

Z-STACK SCANNING CAN IMPROVE AI DETECTION OF MITOSIS: A CASE STUDY OF MENINGIOMAS

Hongyan Gu¹, Ellie Onstott², Wenzhong Yan¹, Tengyou Xu¹, Ruolin Wang¹, Zida Wu¹

Xiang ‘Anthony’ Chen^{1*}, Mohammad Haeri^{2*}

¹ Department of Electrical and Computer Engineering, University of California, Los Angeles

² Department of Pathology and Laboratory Medicine, University of Kansas Medical Center

ABSTRACT

Z-stack scanning is an emerging whole slide imaging technology that captures multiple focal planes alongside the z-axis of a glass slide. Because z-stacking can offer enhanced depth information compared to the single-layer whole slide imaging, this technology can be particularly useful in analyzing small-scaled histopathological patterns. However, its actual clinical impact remains debated with mixed results. To clarify this, we investigate the effect of z-stack scanning on artificial intelligence (AI) mitosis detection of meningiomas. With the same set of 22 Hematoxylin and Eosin meningioma glass slides scanned by three different digital pathology scanners, we tested the performance of three AI pipelines on both single-layer and z-stacked whole slide images (WSIs). Results showed that in all scanner-AI combinations, z-stacked WSIs significantly increased AI’s sensitivity (+17.14%) on the mitosis detection with only a marginal impact on precision. Our findings provide quantitative evidence that highlights z-stack scanning as a promising technique for AI mitosis detection, paving the way for more reliable AI-assisted pathology workflows, which can ultimately benefit patient management.

Index Terms— Z-stack scanning, mitosis detection, Artificial Intelligence, meningiomas

1. INTRODUCTION

In digital pathology, the “z-stack” is a multi-planar scanning technology that captures various focal planes of glass slides along the “z” axis (*i.e.*, perpendicular to the slide surface plane) [1]. Different from traditional “single-layer” scanning with a single focus plane, z-stack scanning preserves more detailed specimen information from multiple planes of focal depths, which, inevitably, also introduces longer scanning time and larger file sizes [2]. To date, the z-stack feature has been available in various digital pathology scanners [1] and has been popularly employed in the analysis of cytopathology smears [3].

Interestingly, there are mixed opinions on whether z-stack scanning can enhance pathology analysis. On one hand, multiple previous works have indicated that the z-stack can improve pathologists’ judgment and AI performance. For instance, Kalinski *et al.* reported the number of focal planes was positively related to pathologists’ correctness in classifying *Helicobacter pylori* [4]. Kim *et al.* showed that z-stack enhanced pathologists’ evaluation of high-grade urothelial carcinomas [5]. Nurzynska *et al.* observed significant performance improvements of AI in identifying low-burden acid-fast mycobacteria on z-stacked WSIs [6]. On the other hand, Sturm *et al.* reported that z-stack scanning did not improve pathologists’ diagnostic accuracy in classifying melanocytic lesions or detecting dermal mitosis [2].

We believe that the discrepancies in prior works’ results are due to two main reasons: (1) variations in the z-stack scanning algorithms from different vendors, and (2) differences in the specimens selected for each study. Therefore, this work aims to provide a more systematic evaluation of AI performance on z-stack vs. single-layer WSIs with three scanners and three AI pipelines on the same set of glass slides. Here, we conduct a case study of mitosis detection in meningiomas. The mitosis is a small-sized ($\sim 10\mu m$, see Figure 2) pattern that is critical in meningioma grading [7]. Despite its importance, mitosis assessment by pathologists is usually time-consuming and suffers from low concordance [8]. In response, recent years have seen the advancement of deep learning models for mitosis detection [9, 10], with some capable of augmenting pathologists in this task [11, 12, 13].

22 meningioma Hematoxylin and Eosin (H&E) slides with 6,350 mitoses were digitized using both single-layer and z-stack settings with three scanners from three vendors. We determined the performance of three deep learning pipelines on these slides by measuring the sensitivity and precision. In all scanner-AI combinations, the deep learning pipelines achieved significantly higher sensitivity, with an average improvement of 17.14% (single-layer: 0.601, z-stack: 0.704). Meanwhile, the impact on precision from z-stack scanning was marginal (single-layer: 0.753, z-stack: 0.757).

* Correspondence: xac@ucla.edu (X.A.C.), mhaeri@kumc.edu (M.H.)

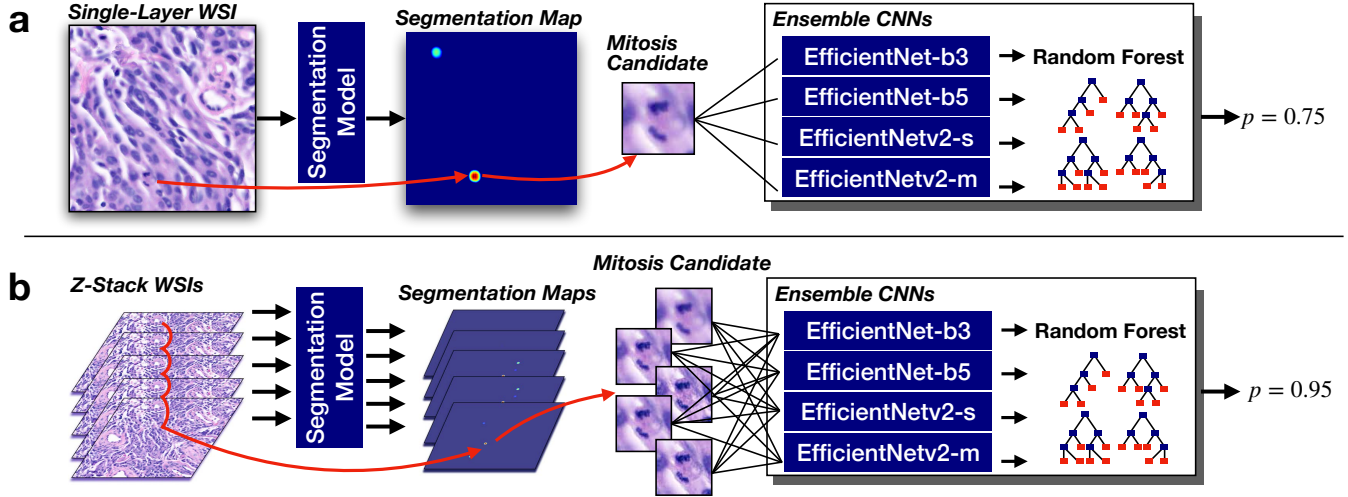


Fig. 1: Deep learning mitosis detection pipeline for (a) single-layer WSIs and (b) z-stacked WSIs.

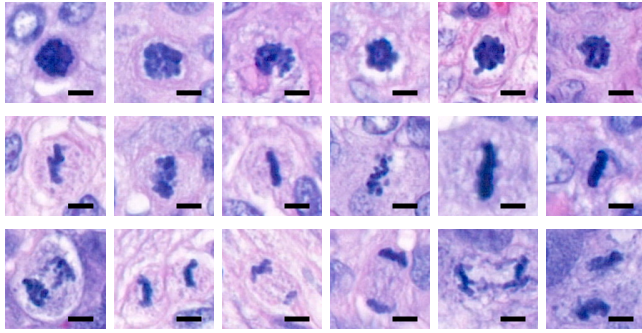


Fig. 2: Examples of mitoses annotated on WSIs from the Panoramic 250 scanner (+0.0 μm focus plane), bar=5 μm .

2. MATERIALS AND METHODS

2.1. Specimen Collection and Mitosis Annotation

22 de-identified H&E meningioma slides were collected from the University of Kansas Medical Center. These slides were initially scanned by Panoramic 250 scanner (3DHISTECH, Hungary) with 41 \times objective (0.121 μm per pixel, hereafter mpp) and z-stack (five planes: -1.2 μm , -0.6 μm , +0.0 μm , +0.6 μm , +1.2 μm) setting. Mitosis annotation was performed on the z-stacked WSIs. Two pathology trainees individually screened the 22 slides and provided the preliminary mitosis annotations. Next, a third neuropathologist reviewed and finalized the annotations. In total, 6,350 mitoses were annotated. Examples of annotated mitoses are shown in Figure 2.

2.2. Slide Scanning

We selected three digital pathology scanners with z-stack features: (1) Panoramic 480DX (3DHISTECH, Hungary,

hereafter P480DX), (2) Aperio GT 450 (Leica, Germany, hereafter GT 450), and (3) AxioScan 7 (Zeiss, Germany). For each scanner, the glass slides were scanned into WSIs with both single-layer and z-stack settings specified in Table 1. The resulting WSIs were exported to the bigTIFF format with JPEG compression (90% quality) using the software provided by the vendor. These WSIs were then rescaled to approximately 0.25 mpp to match the resolution of 40 \times . A two-stage registration process [8] translated the locations of ground truth mitoses acquired from Section 2.1 to the WSIs.

Table 1: Settings used for single-layer and z-stack scanning, mpp: μm per pixel, WI: water immersion.

Scanner	Scanning Settings			
	Z Planes	Objective	Resolution	Interplane Distance
P480DX	1	41 \times , WI	0.121 mpp	N/A
	5	41 \times , WI	0.121 mpp	0.6 μm
GT 450	1	40 \times , Air	0.263 mpp	N/A
	5	40 \times , Air	0.263 mpp	0.75 μm
AxioScan 7	1	40 \times , Air	0.086 mpp	N/A
	5	40 \times , Air	0.086 mpp	0.6 μm

2.3. Deep Learning Inferencing Pipeline

We used a two-stage deep learning-based inferencing pipeline similar to [10] (Figure 1(a)): (1) a segmentation model first selects mitosis candidates, followed by (2) an ensemble of four convolutional neural networks (CNNs) to verify these candidates. The pipeline supports three types of segmentation models: PSPNet [14], Segformer [15], and DeeplabV3+ [16]. The ensemble method was a random forest classi-

fier that predicts based on the output of four CNN models: EfficientNet-b3, EfficientNet-b5 [17], EfficientNetv2-s, and EfficientNetv2-m [18]. This deep learning pipeline was trained using four mitosis datasets of human and animal specimens (MIDOG++ [9], MITOS_WSLCMC [19], MITOS_WSLCCMCT [20], meningioma mitosis [8]). The training set included 38,634 mitoses in total. Note that all slides in the training set were scanned in the single layer.

We further customized the deep learning pipeline for z-stacked WSIs (Figure 1(b)). First, the segmentation model was applied to each layer of the z-stacked WSIs. Then, mitosis candidates from all z-planes were combined. If the distance between any two candidates in the combined set was less than 10 pixels ($2.5 \mu m$), they were considered duplicates and merged. Next, for each mitosis candidate on every layer, four CNN models were applied, resulting in $5 \text{ (z-planes)} \times 4 \text{ (CNNs)} = 20$ predictions. Finally, these 20 predictions were passed to the random forest classifier to predict a mitosis probability.

Considering the variation in imaging quality across different scanners (see Figure 3), for each scanner, we selected one same slide (including 411 mitoses) from 22 slides in Section 2.2 to re-calibrate the random forest classifier. The re-calibration process was performed separately on both single-layer and z-stacked WSIs. We used the remaining 21 WSIs (including 5,939 mitoses) as the test set to evaluate the performance of the mitosis detection pipeline. The inferencing was repeated 20 times for performance evaluation.

2.4. Measures and Statistics

The performance of deep learning pipelines was measured by the sensitivity and precision [10]. For all combinations (*i.e.*, $3 \text{ (scanners)} \times 3 \text{ (deep learning pipelines)} = 9$ conditions per metric) of scanners and deep learning pipelines, we conducted a one-way ANOVA followed by TukeyHSD test to compare each metric on single-layer *vs.* z-stacked WSIs. The mean values were calculated by the bootstrapping method (10,000 times with replacement).

3. RESULTS

The total file size for single-layer WSIs was 87.02 GB. As a comparison, z-stacked WSIs had 418.92 GB, which is $\sim 3.81 \times$ larger. Examples of mitoses extracted from the three scanners are shown in Figure 3.

Table 2 shows the sensitivity results: in all nine conditions, there was a significant increase in AI sensitivity. On average, sensitivity of deep learning improved from 0.601 on the single-layer WSIs to 0.704 on the z-stacked WSIs (+17.14%).

¹We used the same ensemble method for three pipelines. Therefore, the name of the segmentation model was used to represent the pipeline in tables 2 and 3.

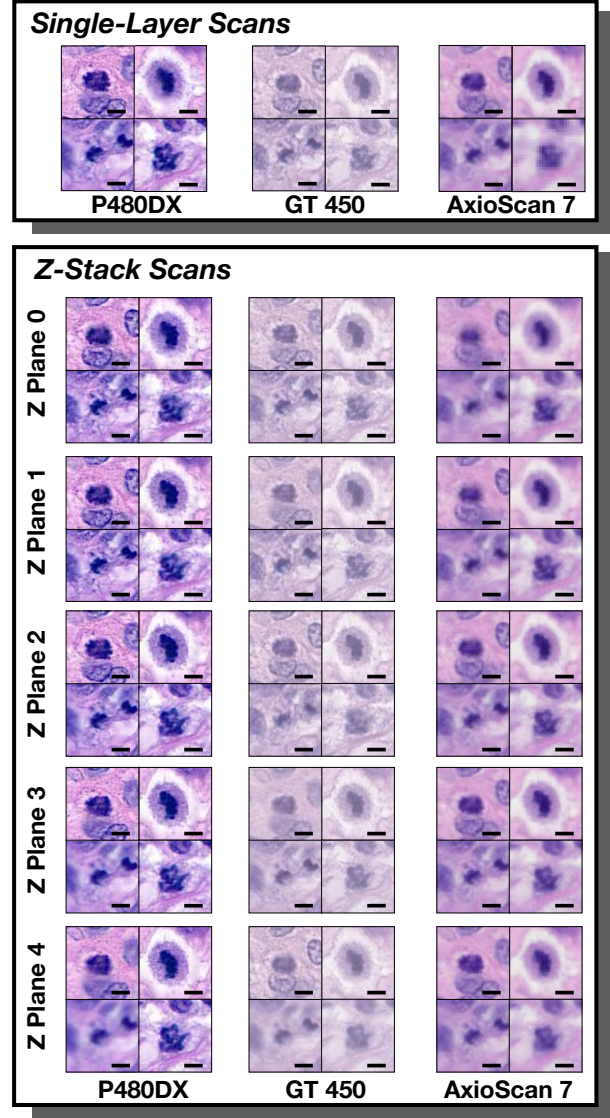


Fig. 3: Examples of mitoses under the single-layer and z-stack scanning with three scanners, bar= $5 \mu m$.

The highest improvement was observed with the PSPNet segmentation on AxioScan 7 WSIs, where its sensitivity was increased from 0.398 to 0.554 (+39.24%). The highest sensitivity was achieved with the DeepLabV3+ segmentation on GT 450 WSIs, with an improvement from 0.681 in the single-layer to 0.773 in the z-stack (+13.52%).

Table 3 presents the precision results: across the nine conditions, statistical significance was not observed in six. Two conditions under the P480DX WSIs showed significant improvements with the Segformer (+3.91%) and DeepLabV3+ (+9.70%) segmentation. However, one condition using GT 450 scanner – Segformer segmentation showed a significant precision decrease, from 0.770 to 0.710 (-7.86%). On average, deep learning achieved a precision of 0.753 on single-layer WSIs and 0.757 on z-stacked WSIs. The highest precision was recorded in the Segformer segmentation on the Ax-

Table 2: Sensitivity of three deep learning pipelines on WSIs from three scanners¹.

Scanner	Segmentation	Single-Layer	Z-Stack	Δ	p
P480DX	PSPNet	0.633	0.726	+14.74%	<0.001
	Segformer	0.664	0.726	+9.32%	<0.001
	DeepLabV3+	0.675	0.717	+6.23%	<0.001
GT 450	PSPNet	0.636	0.739	+16.33%	<0.001
	Segformer	0.616	0.740	+20.09%	<0.001
	DeepLabV3+	0.681	0.773	+13.52%	<0.001
AxioScan 7	PSPNet	0.398	0.554	+39.24%	<0.001
	Segformer	0.447	0.574	+28.28%	<0.001
	DeepLabV3+	0.583	0.713	+22.25%	<0.001
Average		0.601	0.704	+17.14%	N/A

Table 3: Precision of three deep learning pipelines on WSIs from three scanners.

Scanner	Segmentation	Single-Layer	Z-Stack	Δ	p
P480DX	PSPNet	0.768	0.787	+2.46%	0.999
	Segformer	0.763	0.793	+3.91%	<0.001
	DeepLabV3+	0.720	0.790	+9.70%	<0.001
GT 450	PSPNet	0.714	0.735	+3.04%	0.957
	Segformer	0.770	0.710	-7.86%	<0.001
	DeepLabV3+	0.729	0.704	-3.46%	0.825
AxioScan 7	PSPNet	0.815	0.802	-1.59%	0.999
	Segformer	0.845	0.834	-1.31%	0.999
	DeepLabV3+	0.739	0.729	-1.30%	0.999
Average		0.753	0.757	+0.53%	N/A

ioScan 7 scanner, with precision of 0.845 for single-layer and 0.834 for z-stack.

4. DISCUSSION & CONCLUSION

A major gap between the light microscope and regular single-layer whole slide imaging is that pathologists can adjust the fine focus of the microscope, whereas single-layer WSI has only one plane. Z-stack scanning bridges this gap by providing extra z-level information, which has the potential to enhance histology analysis on small-scaled features, cytology, and non-FFPE hematopathology. This study presents the first quantitative evidence that AI can achieve significantly higher sensitivity with only marginal impact on the precision in detecting mitoses in meningiomas. Such improvement is both scanner- and AI-agnostic. Future studies can validate whether the improvement is generalizable by testing on multiple histological patterns (*e.g.*, *H. pylori*) and mitoses on more diverse

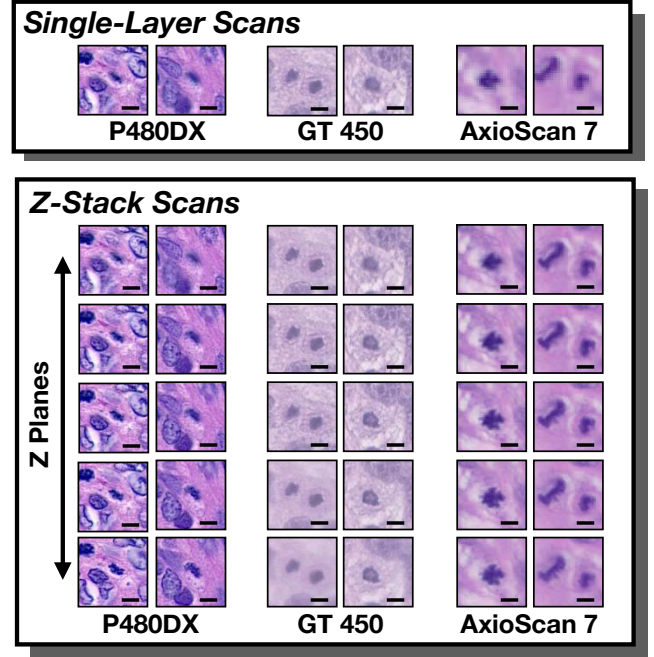


Fig. 4: Examples of mitoses missed by the deep learning pipeline with DeepLabV3+ segmentation model on single-layer WSIs but were captured under z-stacked WSIs, bar=5 μ m.

tumor sites (*e.g.*, breast cancer [10]).

Compared to regular single-layer scanning, z-stack scanning provides enhanced focus control by covering a broader range of depth information, resulting images with higher quality. For instance, as shown in Figure 4, while single-layer WSIs do not exhibit significant out-of-focus issues, z-stack technique captures more nuanced chromosomal features, and might enhance the deep learning performance thereafter.

Meanwhile, it is noteworthy that the file size of z-stacked WSIs increases linearly with the number of z-planes, which brings challenges to storage and file management. More recently, the “extended focus” algorithm, or multi-focus image fusion technology, can collapse z-layers into a single layer, which can potentially reduce the size of z-stacked WSIs without compromising the AI performance. Therefore, we suggest future research explore the optimal setting to balance the file size and the imaging quality – including the number of planes, the interplane distance, as well as the compression settings – while taking the scanner hardware into consideration.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. Ethical approval was exempted as no identifiable human information was involved in this study.

6. ACKNOWLEDGMENTS

This study is supported in part by the National Science Foundation CAREER 2047297 awarded to Xiang ‘Anthony’ Chen and the start up fund award from the University of Kansas Medical Center to Mohammad Haeri. The authors would like to thank Chunxu Yang and Dr. Sallam Alrosan for their help in the data collection process. The authors have no relevant financial or non-financial interests to disclose.

7. REFERENCES

- [1] Ankush Patel, Ulysses GJ Balis, Jerome Cheng, Zaibo Li, Giovanni Lujan, David S McClintock, et al., “Contemporary whole slide imaging devices and their applications within the modern pathology department: a selected hardware review,” *Journal of Pathology Informatics*, vol. 12, no. 1, pp. 50, 2021.
- [2] Bart Sturm, David Creytens, Martin G Cook, Jan Smits, Marcory CRF van Dijk, Erik Eijken, et al., “Validation of whole-slide digitally imaged melanocytic lesions: Does z-stack scanning improve diagnostic accuracy?,” *Journal of Pathology Informatics*, vol. 10, no. 1, pp. 6, 2019.
- [3] Odille Bongaerts, Paul J van Diest, Math Pieters, and Marius Nap, “Working toward consensus among professionals in the identification of classical cervical cytomorphological characteristics in whole slide images,” *Journal of Pathology Informatics*, vol. 6, no. 1, pp. 52, 2015.
- [4] Thomas Kalinski, Ralf Zwönitzer, Saadettin Sel, Matthias Evert, Thomas Guenther, Harald Hofmann, et al., “Virtual 3d microscopy using multiplane whole slide images in diagnostic pathology,” *American journal of clinical pathology*, vol. 130, no. 2, pp. 259–264, 2008.
- [5] David Kim, Robert Burkhardt, Susan A Alperstein, Hamza N Gokozan, Abha Goyal, Jonas J Heymann, et al., “Evaluating the role of z-stack to improve the morphologic evaluation of urine cytology whole slide images for high-grade urothelial carcinoma: results and review of a pilot study,” *Cancer cytopathology*, vol. 130, no. 8, pp. 630–639, 2022.
- [6] Karolina Nurzynska, Dalin Li, Ann E Walts, and Arkadiusz Gertych, “Multilayer outperforms single-layer slide scanning in ai-based classification of whole slide images with low-burden acid-fast mycobacteria (afb),” *Computer Methods and Programs in Biomedicine*, vol. 234, pp. 107518, 2023.
- [7] David N Louis, Arie Perry, Pieter Wesseling, Daniel J Brat, Ian A Cree, Dominique Figarella-Branger, et al., “The 2021 who classification of tumors of the central nervous system: a summary,” *Neuro-oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.
- [8] Hongyan Gu, Chunxu Yang, Issa Al-Kharouf, Shino Magaki, Nelli Lakis, Christopher Kazu Williams, et al., “Enhancing mitosis quantification and detection in meningiomas with computational digital pathology,” *Acta Neuropathologica Communications*, vol. 12, no. 1, pp. 7, 2024.
- [9] Marc Aubreville, Frauke Wilm, Nikolas Stathonikos, Katharina Breining, Taryn A Donovan, Samir Jabari, et al., “A comprehensive multi-domain dataset for mitotic figure detection,” *Scientific data*, vol. 10, no. 1, pp. 484, 2023.
- [10] Marc Aubreville, Nikolas Stathonikos, Taryn A Donovan, Robert Klopffleisch, Jonas Ammeling, Jonathan Ganz, et al., “Domain generalization across tumor types, laboratories, and species—insights from the 2022 edition of the mitosis domain generalization challenge,” *Medical Image Analysis*, vol. 94, pp. 103155, 2024.
- [11] Hongyan Gu, Yuan Liang, Yifan Xu, Christopher Kazu Williams, Shino Magaki, Negar Khanlou, et al., “Improving workflow integration with xpath: design and evaluation of a human-ai diagnosis system in pathology,” *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 2, pp. 1–37, 2023.
- [12] Hongyan Gu, Chunxu Yang, Mohammad Haeri, Jing Wang, Shirley Tang, Wenzhong Yan, et al., “Augmenting pathologists with navipath: design and evaluation of a human-ai collaborative navigation system,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–19.
- [13] Christof A Bertram, Marc Aubreville, Taryn A Donovan, Alexander Bartel, Frauke Wilm, Christian Marzahl, et al., “Computer-assisted mitotic count using a deep learning-based algorithm improves interobserver reproducibility and accuracy,” *Veterinary pathology*, vol. 59, no. 2, pp. 211–226, 2022.
- [14] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.
- [15] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo, “Segformer: Simple and efficient design for semantic segmentation with transformers,” *Advances in neural information processing systems*, vol. 34, pp. 12077–12090, 2021.
- [16] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [17] Mingxing Tan, “Efficientnet: Rethinking model scaling for convolutional neural networks,” *arXiv preprint arXiv:1905.11946*, 2019.
- [18] Mingxing Tan and Quoc Le, “Efficientnetv2: Smaller models and faster training,” in *International conference on machine learning*. PMLR, 2021, pp. 10096–10106.
- [19] Marc Aubreville, Christof A Bertram, Taryn A Donovan, Christian Marzahl, Andreas Maier, and Robert Klopffleisch, “A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research,” *Scientific data*, vol. 7, no. 1, pp. 417, 2020.
- [20] Christof A Bertram, Marc Aubreville, Christian Marzahl, Andreas Maier, and Robert Klopffleisch, “A large-scale dataset for mitotic figure assessment on whole slide images of canine cutaneous mast cell tumor,” *Scientific data*, vol. 6, no. 1, pp. 274, 2019.