimatainen, and CAMELYON16 Consortium, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA Neurology*, vol. 318, pp. 2199–2210, Dec. 2017.
2. G. Li, S. E. Fox, B. Summa, B. Hu, C. Wenk, A. Akmatbekov, J. L. Harbert, R. S. V. Heide, and J. Q. Brown, "Multiscale 3-dimensional pathology findings of covid-19 diseased lung using high-resolution cleared tissue microscopy," *bioRxiv*, 2020.
3. B. Markiefka, A. Pryalukhin, W. Hulla, A. Bychkov, J. Fukuoka, A. Madabhushi, V. Achter, L. Nieroda, R. Büttner, A. Quaas, and Y. Tolkach, "Quality control stress test for deep learning-based diagnostic model in digital pathology," *Modern Pathology*, vol. 34, 06 2021.
4. T. Yeo, S. Ong, R. Sinniah, *et al.*, "Autofocusing for tissue microscopy," *Image and vision computing*, vol. 11, no. 10, pp. 629–639, 1993.
5. A. Santos, C. Ortiz De Solórzano, J. J. Vaquero, J. M. Pena, N. Malpica, and F. del Pozo, "Evaluation of autofocus functions in molecular cytogenetic analysis," *Journal of microscopy*, vol. 188, no. 3, pp. 264–272, 1997.

6. J. F. Brenner, B. S. Dew, J. B. Horton, T. King, P. W. Neurath, and W. D. Selles, "An automated microscope for cytologic research a preliminary evaluation.," *Journal of Histochemistry & Cytochemistry*, vol. 24, no. 1, pp. 100–111, 1976.
7. M. Subbarao, T.-S. Choi, and A. Nikzad, "Focusing techniques," *Optical Engineering*, vol. 32, no. 11, pp. 2824–2836, 1993.
8. S. Yazdanfar, K. B. Kenny, K. Tasimi, A. D. Corwin, E. L. Dixon, and R. J. Filkins, "Simple and robust image-based autofocusing for digital microscopy," *Optics express*, vol. 16, no. 12, pp. 8670–8677, 2008.
9. S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 16, no. 8, pp. 824–831, 1994.
10. Z. Wang, M. Lei, B. Yao, Y. Cai, Y. Liang, Y. Yang, X. Yang, H. Li, and D. Xiong, "Compact multi-band fluorescent microscope with an electrically tunable lens for autofocusing," *Biomedical optics express*, vol. 6, no. 11, pp. 4353–4364, 2015.
11. C. Herrmann, R. S. Bowen, N. Wadhwa, R. Garg, Q. He, J. T. Barron, and R. Zabih, "Learning to autofocus."
12. R. Garg, N. Wadhwa, S. Ansari, and J. T. Barron, "Learning single camera depth estimation using dual-pixels," 2019.
13. L. Wei and E. Roberts, "Neural network control of focal position during time-lapse microscopy of cells," *Scientific reports*, vol. 8, no. 1, p. 7313, 2018.
14. S. Jiang, J. Liao, Z. Bian, K. Guo, Y. Zhang, and G. Zheng, "Transform-and multi-domain deep learning for single-frame rapid autofocusing in whole slide imaging," *Biomedical optics express*, vol. 9, no. 4, pp. 1601–1612, 2018.
15. T. R. Dastidar and R. Ethirajan, "Whole slide imaging system using deep learning-based automated focusing," *Biomedical Optics Express*, vol. 11, no. 1, pp. 480–491, 2020.
16. J. Liao, X. Chen, G. Ding, P. Dong, H. Ye, H. Wang, Y. Zhang, and J. Yao, "Deep learning-based single-shot autofocus method for digital microscopy," vol. 13, no. 1, pp. 314–327.
17. Y. Gan, Z. Ye, Y. Han, Y. Ma, C. Li, Q. Liu, W. Liu, C. Kuang, and X. Liu, "Single-shot autofocusing in light sheet fluorescence microscopy with multiplexed structured illumination and deep learning," *Optics and Lasers in Engineering*, vol. 168, p. 107663, 2023.
18. A. Shajkofci and M. Liebling, "Deepfocus: A few-shot microscope slide auto-focus using a sample invariant cnn-based sharpness function," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 164–168, 2020.
19. Y. Gu, "Single-shot focus estimation for microscopy imaging with kernel distillation," *IEEE Transactions on Computational Imaging*, vol. 9, pp. 542–550, 2023.
20. H. Ding, F. Li, Z. Meng, S. Feng, J. Ma, S. Nie, and C. Yuan, "Auto-focusing and quantitative phase imaging using deep learning for the incoherent illumination microscopy system," *Opt. Express*, vol. 29, pp. 26385–26403, Aug 2021.
21. S. K. Chow, H. Hakozaki, D. L. Price, N. A. MacLean, T. J. Deerinck, J. C. Bouwer, M. E. Martone, S. T. Peltier, and M. H. Ellisman, "Automated microscopy system for mosaic acquisition and processing," *Journal of microscopy*, vol. 222, no. 2, pp. 76–84, 2006.
22. T. Aoyama, S. Takeno, K. Hano, M. Takasu, M. Takeuchi, and Y. Hasegawa, "View-expansive microscope system with real-time high-resolution imaging for simplified microinjection experiments," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12961–12966, IEEE, 2021.
23. Y. Li, R. Zheng, Y. Wu, K. Chu, Q. Xu, M. Sun, and Z. J. Smith, "A low-cost, automated parasite diagnostic system via a portable, robotic microscope and deep learning," *Journal of biophotonics*, vol. 12, no. 9, p. e201800410, 2019.

24. J. T. Collins, J. Knapper, J. Stirling, J. Mduda, C. Mkindi, V. Mayagaya, G. A. Mwakajinga, P. T. Nyakyi, V. L. Sanga, D. Carbery, L. White, S. Dale, Z. J. Lim, J. J. Baumberg, P. Cicuta, S. McDermott, B. Vodenicharski, and R. Bowman, "Robotic microscopy for everyone: the openflexure microscope," *Biomed. Opt. Express*, vol. 11, pp. 2447–2460, May 2020.
25. A. Farshad, Y. Yeganeh, P. Gehlbach, and N. Navab, "Y-net: A spatiospectral dual-encoder network for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 582–592, Springer, 2022.
26. "Swift sw 380t microscope." `https://swiftoptical.com/uploads/pdf/SW380TManual.pdf`. Accessed: 2024-02-28.
27. P. Bandi, O. Geessink, Q. Manson, M. Van Dijk, M. Balkenhol, M. Hermsen, B. E. Bejnordi, B. Lee, K. Paeng, A. Zhong, *et al.*, "From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 550–560, 2018.
28. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.
29. L. Chi, B. Jiang, and Y. Mu, "Fast fourier convolution," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4479–4488, 2020.
30. R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.
31. W. Chen, X. Shen, and Z. Wang, "Microscope autofocus based on difference of gaussians and triplet loss," in *2024 International Conference on Cyber-Physical Social Intelligence (ICCSI)*, pp. 1–6, IEEE, 2024.