2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871996

A Cascaded Deep Learning Framework for Segmentation of Nuclei in Digital Histology Images*

Khadijeh Saednia, *Student Member, IEEE*, William T. Tran, and Ali Sadeghi-Naini, *Senior Member, IEEE*

Abstract— **Accurate segmentation of nuclei is an essential step in analysis of digital histology images for diagnostic and prognostic applications. Despite recent advances in automated frameworks for nuclei segmentation, this task is still challenging. Specifically, detecting small nuclei in large-scale histology images and delineating the border of touching nuclei accurately is a complicated task even for advanced deep neural networks. In this study, a cascaded deep learning framework is proposed to segment nuclei accurately in digitized microscopy images of histology slides. A U-Net based model with customized pixel-wised weighted loss function is adapted in the proposed framework, followed by a U-Net based model with VGG16 backbone and a soft Dice loss function. The model was pretrained on the Post-NAT-BRCA public dataset before training and independent evaluation on the MoNuSeg dataset. The cascaded model could outperform the other state-of-the-art models with an AJI of 0.72 and a F1-score of 0.83 on the MoNuSeg test set.**

Keywords— ***Digital histology images, Nuclei segmentation, Deep neural networks, Cascaded deep learning models***

I. Introduction

Cancer is a major cause of mortality worldwide, accounting for about 10 million deaths in 2020 [1], [2]. Breast cancer is the most prevalent malignancy and the second leading cause of cancer death in women [3]. Early stage diagnosis of breast cancer is effective in preserving treatment options, reducing costs, and improving the survival and quality of life for patients [4]. While standard clinical imaging such as mammography and ultrasound are widely used for breast cancer screening, a biopsy followed by histopathology analysis on the acquired specimens is the gold standard for definitive diagnosis of breast cancer [5]. Recent studies have been shown that quantitative features describing the morphology, distribution, and texture of the tumor nuclei on digital histology images can be used in machine learning frameworks for various diagnostic, tissue characterization and prognostic applications [6], [7]. Accurate nuclei segmentation is an essential step of extracting such features from the histopathology images [8], [9].

Recent advances in deep convolutional networks have led several researchers to propose data-driven frameworks for automated image segmentation [10]. The proposed techniques mostly comprise an encoder-decoder architecture [11]. The requirement for accurate segmentation in addition to the lack of labeled training data raises the importance of developing a task-specific framework for segmenting medical and biomedical images [12]. The U-Net architecture [13] that consists of two symmetric contracting and expanding paths, was specifically introduced for biomedical image segmentation. The proposed model could be trained with a relatively small training set, and it could outperform the other frameworks in the ISBI cell tracking challenge 2015 [14]. However, U-Net struggles with finding the exact border of nuclei, and in practice, the model cannot accurately segment the touching or very close nuclei.

The emergence of U-Net has opened pathways for variants of its architecture to enhance the performance of segmentation tasks in biomedical images. The attention U-Net was proposed to improve the segmentation performance by integrating grid-based attention gates (AGs) on top of U-Net architecture [15]. Models trained with the AGs learn to suppress unnecessary regions in an input image while highlighting important features for a particular task. Each skip connection's gating signal incorporates image features from multiple imaging scales to remove the requirement of having explicit tissue localization modules. The attention U-Net model has been originally proposed for segmenting pancreas in computed tomography (CT) abdominal images [15], but then adapted in other application including nuclei segmentation [16]. The U-Net++ which has a deeply supervised encoder-decoder architecture, is another proposed variant of U-Net to enhance the performance of biomedical image segmentation frameworks [17]. In the proposed model, the encoder and decoder sub-networks are linked through several nested dense skip connections. The redesigned skip paths aim to minimize the semantic gap between the feature maps of the encoder and decoder blocks in the network.

Due to the importance of nuclei segmentation in digital histology images, a dataset was introduced for the multi-organ nuclei segmentation (MoNuSeg) challenge in MICCAI 2018 [18]. Several studies have investigated various approaches to achieve the best result in this challenge. In [8], a contour-aware informative aggregation network (CIA-Net) was presented as an innovative deep neural network with a hierarchical information aggregation module between two task-specific decoders. The proposed architecture could enhance the generalization ability of the framework in segmenting the

*This research was supported by Natural Sciences and Engineering Research Council (NSERC) of Canada, Terry Fox Foundation, and the Lotte and John Hecht Memorial Foundation.

K. Saednia is with the Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada; also, with the Department of Radiation Oncology, Sunnybrook Health Sciences Centre, Toronto, ON, Canada (e-mail: saednia@yorku.ca)

W. T. Tran is with the Department of Radiation Oncology and Biological Sciences Platform, Sunnybrook Health Sciences Centre and University of Toronto, Toronto, ON, Canada.

A. Sadeghi-Naini is with the Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada; also, with the Department of Radiation Oncology and Physical Sciences Platform, Sunnybrook Health Sciences Centre, Toronto, ON, Canada (e-mail: asn@yorku.ca, phone: 416-736-2100 x20590).

unseen organ nuclei. In [12], a pretrained EfficientNet was applied as the backbone of U-Net++ architecture for the breast tumor nuclei segmentation. The adapted backbone resulted in having fewer parameters and outperformed previous convolutional neural networks (CNNs) in terms of efficiency and accuracy. In [19], a novel self-supervised learning framework was introduced that fully exploits the capacity of extensively employed CNNs. The proposed method entails two sub-tasks of scale-wise triplet learning and count ranking, which allow the networks to exploit prior knowledge about nucleus size and number to extract instance-aware feature representations from the raw information. Most of the previous studies have demonstrated suboptimal performance in finding the border of the nuclei accurately, and in practice, touching or very close nuclei are always challenging to be segmented.

In this study, a novel cascaded framework is investigated to overcome the limitations of previous models and improve the accuracy of nuclei segmentation in histology images. The framework consists of a weighted U-Net model followed by a U-Net architecture with a VGG16 backbone and a soft Dice loss function. The weighted pixel-wise masks were generated for the training data and utilized along with the binary masks in calculating the loss function for the weighted U-Net in order to train the model to classify the touching and very close nuclei more accurately. The obtained results demonstrated a considerably better performance of the proposed model compared to each of the single networks and the previously proposed frameworks.

II. MATERIALS AND METHODS

A. Dataset

The proposed framework was pretrained using the Post-NAT-BRCA dataset collected from the cancer imaging archive (TCIA) [20], and trained and evaluated using the multi-organ nuclei segmentation (MoNuSeg) dataset [18]. All the histology images were H&E stained and scanned at 40x magnification. The Post-NAT-BRCA dataset includes histology images of surgical tissue specimens acquired from breast cancer patients following neoadjuvant therapy (NAT). The nuclei in these images were annotated manually using the Sedeen software (Pathcore, Toronto, Canada). The dataset contains 92 histology images of various size, with a total of 25,675 annotated nuclei. The MoNuSeg dataset consists of thirty training histology images and fourteen test images with a size of 1000 × 1000 pixel, collected from seven organs (i.e., breast, liver, kidney, prostate, bladder, colon, stomach). The MoNuSeg training and test sets contain 21,623 and 7,223 annotated nuclei, respectively. In this study, the MoNuSeg training set was randomly split into the training (80%) and validation (20%) sets at patient level.

B. Data Preprocessing

All histology images in the MoNuSeg dataset were zero padded to a size of 1024 × 1024 pixel (12 padded pixels from each side), and then patched to 16 non-overlapped patches of size 256 × 256 pixel. A total of 664 patches with the same size (256 × 256 pixel) were extracted from the Post-NAT-BRCA dataset for model pretraining.

The binary masks were generated for each image patch. The weighted masks were only generated for the patches of the training set. The weight of each pixel (x) in the weighted mask was calculated using the Equation 1 [13]. Each weighted mask was then normalized to the range [0, 1].

$$w(x) = w_c(x) + w_0 \times exp\left(-\frac{\left(d_1(x)+d_2(x)\right)^2}{2\sigma^2}\right) \qquad (1)$$

In Equation 1, w_c is the binary mask, d_1 is the distance to the border of the nearest nucleus, and d_2 is the distance to the border of the second nearest nucleus. w_0 and σ are constants that were selected empirically. Using this equation for generating the weighted masks, the pixels between closely adjacent nuclei are assigned higher weights in order to force the model to learn the separation in these regions with higher priority during the training process when a weighted loss function is applied.

C. Framework

The proposed cascaded framework consists of a U-Net model with a weighted pixel-wised loss function followed by a vanilla U-Net model with a VGG16 backbone and a soft Dice loss function [21], [22]. An effective nuclei segmentation model should learn the separation between adjacent nuclei while pixel-level accuracy is insufficient to evaluate the model on large histology images with many nuclei. This is because the number of pixels between the touching or very close nuclei are typically very few compared to the entire image and misclassifying them does not affect the pixel-level accuracy considerably during the model training. Generally, such separation can be accomplished by performing morphological operations on the images, but these operations are difficult to be incorporated into the model's learning process. As an alternative approach, the network could be forced to learn zone separations exclusively from the data. Following this approach, the pre-calculated weight maps were incorporated into the loss function to penalizes the loss in border areas between the touching or very close nuclei more than anywhere else (Equation 1) [13]. Considering the constant parameters in Equation 1 (w_0, σ), there is a trade-off between the level of loss penalty for the border pixels between touching or very close nuclei, and the priority of not missing the entire (small) nuclei. In this equation, a higher value of w_0 results in increasing the weight difference between the nuclei (foreground) and background regions within the border areas, while increasing the value of σ increases the difference in pixel weights based on the distance to the adjacent nuclei. On this basis, if the weighted U-Net is utilized with a high penalization weight for the areas near the touching or very close nuclei, the model would fail to detect many small nuclei or the center regions of larger objects, and in case a low penalization weight is applied, the model would fail in detecting the border of very close nuclei accurately. The cascaded segmentation framework in this study was proposed to address this issue by balancing the trade-off between detecting the small nuclei and segmenting the border of touching or very close nuclei accurately.

For training the weighted U-Net the loss function was modified from the regular cross entropy to a weighted cross entropy by multiplying $w(x)$ to it. In computing the custom loss function, the weights were applied to the log of the activation before calculating the pixel-wise sum. In the implementation step, as the model needs an ordinary loss function for training optimization, three non-trainable layers were added at the end of the weighted U-Net model for the training phase to embed the weighting of the log of the activation into the model and only leaving the summation step to the custom loss function. In this phase, the weighted U-Net was fed by the training images along with the corresponding binary and weighted masks and the segmentation probability maps were generated as the output of the model. The probability maps along with the binary masks were passed to the second component of the framework, i.e., the vanilla U-Net model with a VGG16 backbone and a soft Dice loss function. The reason for adding the VGG16 backbone is to reduce the number of network parameters for a better generalization. The choice of a smaller convolution kernel (i.e., 3×3) is indeed a key to VGG's remarkable accomplishment. Compared to other networks such as ResNet [23] or AlexNet [24], when acquiring the same receptive field, a smaller convolution kernel can not only use less computation and provide more non-linearity, but also make the model more powerful in fitting ability [21]. The choice of the soft Dice loss function potentially results in penalizing low-confidence predictions of the nuclei, since the ground truth is binary in this model and only the pixels belong to the nuclei are considered in the loss calculation.

The proposed approach needs calculation of weighted masks for each input image, while in the test phase this information is not available. Considering that the last layers of the weighted U-Net model are non-trainable and only used for loss calculation, we address this issue by initializing two different models for the training and test phases. Specifically, the training model applied both the weighted and binary masks as the ground truth for the weighted loss calculation, whereas the test model only applied the binary masks for model evaluation on the unseen samples using the output of the layer before the non-trainable layers.

D. Postprocessing

Two morphological operations of erosion and dilation with the kernel size of five pixels were applied on the predicted segmentation masks to remove very small pixel clusters associated with noise., Considering that the histology slides were scanned at 40x magnification, the nuclei could not be apparent with such a small diameter on the images. It should be noted that since this post processing step only removes very small pixel clusters it would mainly improve the F1-score, with minimal effect on the Aggregated Jaccard Index (AJI) metric. At the end, the segmentation masks generated for the patches of each image were concatenated to reconstruct the whole image masks (1000 ×1000 pixel) for evaluation.

E. Evaluation

Performance of the proposed model was evaluated on the MuNoSeg test set and compared with other state-of-the-art segmentation frameworks. The evaluation was performed at both the object-level (nuclei detection) and pixel-level (accuracy of nuclei segmentation contours), using the F1-score and AJI metrics, respectively. In calculating the F1-score, the true positive (TP), false positive (FP) and false negative (FN) were determined as the number of correctly detected, wrongly detected, and undetected nuclei in each image, respectively.

III. RESULTS

Figure 1 demonstrates weighted masks generated for a representative patch with different constant values. In this study, the values of $w_0 = 3$ and $\sigma = 10$ were selected empirically for these parameters to force the weighed model to differentiate between the nuclei and the background regions within the border area of very close nuclei with high confidence. The generated weighted mask resulted in distinguishing the touching nuclei effectively, while missing some small nuclei and the center region of the larger nuclei. The probability maps generated by the first network were passed to the second network to enhance the final prediction mask. Figure 2 compares the output of different trained model on MoNuSeg test set with the ground truth. The segmentation mask generated by the attention U-Net (Figure 2(a, b)) and vanilla U-Net (Figure 2(c, d)) lacks accuracy in finding the nuclei boarders, particularly for the very close nuclei. The generated mask in Figure 2(e, f) shows the ability of the weighted U-Net model in separating the touching nuclei, however, the model missed to detect a few small nuclei and the center of few larger nuclei. The cascaded model (Figure 2(g, h)) took advantage of both networks in finding the accurate border of very close nuclei, detecting small nuclei and completely annotating the larger nuclei. Table 1 shows the performance of the proposed model on the MoNuSeg dataset compared to the attention U-net, vanilla U-Net with the VGG16 backbone and the weighted U-Net model in terms of AJI and F1-score metrics for the training, validation, and unseen test sets. The cascaded model could outperform each single model in all metrics. Table 2 demonstrates the AJI of the top five state-of-the-art models on the same dataset based on the MoNuSeg test set leaderboard, where the superior performance of the cascaded model in nuclei segmentation task is indicated.

IV. Discussion and conclusion

In this study, a cascaded U-Net based framework was proposed for segmenting the nuclei in digital histology images accurately. A U-Net model with a pixel-wised weighted loss function followed by a vanilla U-Net with a VGG16 backbone were adapted in the framework to annotate the touching nuclei accurately while keeping the number of correctly detected nuclei as high as possible. Three non-trainable layers were added to the weighted U-Net model to calculate the pixel-wised weighted loss function during the training phase. The probability maps generated by the weighted U-Net were passed to the U-Net model with the VGG16 backbone and a soft Dice loss function to generate the final masks. The weighted U-Net was responsible to remove the deceptive texture in the background which could be incorrectly considered as object (nuclei) and detect the border of the touching nuclei with high precision while the second U-Net was in charge of giving the same priority to all pixels to detect small nuclei and the missed (center) regions in the larger nuclei. The proposed framework could generate the binary mask on MoNuSeg test set with an AJI of 0.72, which shows a considerable improvement compared to the previous models. Whereas the cascaded model could outperform the former models, further investigation is still required with the state-of-the-art methods to evaluate and potentially improve the performance of the nuclei segmentation framework on larger datasets.

TABLE I. Performance of the proposed model on the MoNuSeg dataset

Model	Training		Validation		Test	
	AJI	F1	AJI	F1	AJI	F1
Attention U-Net	0.70	0.79	0.67	0.76	0.67	0.74
Vanilla U-Net with VGG16 backbone	0.70	0.80	0.68	0.77	0.68	0.76
Weighted U-Net	0.72	0.82	0.72	0.81	0.70	0.79
Cascaded Model	**0.74**	**0.84**	**0.73**	**0.84**	**0.72**	**0.83**

TABLE II. Comparison of the proposed model performance with other state-of-the-art models

Segmentation Model	AJI
Yunzhi [18]	0.68
Pku.hzq [18]	0.69
BUPT.J.LI [18]	0.69
CIA-Net [8]	0.70
U-Net++ [17]	0.70
SSL [25]	0.71
Proposed Cascaded Model	**0.72**

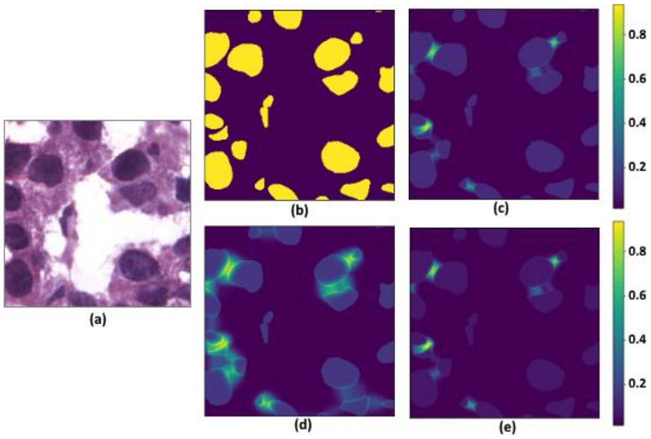

Figure 1. The binary and normalized weighted masks generated for a representative histology image patch (a) with different parameters: (b) binary ground truth ($w_0 = 0$), (c) $w_0 = 5, \sigma = 10$, (d) $w_0 = 3, \sigma = 10$, and (e) $w_0 = 10, \sigma = 5$.

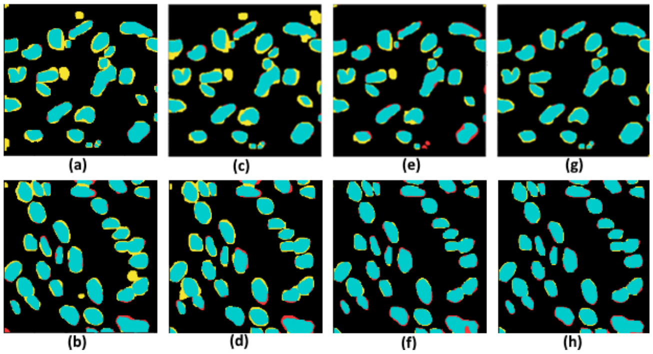

Figure 2. Comparison of the output segmentation masks of different networks with ground truth for two representative patches (256×256 pixel): (a, b) attention U-Net, (c, d) vanilla U-Net with the VGG16 backbone, (e, f) weighted U-Net, (g, h) cascaded framework. The pixel colors indicate true positive (green), true negative (black), false positive (yellow), and false negative (red).

References

[1] F. J, E. M, L. F, C. M, M. L, and P. M, "Global Cancer Observatory: Cancer Today," *International Agency for Research on Cancer*, 2020. https://gco.iarc.fr/today.

[2] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021.

[3] C. E. DeSantis *et al.*, "Breast cancer statistics, 2019," *CA. Cancer J. Clin.*, vol. 69, no. 6, pp. 438–451, Nov. 2019, doi: 10.3322/caac.21583.

[4] A. Shrestha, C. Martin, M. Burton, S. Walters, K. Collins, and L. Wyld, "Quality of life versus length of life considerations in cancer patients: A systematic literature review," *Psychooncology.*, vol. 28, no. 7, pp. 1367–1380, 2019, doi: 10.1002/pon.5054.

[5] C. Loukas, S. Kostopoulos, A. Tanoglidi, D. Glotsos, C. Sfikas, and D. Cavouras, "Breast Cancer Characterization Based on Image Classification of Tissue Sections Visualized under Low Magnification," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–7, 2013.

[6] X. Wang *et al.*, "RaPtomics: integrating radiomic and pathomic features for predicting recurrence in early stage lung cancer," in *Medical Imaging 2018: Digital Pathology*, Mar. 2018, p. 21, doi: 10.1117/12.2296646.

[7] R. Gupta, T. Kurc, A. Sharma, J. S. Almeida, and J. Saltz, "The Emergence of Pathomics," *Curr. Pathobiol. Rep.*, vol. 7, no. 3, pp. 73–84, Sep. 2019, doi: 10.1007/s40139-019-00200-x.

[8] Y. Zhou, O. F. Onder, Q. Dou, E. Tsougenis, H. Chen, and P.-A. Heng, "CIA-Net: Robust Nuclei Instance Segmentation with Contour-Aware Information Aggregation," *Int. Conf. Inf. Process. Med. Imaging*, pp. 682–693, 2019, doi: 10.1007/978-3-030-20351-1_53.

[9] T. Kurc *et al.*, "Segmentation and Classification in Digital Pathology for Glioma Research: Challenges and Deep Learning Approaches," *Front. Neurosci.*, vol. 14, 2020, doi: 10.3389/fnins.2020.00027.

[10] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2021.

[11] M. Khoshdeli, G. Winkelmaier, and B. Parvin, "Fusion of encoder-decoder deep networks improves delineation of multiple nuclear phenotypes," *BMC Bioinformatics*, vol. 19, no. 1, p. 294, 2018.

[12] T. Le Dinh, S.-G. Kwon, S.-H. Lee, and K.-R. Kwon, "Breast Tumor Cell Nuclei Segmentation in Histopathology Images using EfficientUnet++ and Multi-organ Transfer Learning," *J. Korea Multimed. Soc.*, vol. 24, pp. 1000–1011, 2021.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Med. Image Comput. Comput. Interv. (MICCAI)*, pp. 234–241, 2015.

[14] "2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), Cell Tracking Challenge," *IEEE*, [Online]. Available: https://ieeexplore.ieee.org/xpl/conhome/7150573/proceeding.

[15] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," *MIDL 2018 Conf.*, 2018.

[16] J. Fang, Q. Zhou, and S. Wang, "Segmentation Technology of Nucleus Image Based on U-Net Network," *Sci. Program.*, vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/1892497.

[17] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," *Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support*, vol. 11045, pp. 3–11, 2018.

[18] N. Kumar *et al.*, "A Multi-Organ Nucleus Segmentation Challenge," *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1380–1391, 2020.

[19] X. Xie, J. Chen, Y. Li, L. Shen, K. Ma, and Y. Zheng, "Instance-Aware Self-supervised Learning for Nuclei Segmentation," *Med. Image Comput. Comput. Assist. Interv. – MICCAI 2020.*, vol. 12265, pp. 341–350, 2020, doi: 10.1007/978-3-030-59722-1_33.

[20] A. L. Martel, S. Nofech-Mozes, S. Salama, S. Akbar, and M. Peikari, "Assessment of Residual Breast Cancer Cellularity after Neoadjuvant Chemotherapy using Digital Pathology [Data set]," *Cancer Imaging Arch.*, 2019, doi: 10.7937/TCIA.2019.4YIBTJNO.

[21] R. Zhang, L. Du, Q. Xiao, and J. Liu, "Comparison of Backbones for Semantic Segmentation Network," *J. Phys. Conf. Ser.*, vol. 1544, no. 1, p. 012196, May 2020, doi: 10.1088/1742-6596/1544/1/012196.

[22] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Oct. 2020, pp. 1–7, doi: 10.1109/CIBCB48159.2020.9277638.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, vol. 25, pp. 1097–1105.

[25] X. Xie, J. Chen, Y. Li, L. Shen, K. Ma, and Y. Zheng, "Instance-Aware Self-supervised Learning for Nuclei Segmentation," *Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, pp. 341–350, 2020.