Multi-Attention Feature Fusion for Nuclei Segmentation and Classification

Xingpeng Zhang [1, a], Peng Guo [1, b]

[1]School of Computer and Software, Southwest Petroleum University, Chengdu, 610500, China;

[a] xpzhang@swpu.edu.cn, [b] pengg_0314@163.com

Abstract. Nucleus instance segmentation is crucial in the digital pathology, serving as a foundational step for subsequent tasks like precision medicine and cancer prognosis. In this paper, we introduce MAP-SegNet, a new network architecture designed for the segmentation and classification of nuclei in H&E stained images from multiple tissues. The key innovation of MAP-SegNet lies in its Multi-Attention Feature Fusion Module, which enhances the network's ability to efficiently combine multi-scale feature aggregation mechanisms with local and global features. The core component is the dual attention fusion module, which possesses the capability to adjust spatial attention weights dynamically by the input contextual information. This functionality enables the model to concentrate on various image regions, effectively capture long-range dependencies, and mitigate the risk of information loss. MAP-SegNet's uplifting accuracy in nuclei segmentation and classification is validated by rigorous quantitative experiments on the standard PanNuKe dataset.

Keywords: Nuclei segmentation; Instance segmentation; Dual attention; Deep learning.

1. Introduction

In the domain of medicine, the hematoxylin and eosin (H&E) staining is extensively employed for visualizing the structures of cellular tissues and their morphological attributes in tissue sections. To perform downstream analysis using feature information, nuclei segmentation must be conducted as an initial step. Segmentation of nuclei is crucial in pathological diagnosis as it involves identifying and labeling each pixel associated with distinct nuclei in histopathological images [1]. However, the traditional manual screening methods are time and effort-consuming, attributable to the large scale and intricate cellular structures of histopathological images. Furthermore, the quality of outcomes is heavily relies on the experience of the individual screeners and their conditions at the time of analysis, which can lead to significant inter-observer variability [2].

In recent years, a variety of computer-assisted approaches have been introduced, marking significant advancements over traditional manual screening. For instance, the Mask R-CNN [3], typically utilized for object detection, often results in models that are excessively complex and intricate, which in turn leads to a noticeable decrease in the speed of real-time inference performance. On the other hand, the encoder-decoder architecture known as U-Net [4] has given rise to a plethora of different variants that are widely applied in medical image analysis [5]. In addition, a significant number of studies have incorporated attention mechanisms into their methodologies for medical image analysis [6], with channel attention being the most commonly implemented type that demonstrates exceptional performance under specific circumstances and conditions. Moreover, several methods utilize the contours of nuclei [7, 8] for analysis. For example, certain methodologies based on regression modeling effectively perform nuclei segmentation and classification tasks by employing regression distance maps [9] as well as heat maps [10].

It is worth noting that the significant portion of the information present in H&E-stained histological images is encapsulated within their respective channels, while the remaining information can be found within the spatial domain of the images. Therefore, We postulate that the integration of multiple attention mechanisms (channel attention mechanisms and spatial attention mechanisms) will yield additional performance improvements. Based on this insight, we have developed the Multi-Attention Feature Fusion (MAFF) module. Our aim with this module is to seamlessly merge feature information from neighboring layers by strategically focusing attention on

the relevant feature regions through the Dual Attention Fusion Module(DAFM). Through this attention-based feature fusion approach, we can obtain robust features that significantly enhance the quality of segmentation and classification outputs. The output generated by the MAFF consists of the aggregation of multi-scale features, which are subsequently utilized to derive the final segmentation and classification results via keypoint estimation methods. We have designated our proposed model as MAP-SegNet for Multi-Attention Fusion Keypoint Nuclei Segmentation Network.

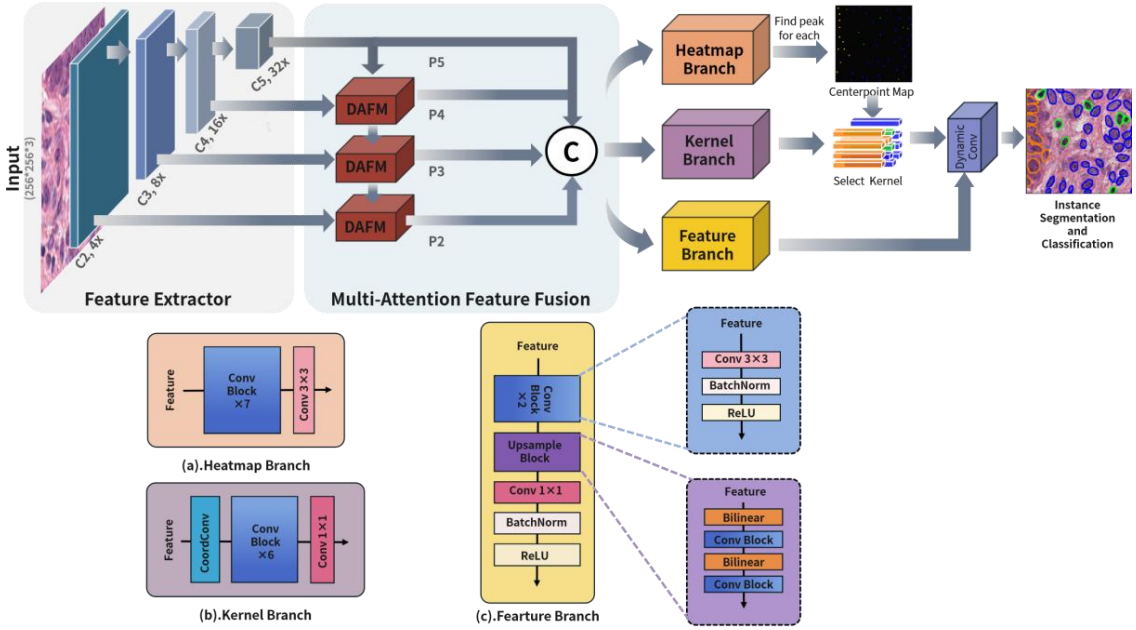

Fig. 1 Framework of the proposed MAP-SegNet. DAFM denotes Dual Attention Fusion Module.

2. Methodology

2.1 Proposed Method

Our proposed model MAP-SegNet, as illustrated in Fig. 1, is closely related to the PointNu-Net architecture [10]. PointNu-Net utilizes the Joint Pyramid Fusion Module (JPFM) for feature fusion. This module directly cascades features from different scales together and employs dilated convolution to model the dependencies between these multi-scale features. However, a challenge arises when feature information from different scales is directly connected, as it can lead to interference, making it difficult for the model to accurately extract critical location and boundary information of the target nuclei. While dilated convolution is effective in expanding the receptive field, it struggles to simultaneously capture both global context and local details. To address these issues, we introduce the Multi-Attention Multi-Feature Fusion (MAFF) module.

After the MAFF processing, the features learned from this stage are utilized as inputs to three independent branches designed for specific prediction tasks. The heatmap branch is responsible for predicting the centroid of each nucleus, facilitating detection and classification. The kernel branch incorporates normalized coordinates to effectively embed positional information. The feature branch utilizes bilinear interpolation to generate high-resolution features from the learned multi-scale representations, enhancing the detail and quality of the feature maps. As depicted in Fig. 1, After the nuclei centroids are predicted by the heatmap branch, corresponding nuclear vectors are selected from the nuclear vectors generated by the kernel branch, based on the location of these centroids. Finally, the instance mask prediction is produced by dynamically convolving the selected kernel vectors with the high-resolution features obtained from the feature branch.

2.2 Feature Fusion

The MAFF module comprises several novel Dual Attention Fusion Module(DAFM) components. With the implementation of these modules, the attention mechanism network can dynamically adjust the attention levels assigned to neighboring scale features. This dynamic adjustment allows the model to better focus on crucial regions within the image, and effectively fuse the rich contextual information present in multi-scale feature maps, thereby enhancing the robustness and performance of the overall model.

2.2.1 Dual Attention Fusion Module

Inspired by the Global Attention Mechanism(GAM) [11], we introduce a unified and multi-purpose structure called the Dual Attention Fusion Module(DAFM). The DAFM is designed to effectively integrate features that have both semantic and scale differences. It comprises two tandem components: one for capturing feature information in the channel dimension and another for the spatial dimension. The DAFM structure is depicted in Fig. 2. The input features C_i and C_{i+1} are initially summed to form a combined input for the channel attention submodule.

Channel Attention Module: This submodule employs the dimension operation of the exchanged to preserve cross-latitude information. Cross-dimensional dependencies in channel space are enhanced by a Multi-Layer Perceptron (MLP).

Spatial Attention Module: Comprising two convolutional blocks, this submodule functions to adjust channel quantities for computational efficiency. Initially, a 7×7 convolution kernel in the first convolutional layer reduces channel numbers to lessen computational demands. The second convolutional layer, also using a 7×7 convolution kernel, increases the number of channels back to maintain consistency with the original feature map dimensions.

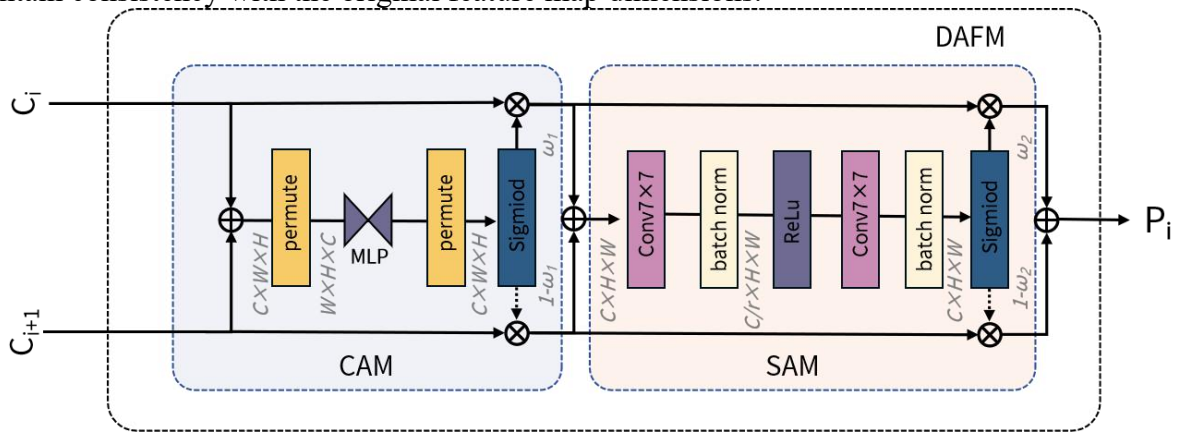

Fig. 2 Structure of DAFM. In this picture, the dotted lines symbolize 1-ω, ⊕ denote the addition, and ⊗ denote the multiplication.

3. Experiment

3.1 Dataset and Metrics

Dataset: PanNuke [12] is recognized as the largest publicly accessible dataset focused on nuclei segmentation and classification within H&E stained tissue images. It comprises 7904 images, each measuring 256 × 256 pixels, derived from over 20,000 complete section images at varying levels of magnification. The dataset provides extensive annotations for 19 distinct tissue types and 5 different nuclei types. We categorized all images into three distinct groups in accordance with the evaluation procedures described in the relevant literature [13].

Metrics: Panoramic Quality(PQ) was introduced in [9] for nuclei segmentation. To measure the performance of nuclei instance segmentation, we follow the approach, which is defined as follows:

$$PQ = \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{DQ} \times \underbrace{\frac{\sum_{(p,g)\in TP} IOU(p,g)}{|TP|}}_{SQ}$$

The first component pertains to the Detection Quality (DQ), and the second component corresponds to the Segmentation Quality (SQ). TP indicates the number of correctly predicted instances, FP refers to the number of mispredicted positive instances, and FN signifies the number of negative instances with inaccurate predictions, which is employed to assess the precision of the segmentation results. The $\sum_{(p,g)\in TP} IOU(p,g)$ symbolizes the IOU value of the cumulatively correctly matched objects. To evaluate performance of nuclei segmentation and classification, we adhered to the methodology detailed in PointNu-Net [10].

Table 1. Comparison of experimental results on the PanNuke dataset.

Tissue	Micro-Net		Hover-Net		CPP-Net		PointNu-Net		Ours	
	mPQ	bPQ	mPQ	bPQ	mPQ	bPQ	mPQ	bPQ	mPQ	bPQ
Adrenal	0.4153	0.6440	0.4812	0.6962	0.4922	0.7031	0.5115	0.7134	**0.5145**	**0.7152**
Bile-duct	0.4124	0.6232	0.4714	0.6696	0.4650	0.6739	**0.4868**	**0.6814**	0.4831	0.6803
Bladder	0.5357	0.6488	0.5792	0.7031	0.5932	0.7057	**0.6065**	0.7226	0.5660	**0.7266**
Breast	0.4407	0.6029	0.4902	0.6470	0.5066	0.6718	0.5147	0.6709	**0.5238**	**0.6744**
Cervix	0.3795	0.6101	0.4438	0.6652	0.4779	0.6880	**0.5014**	**0.6899**	0.4988	0.6850
Colon	0.3414	0.4972	0.4095	0.5575	0.4269	0.5888	0.4509	0.5945	**0.4628**	**0.5981**
Esophagus	0.4668	0.6011	0.5085	0.6427	0.5410	0.6755	0.5504	0.6766	**0.5578**	**0.6874**
H&N	0.3668	0.5242	0.4530	0.6331	0.4667	0.6468	**0.4838**	**0.6546**	0.4709	0.6503
Kidney	0.4165	0.6321	0.4424	0.6836	0.5092	0.7001	0.5066	0.6912	**0.5517**	**0.7090**
Liver	0.4365	0.6666	0.4974	0.7248	0.5099	0.7271	0.5174	**0.7314**	**0.5355**	0.7288
Lung	0.3370	0.5588	0.4004	0.6302	0.4234	0.6364	0.4048	0.6352	**0.4261**	**0.6482**
Ovarian	0.4387	0.6013	0.4863	0.6309	0.5276	0.6792	0.5484	0.6863	**0.5774**	**0.6946**
Pancreatic	0.4041	0.6074	0.4600	0.6491	0.4680	0.6742	0.4804	0.6791	**0.5258**	**0.6811**
Prostate	0.4341	0.6049	0.5101	0.6615	0.5261	0.6903	0.5127	0.6854	**0.5465**	**0.6940**
Skin	0.3223	0.5817	0.3429	0.6234	0.3547	0.6192	0.4011	**0.6494**	**0.4078**	0.6093
Stomach	0.3872	0.6293	0.4726	0.6886	0.4553	0.7043	0.4517	0.7010	**0.4949**	**0.7051**
Testis	0.4088	0.6300	0.4754	0.6890	0.4917	0.7006	0.5334	**0.7058**	**0.5358**	0.6994
Thyroid	0.3712	0.6555	0.4315	0.6983	0.4344	0.7094	0.4508	0.7076	**0.4715**	**0.7326**
Uterus	0.3965	0.5821	0.4393	0.6393	0.4790	0.6622	**0.4846**	0.6634	0.4664	**0.6791**
Average	0.4059	0.6053	0.4629	0.6596	0.4817	0.6767	0.4957	0.6808	**0.5062**	**0.6849**

Specifically, PQs were reported for 19 different tissue types. Multi-category PQ(mPQ) assesses detection, segmentation, and classification performance by evaluating each nucleus class independently. The overall performance is the average PQ across the 5 nuclei classes. Binary PQ(bPQ) focuses on detection and segmentation performance without considering classification. It assumes that all nuclei belong to a single category and computes performance accordingly, ignoring the specific class distinctions.

3.2 Result

On the large dataset PanNuke, the performance of the proposed MAP-SegNet is evaluated against various advanced deep learning techniques, such as MicroNet [14], Hover-Net [9], CPP-Net [15], and PointNu-Net [10] networks. Table 1 illustrates that the proposed MAP-SegNet exhibits top-tier performance in the mPQ and bPQ metrics across 19 different types of organizations. In the mPQ metric, our MAP-SegNet's organizational performance is slightly lower than that of PointNu-Net for Bile Duct, Bladder, Head and Neck, Uterus, and Cervix, and outperforms all other models for the remaining organizations. In particular, in terms of average mPQ across organizations, our MAP-SegNet outperforms the third-ranked CPP-Net and the second-ranked PointNu-Net by 2.45% and 1.05% (|mPQ(ours) − mPQ(others)| ∗ 100), respectively. In addition, MAP-SegNet

surpasses the majority of existing methods in the bPQ evaluation metric, both in terms of overall average performance and across various organizational benchmarks. Specifically, in terms of average bPQ by organization, our MAP-SegNet outperforms Hover-Net, CPP-Net, and PonitNu-Net by 2.53%, 0.82%, and 0.41% ($|bPQ(ours) - bPQ(others)| * 100$), respectively. Fig. 3 shows some qualitative results of MAP-SegNet on the large dataset PanNuKe. The proposed MAP-SegNet attains remarkable outcomes on the PanNuke for the segmentation and classification of nuclei across 19 tissue.

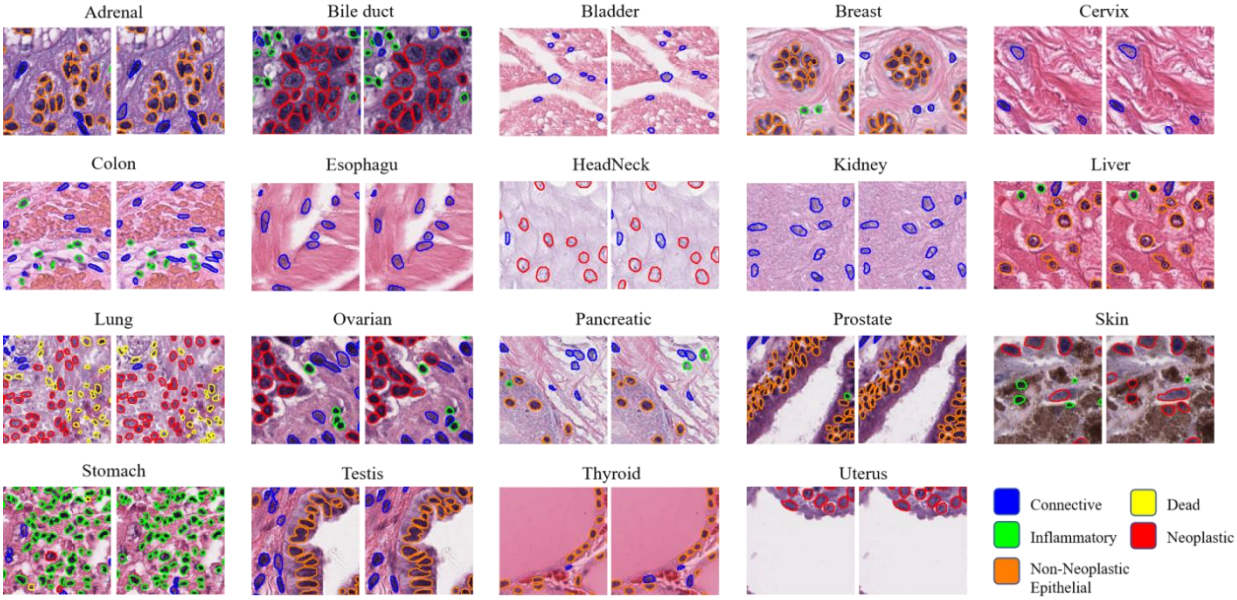

Fig. 3 Examples of MAP-SegNet results for 19 organisations on PanNuke. The ground truth is on the left and the MAP-SegNet prediction is on the right.

Table 2. Average PQ of each nuclei type for the three divisions on the PanNuke dataset.

	Neo	Epi	Inflam	Conn	Dead
DIST	0.439	0.290	0.343	0.275	0.000
Mask-RCNN	0.472	0.403	0.290	0.300	0.069
Micro-Net	0.504	0.442	0.333	0.334	0.051
Hover-Net	0.551	0.491	0.417	0.388	0.139
PointNu-Net	0.578	0.577	0.433	0.409	**0.154**
Ours	**0.582**	**0.581**	**0.439**	**0.422**	0.148

Table 2 further presents the average PQ for each nuclei category within the PanNuke. As observed, our MAP-SegNet surpasses the latest models in performance, with the exception of the dead cell category. In some cases, distinguishing between neoplastic and nonneoplastic cells is challenging, but our MAP-SegNet improves the average PQ of these two categories of nuclei by 0.004 compared to the PointNu-Net method. Dead cells exhibit the lowest PQ among all models, which may be attributed to the imbalanced representation of categories within the PanNuke. This difficulty contributes to the observed poor performance in this category.

4. Conclusion

In this paper, we introduce MAP-SegNet, a novel network architecture designed for the segmentation and classification of nuclei. The core component of our design is the dual-attention fusion module, which enables the network to dynamically adjust spatial attention weights based on the input's contextual information. This selective attention mechanism enables the model to focus on diverse regions in the image and capture long-range dependencies, mitigating the risk of information loss. By assigning attention weights across different channels and spatial regions, the module ensures that the network can handle diverse and complex features within the data more

efficiently. Our approach has been rigorously validated through quantitative experiments, which demonstrate that MAP-SegNet delivers leading performance in accuracy when tested on standard nuclei segmentation datasets.

Acknowledgements

This work is supported by the Natural Science Starting Project of SWPU (No. 2022QHZ023, 2022QHZ013), the Sichuan Science and Technology Program (No.2024YFH0022), the Sichuan Scientific Innovation Fund (No.2022JDRC0009).

References

[1] F. Clayton, Pathologic correlates of survival in 378 lymph node-negative infiltrating ductal breast carcinomas. mitotic count is the best single predictor, Cancer 68 (6) (1991) 1309–1317.

[2] E. Davey, A. Barratt, L. Irwig, S. F. Chan, P. Macaskill, P. Mannes, A. M. Saville, Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review, Lancet 367 (9505) (2006) 122–132.

[3] J. W. Johnson, Automatic nucleus segmentation with mask-rcnn, in: Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), 2020, pp. 399–407.

[4] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241

[5] S. Yildiz, A. Memi s, S. Varl, Nuclei segmentation in colon histology images by using the deep cnns: A u-net based multi-class segmentation analysis, in: Medical Technologies Congress (TIPTEKNO), 2022, pp. 1–4.

[6] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, C. Fan, Sa-unet: Spatial attention u-net for retinal vessel segmentation, in: International Conference on Pattern Recognition (ICPR), 2021, pp. 1236–1242.

[7] H. Chen, X. Qi, L. Yu, P.-A. Heng, Dcan: deep contour-aware networks for accurate gland segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2487–2496.

[8] Q. D. Vu, S. Graham, T. Kurc, M. N. N. To, M. Shaban, T. Qaiser, N. A. Koohbanani, S. A. Khurram, J. Kalpathy-Cramer, T. Zhao, et al., Methods for segmentation and classification of digital microscopy tissue images, Frontiers in Bioengineering and Biotechnology (2019) 53.

[9] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, N. Rajpoot, Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images, Medical Image Analysis 58 (2019) 101563.

[10] K. Yao, K. Huang, J. Sun, A. Hussain, Pointnu-net: Keypoint-assisted convolutional neural network for simultaneous multi-tissue histology nuclei segmentation and classification, IEEE Transactions on Emerging Topics in Computational Intelligence 8 (2024) 802–813.

[11] Y. Liu, Z. Shao, N. Hoffmann. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. arXiv preprint arXiv:2112.05561, 2021.

[12] J. Gamper, N. A. Koohbanani, K. Benet, A. Khuram, N. Rajpoot, Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification, in: European Congress on Digital Pathology, 2019, pp. 11–19.

[13] J. Gamper, N. A. Koohbanani, S. Graham, M. Jahanifar, S. A. Khurram, A. Azam, K. Hewitt, N. Rajpoot, Pannuke dataset extension, insights and baselines, arXiv preprint arXiv:2003.10778 (2020).

[14] S. E. A. Raza, L. Cheung, M. Shaban, S. Graham, D. Epstein, S. Pelengaris, M. Khan, N. M. Rajpoot, Micro-net: A unified model for segmentation of various objects in microscopy images, Medical Image Analysis 52 (2019) 160–173.

[15] S. Chen, C. Ding, M. Liu, J. Cheng, D. Tao, Cpp-net: Context-aware polygon proposal network for nucleus segmentation, IEEE Transactions on Image Processing 32 (2023) 980–994.