

Robust Multiresolution and Multistain Background Segmentation in Whole Slide Images

Artur Jurgas^{1,2}, Marek Wodzinski^{1,2}, Manfredo Atzori^{1,3}, and Henning Müller^{1,4}

¹ University of Applied Sciences Western Switzerland (HES-SO Valais), Information Systems Institute, Sierre, Switzerland

² Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, Krakow, Poland

³ Department of Neuroscience, University of Padova, Padova, Italy

⁴ Medical Faculty, University of Geneva, Geneva, Switzerland

Abstract. Background segmentation is an important step in analysis of histopathological images. It allows one to remove irrelevant regions and focus on the tissue of interest. However, background segmentation is challenging due to the variability of stain colors and intensity levels across different images, modalities, and magnification levels. In this paper, we present a learning-based model for histopathology background segmentation based on convolutional neural networks. We compare two multiresolution approaches to deal with the variability of magnification in histopathology images: (i) model that uses upscaling of smaller patches of the image, and (ii) model simultaneously trained on multiple resolution levels. Our model is characterized by solid performance both in multiresolution and multistain dyes (H&E and IHC), achieving good performance on publicly available dataset. The quantitative scores are, in terms of the Dice score, close to 94.71. The qualitative analysis presents strong performance on previously unseen cases from different distributions and various dyes. We freely release the model, weights, and ground-truth annotations to promote the open science and reproducible research.

Keywords: Computational Pathology, Deep Learning, Digital Pathology Segmentation, Whole-Slide Images, WSI

1 Introduction

Background segmentation is a basic step in most preprocessing tasks for whole slide images (WSIs), which are digital scans of tissue slides used for cancer diagnosis and prognosis. Segmentation aims to separate the foreground tissue regions from the background glass regions, which can reduce computational costs and improve accuracy for subsequent analysis such as classification, detection, grading, and registration [10]. However, background segmentation of WSIs is challenging due to variations in tissue appearance, staining quality, illumination conditions,

and scanning artifacts. Existing methods for background segmentation of WSIs are either based on handcrafted features or deep learning models. Up to our knowledge, the currently existing methods are limited to H&E (hematoxylin and eosin) staining. Moreover, most existing methods for background segmentation are either not publicly available or require manual tuning of parameters for different datasets.

Many studies have explored histopathology segmentation, which is a technique to identify different regions in tissue images like [5,1,13] However, most of these studies focus on segmenting nuclei, which are small and distinct structures in the tissue. Segmenting the whole tissue is also challenging because it involves large areas that have low contrast and that are often similar to the background, especially when immunohistochemistry (IHC) staining is used. This type of staining is much less normalized and quality controlled [7,17]. Even for H&E staining, the differences in dyes are problematic [16]. Therefore, existing methods for segmentation cannot be easily adapted to whole tissue segmentation.

Some previous studies [12] used a conventional method to segment tissue regions of interest (ROI) from histopathological images. However, this method works well only for images stained with H&E, and not for those stained with immunohistochemistry (IHC). Moreover, our goal is different from theirs. We want to segment the background from the tissue, not just the ROI within the tissue. This is important for some preprocessing techniques that require selecting the entire tissue area, including its folds and artifacts.

Recent studies have demonstrated that deep learning can achieve remarkable results in histopathology, enabling more accurate and detailed predictions for various diseases [6,11,15,13] However, most of the existing work focuses on specific types of tissue, such as epidermal tissue [14]. While there is some research in background segmentation [2], the proposed models are not publicly available or accessible, limiting their reproducibility and applicability.

In this paper, we propose a deep learning-based pipeline for background segmentation of WSIs that can handle diverse types of tissues and stains without requiring any prior information or user intervention. Our framework consists of two main components: a patch-level segmentation network that predicts foreground probability maps for small patches extracted from WSIs, and a slide-level fusion during inference that combines the patch-level predictions into a final binary mask for the whole slide. We evaluate our framework on a public dataset from ACROBAT challenge, covering mainly breast cancer patients. We show that our framework achieves solid performance and high generalizability across different tissues. Additionally, we show solid performance in multiple resolutions of those images, as well as both H&E and IHC staining.

One of the applications of our background segmentation method is to improve the quality and efficiency of other algorithms that process histopathology images. Software tools like QuPath [3] and HistoQC [8] are widely used for various tasks such as tissue detection, annotation, classification, and quantification and are proven to benefit histopathology analysis [4]. However, these tools often rely on manual or semi-automatic methods to remove the background regions from

the images, which can be time-consuming and inconsistent. Our background segmentation method can help streamline this process.

The source code, ground-truth segmentation masks, and model weights will be made publicly available [9].

2 Materials and Methods

2.1 Dataset

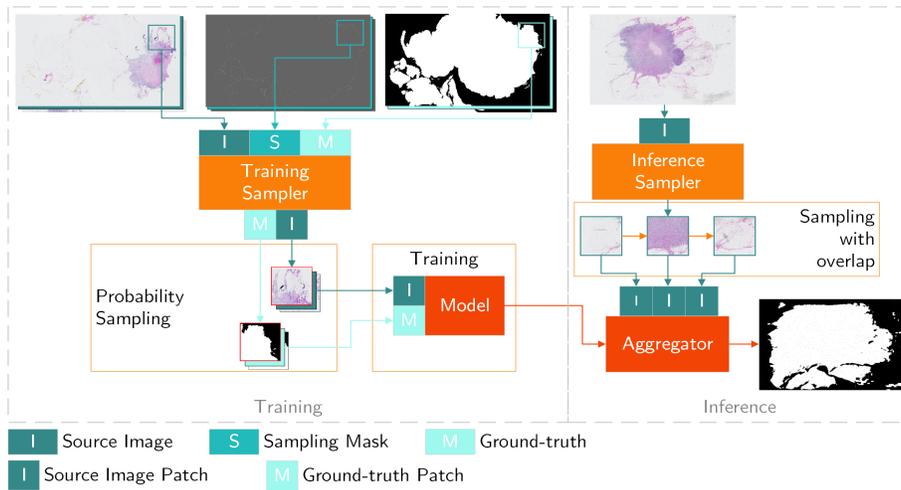


Fig. 1. A schematic diagram of the proposed deep learning pipeline. The pipeline consists of two stages: training and inference. Each block has marked inputs described in the legend.

The ground-truth used for training is defined by manual segmentation of nine different tissues from the ACROBAT dataset [18]. The dataset is a collection of tissues in a pyramidal tiff format with varying resolutions at each level of the pyramid. Starting at 10x, authors provide 7 to 9 lower resolutions per image, with a downsampling factor of 2 in between. The second level which made up most of the training data has an average resolution of 12146 ± 2189 pixels along the X axis, and 24007 ± 4854 pixels along the Y axis. Each slide at this resolution can potentially generate up to 4449 unique, non-overlapping patches of size 256 by 256 pixels. We utilize a random sampling strategy. In each iteration, we sample 256 patches in random locations. This means over the course of training, there is some overlap of those patches. The final number of patches is not defined and depends on how long we train the network.

The dataset includes various artifacts such as markers, scratches, out-of-focus regions and coverslips. It also consists of two types of staining: H&E and IHC. We use three distinct IHC dyes during training and testing. We selected this dataset because it represents the real-world challenges of image analysis. We used eight images for training, one image for validation, and four images each in its two staining variants for qualitative evaluation without prior manual segmentations. We aimed to achieve high-quality segmentation by using a patch-based approach for both training and inference. This allows us to exploit the heterogeneity of the tissue structures across different slides and to generalize better with less data.

We used Nvidia Tesla V100 graphics card configured with 300W TDP and 32 GB of memory hosted on PLGrid HPC cluster Prometheus. We did not utilize the tensor cores as of the time of writing, which means the inference times in Table 2 could be further improved.

2.2 Model

We present an encoder-decoder Convolutional Neural Network (CNN) as our model for image segmentation in Fig 2. The model architecture and the training pipeline are illustrated in Fig 1. We trained our model using binary cross-entropy (BCE) loss and Adam optimizer, and we measured its performance on both training and validation datasets using soft Dice score.

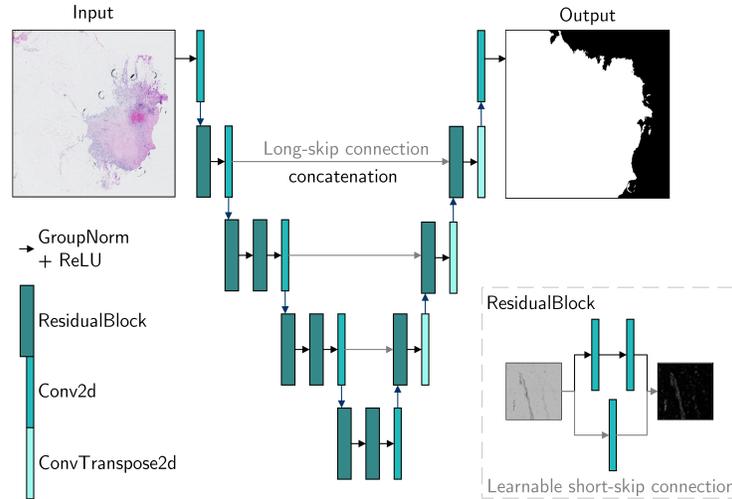


Fig. 2. A schematic illustration of the UNet-like model used in this study. The model consists of an encoder-decoder architecture with both short skip connections described in ResidualBlock and long skip connections based on concatenating feature maps.

Table 1. Models’ performance during training.

Dataset patches	Model	Dice
Training	Multiresolution	96.14
Validation	Multiresolution	92.27
Training	Upscaled	96.31
Validation	Upscaled	91.68

We adopted a UNet-like architecture that consists of an encoder-decoder structure with both short and long skip connections. The short skip connections allow the network to preserve spatial information across different levels of abstraction, while the long skip connections enable the network to recover fine details from the encoder output. We present the architecture details of the model in Table 3.

Table 2. Models’ performance on validation image on different resolution levels.

Pyramid level	<i>Multiresolution model</i>		<i>Upscaled model</i>	
	Dice	Latency [mm:ss]	Dice	Latency [mm:ss]
2nd	93.66	1:08	94.43	1:06
3rd	95.09	0:16	93.21	0:49
4th	95.61	0:03	92.85	0:44
5th	94.62	0:01	81.00	0:47
6th	94.57	0:00	11.16	0:45
Mean Dice	94.71		74.53	

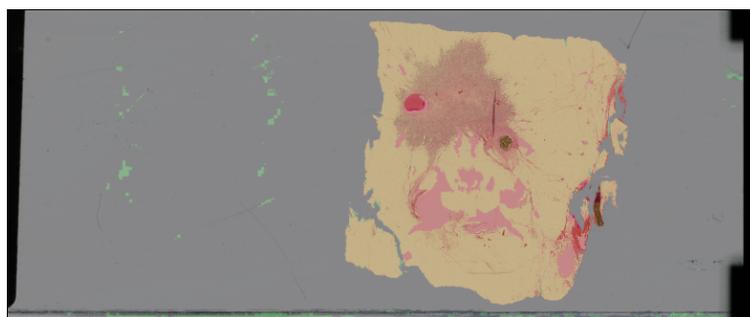
Our training pipeline involves extracting image patches of size 256×256 from: (i) different resolutions of the image pyramid, or (ii) only the second level of the pyramid, and then inferring on other levels by extracting smaller patches proportional to the downsampling level and upscaling them using bilinear interpolation to 256×256 pixels.

Our random sampling strategy is dependent on previously generated probability maps. We apply gaussian blurring with a 3×3 kernel on segmentation maps and then subtract that from the original mask. This gives us a blurred edge of the segmentation mask. Then we transform this output so that the background and foreground of the tissue has a 25% sampling probability and the edge 50%.

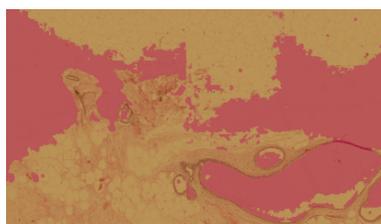
During the validation and inference steps, we employ an aggregator of individual patches by merging them together with a 20% overlap. In this overlap region, we average the value of pixels.

Table 3. Model architecture with each convolution kernel size, the output shape of the layer, and the layer’s number of parameters.

Layer type	Kernel Shape	Output Shape	Param #
UNet	–	[1, 1, 256, 256]	–
+ Sequential	–	[1, 1, 128, 128]	–
+ Conv2d	[4, 4]	[1, 1, 128, 128]	17
+ GroupNorm	–	[1, 1, 128, 128]	2
+ LeakyReLU	–	[1, 1, 128, 128]	–
+ Sequential	–	[1, 32, 64, 64]	–
+ ResidualBlock	[3, 3]	[1, 32, 128, 128]	9,760
+ Conv2d	[4, 4]	[1, 32, 64, 64]	16,416
+ GroupNorm	–	[1, 32, 64, 64]	64
+ LeakyReLU	–	[1, 32, 64, 64]	–
+ Sequential	–	[1, 64, 32, 32]	–
+ ResidualBlock	[3, 3]	[1, 64, 64, 64]	57,792
+ ResidualBlock	[3, 3]	[1, 64, 64, 64]	78,272
+ Conv2d	[4, 4]	[1, 64, 32, 32]	65,600
+ GroupNorm	–	[1, 64, 32, 32]	128
+ LeakyReLU	–	[1, 64, 32, 32]	–
+ Sequential	–	[1, 128, 16, 16]	–
+ ResidualBlock	[3, 3]	[1, 128, 32, 32]	230,272
+ ResidualBlock	[3, 3]	[1, 128, 32, 32]	312,192
+ Conv2d	[4, 4]	[1, 128, 16, 16]	262,272
+ GroupNorm	–	[1, 128, 16, 16]	256
+ LeakyReLU	–	[1, 128, 16, 16]	–
+ Sequential	–	[1, 128, 32, 32]	–
+ ResidualBlock	[3, 3]	[1, 128, 16, 16]	312,192
+ ResidualBlock	[3, 3]	[1, 128, 16, 16]	312,192
+ ConvTranspose2d	[4, 4]	[1, 128, 32, 32]	262,272
+ GroupNorm	–	[1, 128, 32, 32]	256
+ LeakyReLU	–	[1, 128, 32, 32]	–
+ Sequential	–	[1, 64, 64, 64]	–
+ ResidualBlock	[3, 3]	[1, 64, 32, 32]	160,192
+ ConvTranspose2d	[4, 4]	[1, 64, 64, 64]	65,600
+ GroupNorm	–	[1, 64, 64, 64]	128
+ LeakyReLU	–	[1, 64, 64, 64]	–
+ Sequential	–	[1, 32, 128, 128]	–
+ ResidualBlock	[3, 3]	[1, 32, 64, 64]	40,160
+ ConvTranspose2d	[4, 4]	[1, 32, 128, 128]	16,416
+ GroupNorm	–	[1, 32, 128, 128]	64
+ LeakyReLU	–	[1, 32, 128, 128]	–
+ Sequential	–	[1, 1, 256, 256]	–
+ ResidualBlock	[3, 3]	[1, 1, 128, 128]	346
+ ConvTranspose2d	[4, 4]	[1, 1, 256, 256]	17
+ GroupNorm	–	[1, 1, 256, 256]	2
+ LeakyReLU	–	[1, 1, 256, 256]	–
+ Sequential	–	[1, 1, 256, 256]	–
+ Conv2d	[1, 1]	[1, 1, 256, 256]	2



(a) Full image



(b) Example of segmentation details



(c) Example of failure on out-of-focus regions

Fig. 3. Visualization of segmentation on the validation dataset. Segmentation output is marked with green color. Ground-truth masks are marked with red color. Overlap of the prediction with ground-truth is marked with orange color.

3 Results

The Figures 3,4,5 illustrate some examples of our method’s output. We compared the results of the upscaled and multiresolution model in Table 1 and Table 2. The scores reported in those tables are the average scores of all patches in each batch.

Results in Table 1 indicated that the multiresolution model had a similar performance on the training dataset but a slightly higher performance on the validation dataset than the single-resolution model. This indicates that the multiresolution model reduced overfitting and improved generalization. Fig 4 and Fig 5 illustrate some examples of image reconstruction from both models on the test set.

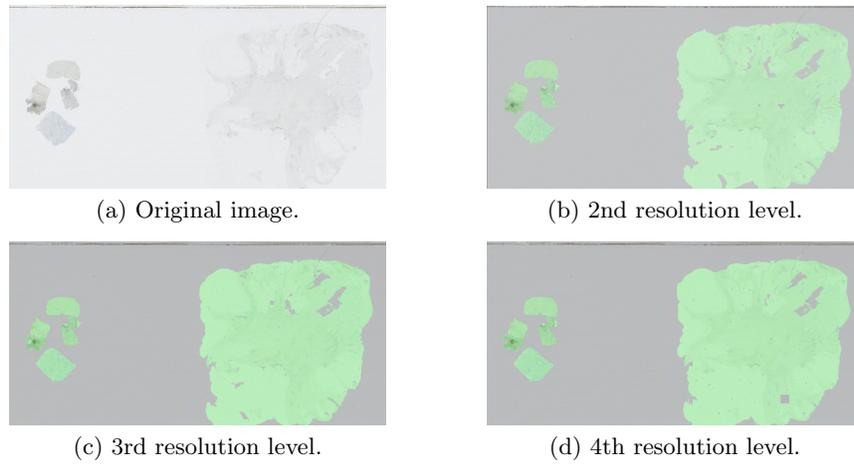


Fig. 4. Exemplary visualization from the test dataset (IHC staining). Segmentation output is marked with green color.

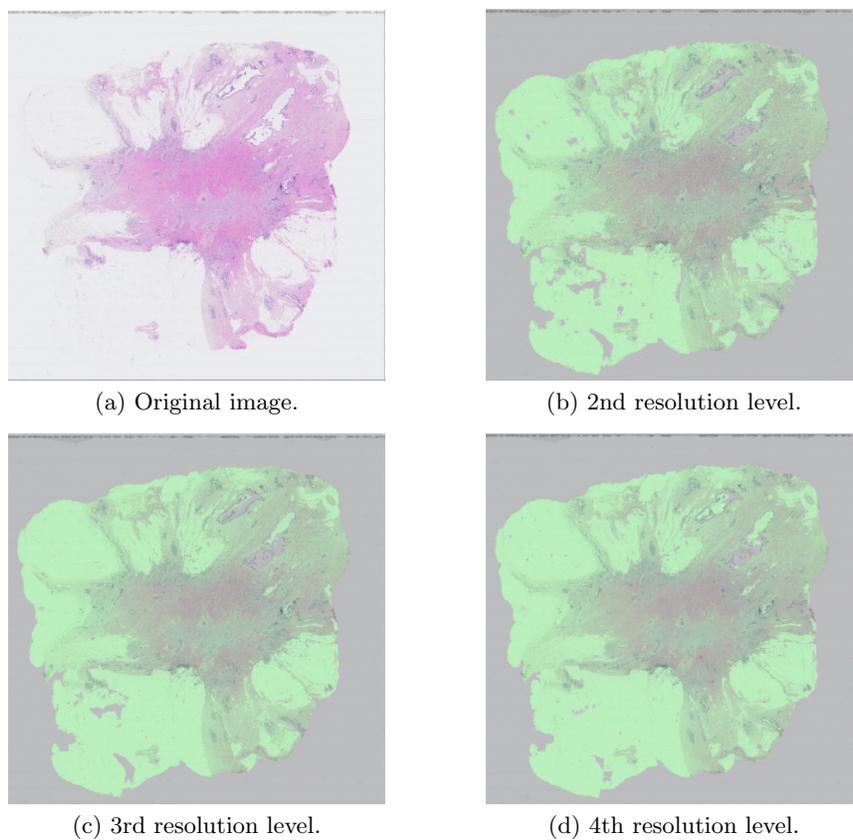


Fig. 5. Exemplary visualization from the test dataset (H&E staining). Segmentation output is marked with green color.

Table 4. Qualitative analysis of the model’s output in the test set. Each case’s number matches the number in the original dataset. The images will be published as the supplementary material in the associated repository.

Case from Dataset	Output Commentary
4	Small artifacts from aggregation in blurry regions in H&E variant. They disappear while going down the magnification levels. On the 3rd level, artifacts from scratches appear. IHC variant has no artifacts, even though they are present in the image.
5	While correctly segmenting tissue, there is a larger artifact area in the HE variant. It disappears on lower resolution levels. IHC variant is segmented similarly to H&E one.
8	Solid performance on both HE and IHC variants. A little noise in the largest IHC sample where coverslip is present in the original, though most of the coverslip is properly not segmented by the model. The artifact disappears at lower magnifications.
14	Solid HE segmentation. Similar performance in IHC, although some noise is present on the highest resolution in less visible regions of the tissue.
29	Solid HE segmentation. The coverslip on the IHC variant has been partly segmented.
30	Solid HE and IHC segmentation. The coverslip on the IHC variant was properly not segmented.

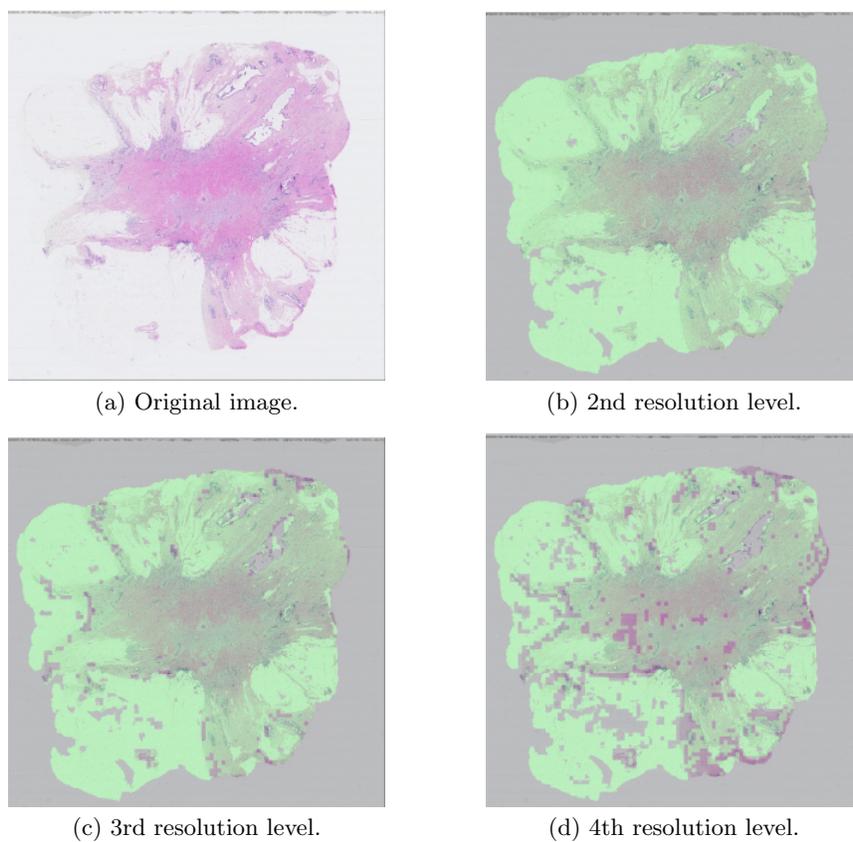


Fig. 6. Exemplary visualization from the test dataset (H&E staining) on the upscaling model. Segmentation output is marked with green color.

4 Discussion and Conclusions

We proposed two methods for patch-based segmentation of histopathology images: (i) training and testing on a single pyramid level with patch resampling at inference time, and (ii) training and testing on multiple pyramid levels without patch resampling. We found that the second method achieved better performance across different resolutions and significantly reduced the inference time. Our method is a general approach for histopathology slide analysis that does not target any specific regions of tissues. Instead, it aims to segment the background of the slide by inverting the segmentation of other structures, such as cells, nuclei, glands, vessels, etc.

According to Table 2 multiresolution model achieved substantially better performance on lower magnification levels, which is supported by comparing visualizations in Fig 5 and Fig 6. The multiresolution model had also a faster inference time than the single-resolution model, which is beneficial for practical applications.

The second method required a larger model with a wider receptive field, but it is still feasible to deploy on even IoT class hardware such as Nvidia Jetson. This allows for deployment on modern WSIs digitization hardware or integration into most computational histopathology software.

Additionally, the first method incurred additional computational costs due to two interpolation steps. The first method suffered from information loss and artifact introduction when upscaling patches from lower resolutions. The difference in performance between the two methods can be intuitively explained by considering that the second method used more information from the original image and did not introduce any artificial structures due to patches interpolation.

A common problem in histopathology is the variability of staining techniques and dyes used to visualize different tissue structures and biomarkers. Different laboratories may use different protocols, reagents, and equipment to perform immunohistochemistry staining, which can result in inconsistent and non-reproducible results. This poses a challenge for developing computational models that can analyze histopathological images and extract meaningful features from them. Our model addresses this challenge by being stain-invariant, meaning that it can work on images stained with a wide range of different dyes without requiring any pre-processing or normalization steps. This makes our model more robust and generalizable to different datasets and applications. Segmentations of the same tissue dyed by IHC and HE staining is presented in Fig 4 and Fig 5.

One of the limitations of our method is the sensitivity of the model to the resolution of the input images. Our model was trained on images with a fixed set of magnifications, and it may not perform well on images with a very different magnification level. Another limitation is the possibility of large or significantly different artifacts than in our dataset influencing the segmentation results. Particularly with higher resolutions, such artifacts could be segmented as part of the background, leading to inaccurate or incomplete segmentation.

In this paper, we proposed a novel method for fast and generalizable background segmentation in histopathological images. We tackled the problem from two perspectives: an upscaling of lower resolution images, and training the model on all magnification levels. We demonstrated that our method can achieve high accuracy and robustness on various tissues with different resolution levels and staining dyes. Compared to other works like [12] our method does not require adjusting any hyperparameters. We also showed that our deep learning model can learn to segment tissues effectively even with a small amount of data, thanks to the patch-based approach. Our method can be useful for preprocessing histopathological images for further analysis and diagnosis.

Acknowledgments. This work was done as a part of the IMI BigPicture project (IMI945358). We gratefully acknowledge Poland’s high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2023/016239.

References

1. Al-Kofahi, Y., Lassoued, W., Lee, W., et al.: Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images. *IEEE Transactions on Biomedical Engineering* **57**(4) (April 2010) 841–852
2. Bándi, P., Balkenhol, M., van Ginneken, B., et al.: Resolution-agnostic tissue segmentation in whole-slide histopathology images with convolutional neural networks. *PeerJ* **7** (2019) e8242
3. Bankhead, P., Loughrey, M.B., Fernández, J.A., et al.: QuPath: Open source software for digital pathology image analysis. *Scientific Reports* **7**(1) (December 2017) 16878
4. Chen, Y., Zee, J., Smith, A., et al.: Assessment of a computerized quantitative quality control tool for whole slide images of kidney biopsies. *The Journal of pathology* **253**(3) (March 2021) 268–278
5. Cui, Y., Zhang, G., Liu, Z., et al.: A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images. *Medical & Biological Engineering & Computing* **57**(9) (September 2019) 2027–2043
6. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., et al.: Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* **318**(22) (December 2017) 2199–2210
7. Elias, J.M., Gown, A.M., Nakamura, R.M., et al.: Special Report: Quality Control in Immunohistochemistry: Report of a Workshop Sponsored by the Biological Stain Commission. *American Journal of Clinical Pathology* **92**(6) (December 1989) 836–843
8. Janowczyk, A., Zuo, R., Gilmore, H., et al.: HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. *JCO clinical cancer informatics* **3** (April 2019) 1–7
9. Jurgas, A.: Jarartur/pcbbe23-histseg: Multiresolution and Multistain Background Segmentation in WSIs (May 2023)
10. Levy, J.J., Jackson, C.R., Haudenschild, C.C., et al.: PathFlow-MixMatch for Whole Slide Image Registration: An Investigation of a Segment-Based Scalable Image Registration Method (March 2020)

11. Litjens, G., Sánchez, C.I., Timofeeva, N., et al.: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports* **6** (May 2016) 26286
12. Muñoz-Aguirre, M., Ntasis, V.F., Rojas, S., et al.: PyHIST: A Histological Image Segmentation Tool. *PLOS Computational Biology* **16**(10) (October 2020) e1008349
13. Naylor, P., Laé, M., Reyal, F., et al.: Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map. *IEEE Transactions on Medical Imaging* **38**(2) (February 2019) 448–459
14. Oskal, K.R.J., Risdal, M., Janssen, E.A.M., et al.: A U-net based approach to epidermal tissue segmentation in whole slide histopathological images. *SN Applied Sciences* **1**(7) (June 2019) 672
15. Shelhamer, E., Long, J., Darrell, T.: Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(4) (April 2017) 640–651
16. Tellez, D., Litjens, G., Bándi, P., et al.: Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis* **58** (December 2019) 101544
17. Tsutsumi, Y.: Pitfalls and Caveats in Applying Chromogenic Immunostaining to Histopathological Diagnosis. *Cells* **10**(6) (June 2021) 1501
18. Weitz, P., Valkonen, M., Solorzano, L., et al.: ACROBAT - a multi-stain breast cancer histological whole-slide-image data set from routine diagnostics for computational pathology. *ArXiv* **abs/2211.13621** (2022)