Attention-based Mask R-CNN for Microvascular Segmentation

Yufei Zhang*
Beijing University of Technology
Faculty of Information Technology
Beijing, China
denny_0601@126.com

Henan Zhao*
Beijing University of Technology
Faculty of Information Technology
Beijing, China

Zekai Yang*
Beijing University of Technology
Faculty of Information Technology
Beijing, China

Tong Mo*
Beijing University of Technology
Faculty of Information Technology
Beijing, China

Yiya Yao*
Beijing University of Technology
Faculty of Information Technology
Beijing, China

* These author *above contributed equally to this work.*

***Abstract*—The normal functioning of human organs and tissues depends on the spatial organization, specialization, and interaction between cells. With a total of 37 trillion cells in the human body, determining their functions and relationships is highly challenging. The Vascular Coordinate Common Framework (VCCF) is widely used for cell mapping, but it is limited by a lack of knowledge about the microvascular system. Therefore, we propose a novel framework called Attention-based Mask R-CNN for Microvascular Segmentation to achieve automated segmentation of microvascular arrangements. In terms of structure, our model replaces the original fully connected layers with the Feature Pyramid Network (FPN) as the segmentation head network and integrates feature maps from different levels. To enhance performance, we introduce RoI (Region of Interest) feature alignment, employing techniques such as bilinear interpolation to accurately align features within the RoI region. Furthermore, to better focus on the regions of interest, we introduce an attention mechanism specifically designed for Human Vasculature, which improves the segmentation accuracy of critical areas. We train our model using 2D PAS-stained histological images of healthy human kidney tissue sections. Experimental results demonstrate the superior segmentation outcomes of our approach for microvascular structures, including capillaries, arterioles, and venules.**

***Index Terms*—Instance Segmentation, Medical Imaging, Mask R-CNN, Attention,**

I. Introduction

Cells are the fundamental units of life, existing in astonishing numbers and diversity within our bodies. Each cell has specific tasks and functions, such as providing energy, synthesizing proteins, and transmitting signals [1], [2]. However, cells do not exist in isolation; they are closely interconnected, forming complex tissues and organs. The normal functioning of human organs and tissues also relies on the spatial organization, specialization, and interaction of cells. Researchers attempt to understand the working mechanisms of human organs by mapping the cellular landscape. However, accomplishing this task is highly challenging due to the staggering number of cells in the human body, totaling 37 trillion [3].

Currently, the work of creating cellular atlases utilizes the Vascular Common Coordinate Framework (VCCF) [4], which provides a unique method of using capillary structures as addresses to identify cell locations. Within this framework, microvascular systems play an extremely important role. Accurate segmentation of microvascular networks [5] is of significant importance for researchers to understand the structure, function, and disease diagnosis of biological tissues. However, due to the complexity and diversity of microvascular systems, achieving efficient and accurate automatic segmentation remains a challenging task. Additionally, researchers have limited knowledge about the microvascular system, resulting in gaps in the VCCF. Therefore, if we can automatically segment the arrangement of microvascular systems, real-world tissue data can be utilized for vascular system mapping.

In the past few decades, the field of computer vision has made significant advancements, particularly in the areas of object detection and image segmentation [6], [7]. However, existing methods face challenges when it comes to segmenting microvascular systems [8]. Firstly, microvascular systems often exhibit complex shapes and varying scales, posing a challenge for traditional feature-based methods. Secondly, due to the subtle texture and weak edges of microvascular systems, traditional segmentation methods often struggle to accurately capture the boundaries of microvessels. Figure 1 shows examples of sliced images from a microvascular system, providing a more intuitive view of the complex cell distribution and morphology.

To address these issues, we propose an Attention-based Mask R-CNN (Mask Region Convolutional Neural Network) method for automatic segmentation of microvascular systems. Attention mechanisms have been widely applied in natural language processing and computer vision, as they can learn the importance of different regions in an image, thereby improving the accuracy of microvascular system segmentation. Specifically, our method utilizes deep convolutional neural networks to extract image features and introduces an attention module

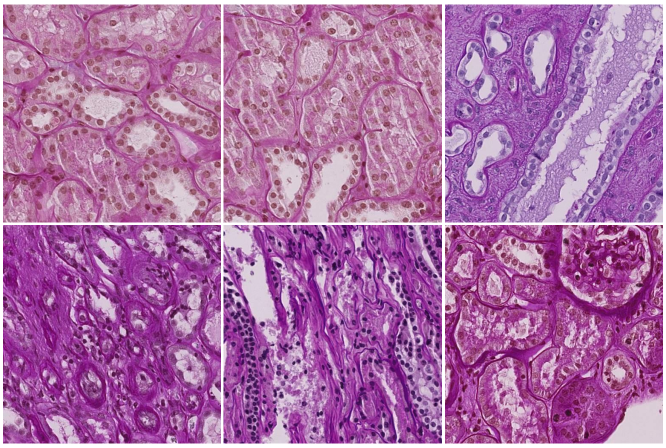

Fig. 1. Examples of microvascular system.

on top of it. This attention module dynamically selects the regions of interest within the microvascular system, enabling better capture of the subtle features of microvessels. Then, we make improvements based on the Mask R-CNN framework. The model adopts Feature Pyramid Network (FPN) to replace the original fully connected layers as the segmentation head network and fuses feature maps from different levels. Additionally, we introduce RoI (Region of Interest) feature alignment, which accurately aligns the features within the RoI region using bilinear interpolation and other techniques. This framework combines the capabilities of object detection and instance segmentation, enabling simultaneous localization and segmentation of different instances within the microvascular system. We conducted experimental evaluations of our method on extensively validated datasets and compared it with existing methods. The experimental results demonstrate significant improvements achieved by our proposed Attention-based Mask R-CNN method in microvascular system segmentation. Our method not only accurately detects the positions of microvascular systems but also captures the details of microvessels, achieving more precise segmentation.

The main contributions of this paper are as follows: 1) Proposing an Attention-based Mask R-CNN method for automatic segmentation of microvascular systems; 2) Designing an attention module that effectively captures the detailed features of microvessels; 3) Conducting experimental evaluations on extensively validated datasets, demonstrating the superior performance of our method in microvascular system segmentation.

II. Related Work

A. Instance Segmentation

Instance segmentation is an important task in the field of computer vision and has seen many influential research works [8], [9]. One classic method is Fully Convolutional Networks (FCN). Long et al. [10] proposed the FCN method, which extends traditional convolutional neural networks to pixel-level segmentation tasks. The FCN method treats image segmentation as a pixel classification problem and uses deconvolutional layers to recover the spatial dimensions of the original image. However, the FCN method does not consider the distinction between different instances, resulting in poor performance in multi-instance scenarios. To address this issue, He et al. [7] proposed the Mask R-CNN method, which combines object detection and instance segmentation. Mask R-CNN introduces an additional segmentation branch based on Faster R-CNN, enabling the model to simultaneously predict the class, bounding box, and segmentation mask of each pixel. Mask R-CNN improves accuracy by introducing RoIAlign operation and Mask Head, and achieves outstanding performance on the COCO dataset. However, Mask R-CNN has a high computational complexity, leading to slow inference speeds. Apart from Mask R-CNN, there have been significant advancements in instance segmentation with other methods. For example, U-Net, proposed by Ronneberger et al. [11] adopts an encoder-decoder architecture and skip connections to capture features at different scales. U-Net is widely used in medical image segmentation and has achieved excellent performance. In this paper, we draw on the strengths of FCN, Mask R-CNN, U-Net, and attention mechanisms to propose a novel method to address challenges in instance segmentation tasks. We validate the proposed method on specific domain instance segmentation problems.

B. Microvascular Segmentation

Microvascular and its segmentation is an important research direction in the field of biomedical imaging [12]. In this domain, some traditional computer vision methods have been applied [13], [14]. For example, methods based on filtering and edge detection can extract the shape and boundary information of microvessels. However, these methods have limited effectiveness in dealing with the complex structure and weak boundaries of microvascular systems, making accurate segmentation challenging. In recent years, deep learning methods have made significant progress in microvascular segmentation tasks. For instance, Chen et al. [15] proposed the U-Net method, which effectively captures the detailed features of microvessels through an encoder-decoder architecture and skip connections. Lu et al. [16] developed clinical radiomics models, radiomics models, and combined models for microvascular segmentation using stepwise regression and multinomial logistic regression. Wu et al. [17] proposed an optimized vessel segmentation method based on an improved Vesselness-based Connectivity Analysis (VCA) algorithm, combined with morphological characterization and noise and artifact elimination. Despite the significant progress of deep learning methods in microvascular segmentation tasks, there are still challenges to overcome. Microvascular systems exhibit complex shapes and diverse scales, requiring network architectures and algorithms that are better suited to microvascular features. Additionally, the lack of large-scale and accurately annotated microvascular segmentation datasets limits algorithm development. In this paper, we aim to address the challenges of microvascular segmentation by drawing on the strengths of traditional computer vision methods, deep learning methods, and attention mechanisms. We propose a novel microvascular segmentation method and validate and evaluate it using benchmark datasets.

III. Method

In this paper, we propose a novel framework for automatic segmentation of microvasculature arrangements, named Attention-based Mask R-CNN. The framework combines Feature Pyramid Network (FPN) and attention mechanism to achieve accurate microvessel segmentation. Next, we will explain our method from two perspectives: framework overview and attention mechanism.

A. Framework overview

The framework proposed in this article combines the Feature Pyramid Network (FPN) and attention mechanism to achieve accurate microvasculature segmentation. In terms of structure, the model replaces the original fully connected layers with FPN as the segmentation head network and integrates feature maps from different levels. In terms of performance, the RoI feature alignment technique is introduced to accurately align the features within the RoI regions using methods such as bilinear interpolation. Additionally, to better focus on the regions of interest, an attention mechanism specifically designed for human blood vessels is introduced to improve the segmentation accuracy of critical areas.

First, the input microvasculature image I to be segmented is fed into the Backbone network. This structure uses a pre-trained convolutional neural network to extract low-level features of the image, denoted as $\{C_1, C_2, C_3, C_4, C_5\} = Backbone(I)$, where C_i represents the i-th feature map. Next, the Feature Pyramid Network (FPN) is utilized to fuse feature maps from different levels, allowing the model to capture multi-scale semantic information and handle microvasculature structures of different scales more effectively. Specifically, for each feature map C_i, a 1×1 convolution operation is applied to adjust the number of channels, resulting in the feature map $M_i = Conv1 \times 1(C_i)$. For each level i from 4 to 1, bilinear interpolation is performed to upsample $U_{i+1} = Upsample(Mi + 1)$, and the feature maps are fused as $M_i = M_i + U_{i+1}$. By applying this process to each feature map C_i, a series of feature maps with different scales $\{P_1, P_2, P_3, P_4, P_5\}$ is generated through upsampling and fusion operations. Once the feature maps are obtained, a region proposal network (RPN) is used to generate candidate microvasculature box feature maps $\{P_1, P_2, P_3, P_4, P_5\}$ at each level of the feature pyramid. The RPN applies convolutional operations to compute classification scores and bounding box regression scores: $cls_scores, bbox_deltas = RPN(P_1, P_2, P_3, P_4, P_5)$, and assigns classification scores and bounding box regression scores to each box. These candidate boxes serve as input for subsequent segmentation operations to locate the position and shape of microvasculature.

Next, RoI feature alignment is performed. For each candidate box ($proposal_box$), the corresponding feature map P_i from the feature pyramid is selected based on its size. For each position in P_i, bilinear interpolation is used to generate a fixed-sized feature map $RoI_feature$ within the $proposal_box$. RoI feature alignment ensures the spatial consistency of features within the RoI region, enabling more accurate representation of microvasculature features inside the candidate boxes and improving the accuracy of subsequent segmentation operations. Finally, the Segmentation Head network takes the fused $RoI_feature$ as input and performs feature extraction and upsampling through a series of convolutional and upsampling layers, resulting in the final segmentation mask ($segmentation_mask$). This mask indicates the position of the microvasculature regions in the image and is used for automatic microvasculature segmentation. In addition, an attention mechanism is introduced in this process to weight the importance of different regions. These weights can be used to allocate weights to the segmentation mask, improving the segmentation accuracy of critical areas such as human blood vessels. Thus, accurate segmentation of complex microvasculature systems is achieved.

B. Attention mechanism

In traditional Mask R-CNN, the generation of segmentation masks is achieved by applying a fully convolutional network on each candidate region. However, for certain targets with complex microvascular systems, such as human blood vessels, the traditional fully convolutional network may fail to capture small and important details, leading to inaccurate segmentation results.

To address this issue, we introduce adaptive weights in Mask R-CNN to quantify the importance of different regions. Taking into account the issue of class imbalance, we combine attention mechanisms to further improve the updating process of adaptive weights. This mechanism allows us to focus more on key regions and enhance the segmentation accuracy of important targets, such as human blood vessels. Specifically, we introduce a set of adaptive weights to control the importance allocation in the generation of segmentation masks. These adaptive weights can be dynamically adjusted based on attention mechanisms, considering the features and contextual information of the targets, in order to better capture small and important details.

When considering the issue of class imbalance and incorporating the attention mechanism, we can update the adaptive weights using a weighted attention loss function to improve the segmentation accuracy of the key regions.

Specifically, we can define a loss function that consists of two components: the attention loss term L_{att} and the segmentation loss term L_{seg}. The attention loss term L_{att} constrains the generation of attention weights to ensure appropriate focus on the key regions. A commonly used attention loss function

TABLE I
PERFORMANCE OF COMPARATIVE EXPERIMENT (ACCURACY(%)).

Models	**Backbone**	**Accuracy(%)**
Mask R-CNN	Conv4	82.820 ± 0.832
U-net	Conv4	83.982 ± 0.389
Blendmask	Conv4	85.238 ± 0.673
Polarmask	Conv4	80.832 ± 0.524
Solov2	Conv4	79.289 ± 0.903
Yolact	Conv4	81.329 ± 0.621
Centermask	Conv4	85.251 ± 0.102
Ours	Conv4	86.544 ± 0.329
Mask R-CNN	Resnet50	85.239 ± 0.283
U-net	Resnet50	86.389 ± 0.298
Blendmask	Resnet50	86.936 ± 0.918
Polarmask	Resnet50	85.932 ± 0.238
Solov2	Resnet50	85.128 ± 0.356
Yolact	Resnet50	88.839 ± 0.652
Centermask	Resnet50	89.932 ± 0.152
Ours	Resnet50	89.200 ± 0.415

is the cross-entropy loss, which can be defined as:

$$L_{att} = -\frac{1}{N}\sum_{p=1}^{N}\sum_{c=1}^{C} Y_{att}^{(p,c)} log(A_i^{(p,c)}) + (1 - Y_{att}^{(p,c)}) log(1 - A_i^{(p,c)}) \tag{1}$$

Here, $Y_{att}^{(p,c)}$ represents the ground truth labels for the key regions (e.g., 1 for key regions and 0 for other regions), and $A_i^{(p,c)}$ denotes the adaptive attention weights.

The segmentation loss term L_{seg} measures the difference between the segmentation mask and the ground truth labels. We can use the cross-entropy loss or other appropriate segmentation loss functions.

Taking into account both the attention loss term and the segmentation loss term, the loss function for the adaptive weights can be defined as:

$$L = \lambda_{att} \cdot L_{att} + \lambda_{seg} \cdot L_{seg} \tag{2}$$

Here, L_{att} and L_{seg} are weight factors for controlling the attention loss term and the segmentation loss term, respectively. These factors can be adjusted according to the specific task requirements.

By optimizing the above loss function for the adaptive weights, we can learn appropriate attention weights for the key regions, thereby improving the segmentation accuracy of the key objects and achieving better performance in the presence of class imbalance.

IV. EXPERIMENT

To evaluate the effectiveness of our method, we conduct extensive experiments in this section. We detail the adopted datasets, baseline algorithms, and experimental results in this section.

A. Dataset

Our goal was to localize microvascular structures (vessels) in human kidney histology slides. We choose the dataset for microvessel segmentation. It contains tiles extracted from five full slide images (WSI), each of size 512x512, divided into two datasets. The tiles in Dataset 1 have annotations reviewed by experts. Dataset 2 contains remaining tiles from these same WSIs and contains sparse annotations that have not been reviewed by experts. Two WSIs constitute the training set, two WSIs constitute the public test set, and one WSI constitutes the private test set. For the test set, all 600 tiles are from dataset 1. For the training set, the data includes Dataset 2 tiles from the public testing WSI, while excluding Dataset 2 tiles from the private testing WSI. Also included in this dataset are tiles extracted from another 9 WSIs as Dataset 3, which are not yet annotated. But for this study, we did not consider unsupervised information, so we did not use Dataset 3 as the research content.

B. Baseline

- Mask R-CNN [7] a conceptually simple, flexible and general framework for object instance segmentation. It is one of the most classic instance segmentation framework. Mask R-CNN extends Faster R-CNN by adding a branch for predicting object masks in parallel with the existing bounding box recognition branch. Mask R-CNN is simple to train, adds only a small overhead to Faster R-CNN, and runs at 5 fps.
- U-net [18] is a network and training strategy that relies on the brute force use of data augmentation to use available annotated samples more efficiently. The architecture consists of a contraction path that captures context and a symmetric expansion path that enables precise localization. This network is specially used for medical impact segmentation, but the accuracy of complex tiny targets may not be as good as traditional instance segmentation.
- Blendmask [19] achieves improved mask prediction by effectively combining instance-level information with semantic information with lower-level fine-grainedness. The framework adopts a mixer module, which draws inspiration from top-down and bottom-up instance segmentation methods. Therefore, it can efficiently predict dense per-pixel position-sensitive instance features with few channels, and learn a per-instance attention map with only one convolutional layer.
- Polarmask [20] is an anchor-free single-shot instance segmentation method that is conceptually simple, fully convolutional, and usable by easily embedding it into most off-the-shelf detection methods. It formulates the instance segmentation problem as predicting the contours of instances via instance center classification and dense distance regression in polar coordinates.
- Solov2 [21] is a simple, straightforward, fast and performant framework for instance segmentation. The framework is powered by an efficient and holistic instance mask representation scheme that dynamically segments each instance in an image without resorting to bounding box detection. Specifically, object mask generation is decoupled into mask kernel prediction and mask feature learning, which are responsible for generating convolution kernels and feature maps to be convoluted, respectively. At the

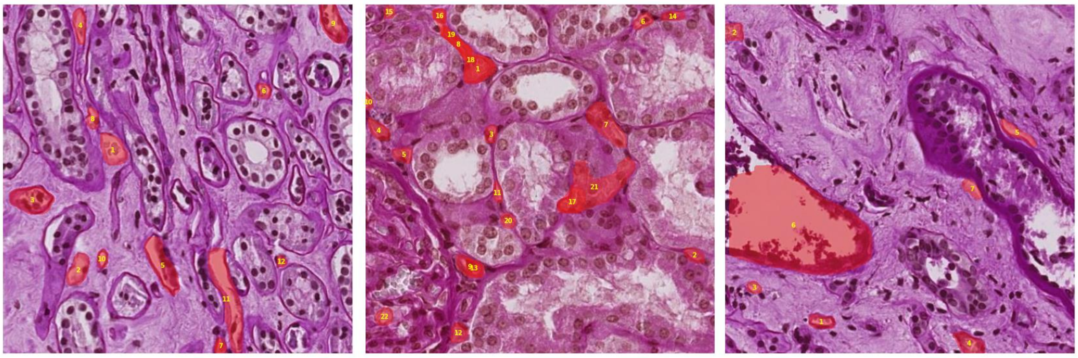

Fig. 2. Examples of sequential segmentation at different positions of the slice image in the microvasculature.

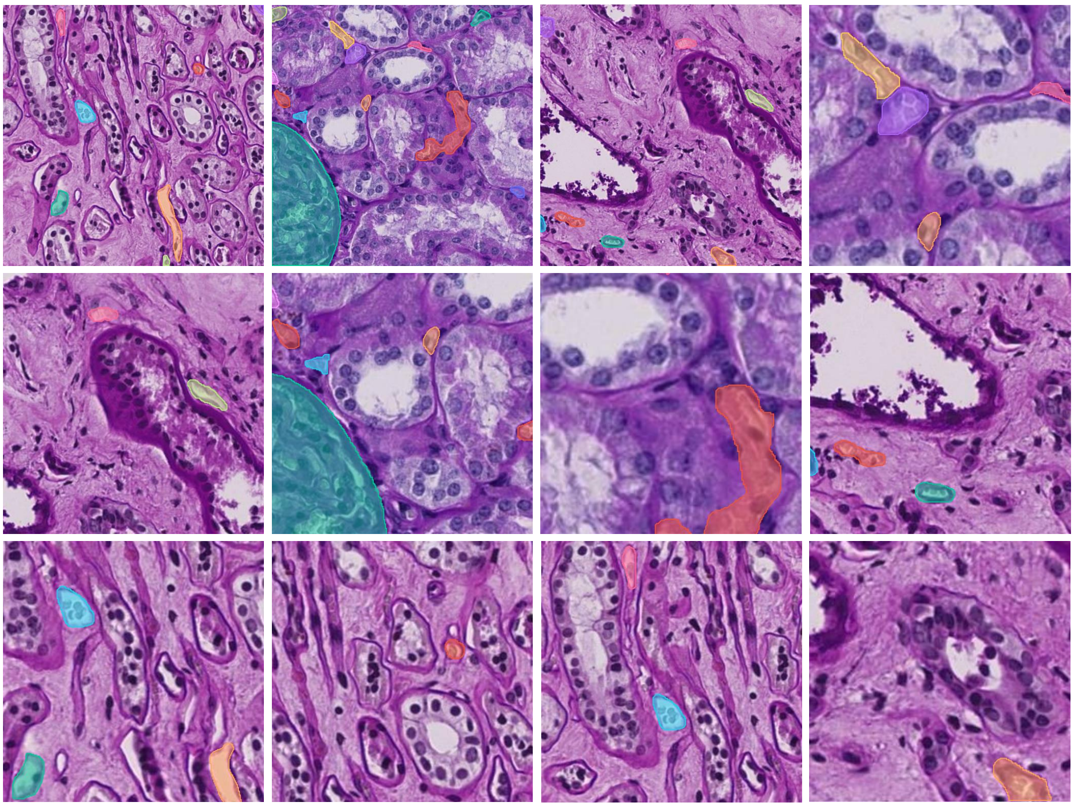

Fig. 3. The instance segmentation results of different categories of regions of the Weimaiguan system.

same time, it significantly reduces the inference overhead by utilizing the matrix non-maximum suppression (NMS) technique.

- Yolact [9] is a simple fully convolutional model for real-time instance segmentation. It greatly improves model efficiency by splitting instance segmentation into two parallel subtasks (1) generating a set of prototype masks and (2) predicting per-instance mask coefficients. We then incorporate fast NMS to generate instance masks by linearly combining the prototypes with the mask coefficients.
- Centermask [22] is a simple yet efficient anchor-free instance segmentation that, like Mask R-CNN, adds a novel spatial attention-guided mask (SAG-Mask) to anchor-free one stage object detector (FCOS) branch. After plugging into the FCOS object detector, the SAG-Mask branch uses the spatial attention map to predict a segmentation mask on each box, which helps to focus on informative pixels and suppress noise.

C. Result

Figure 2 and Figure 3 show some results of cell segmentation. Among them, Figure 2 shows an example of sequential segmentation at different positions of the slice image in the microvasculature. Figure 3 shows the instance segmentation results of different categories of regions of the Weimaiguan system. We can intuitively find that the proposed method has clear segmentation boundaries for different categories of regions, and achieves the expected goal well. In addition, in order to better quantify the effect of the model, we compared the model with the baseline algorithm. The results in Table 1 illustrate that our method is more stable than other algorithms.

V. Conclusion

Our study proposes a novel framework, Attention-based Mask R-CNN, for Microvascular Segmentation, which aims to address the challenges in human cell atlas mapping. Through structural and performance improvements, we successfully achieved automatic segmentation of microvasculature arrangements and achieved superior segmentation results in healthy human kidney tissue slices. This work is of great significance for a deeper understanding of the microvascular structure and function of the human body, and provides a powerful tool and basis for future medical research and diagnosis.

Acknowledgment

The completion of this paper would not have been possible without the contributions and support of various individuals whom I wish to acknowledge with gratitude. I extend our deepest appreciation to our co-authors, whose expertise and collaborative efforts significantly enriched the quality of this research.

I would like to express our heartfelt thanks to Zhao Henan, Yang Zekai, MoTong, and Yao Yiya for their dedicated engagement in the conception, execution, and analysis of this study. Their insightful discussions, diligent reviewing, and constructive feedback played an essential role in shaping the direction and scope of this work.

References

[1] P. C. Georges and P. A. Janmey, "Cell type-specific response to growth on soft materials," *Journal of applied physiology*, vol. 98, no. 4, pp. 1547–1553, 2005.

[2] M. Malumbres and M. Barbacid, "Cell cycle, cdks and cancer: a changing paradigm," *Nature reviews cancer*, vol. 9, no. 3, pp. 153–166, 2009.

[3] M. Zolfaghari and H. Sajedi, "A survey on automated detection and classification of acute leukemia and wbcs in microscopic blood cells," *CoRR*, vol. abs/2303.03916, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2303.03916

[4] J. E. Rood, T. Stuart, S. Ghazanfar, T. Biancalani, E. Fisher, A. Butler, A. Hupalowska, L. Gaffney, W. Mauck, G. Eraslan *et al.*, "Toward a common coordinate framework for the human body," *Cell*, vol. 179, no. 7, pp. 1455–1467, 2019.

[5] S. Kim, H. Lee, M. Chung, and N. L. Jeon, "Engineering of functional, perfusable 3d microvascular networks on a chip," *Lab on a Chip*, vol. 13, no. 8, pp. 1489–1500, 2013.

[6] W. Jingyao, Y. Naigong, and E. Firdaous, "Gesture recognition matching based on dynamic skeleton," in *2021 33rd Chinese Control and Decision Conference (CCDC)*. IEEE, 2021, pp. 1680–1685.

[7] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, 2020. [Online]. Available: https://doi.org/10.1109/TPAMI.2018.2844175

[8] J. Wang and N. Yu, "So-perm: Pose estimation and robust measurement for small objects," in *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022, pp. 1–7.

[9] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "Yolact: Real-time instance segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9157–9166.

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[11] O. Ronneberger, "Invited talk: U-net convolutional networks for biomedical image segmentation," in *Bildverarbeitung für die Medizin 2017 - Algorithmen - Systeme - Anwendungen. Proceedings des Workshops vom 12. bis 14. März 2017 in Heidelberg*, ser. Informatik Aktuell, K. H. Maier-Hein, T. M. Deserno, H. Handels, and T. Tolxdorff, Eds. Springer, 2017, p. 3. [Online]. Available: https://doi.org/10.1007/978-3-662-54345-0_3

[12] S. Gur, L. Wolf, L. Golgher, and P. Blinder, "Unsupervised microvascular image segmentation using an active contours mimicking neural network," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 10 722–10 731.

[13] H. Zhao, J. Huang, Q. Zhou, N. Chen, L. Liu, X. Wang, T. Wang, L. Chen, C. Liu, C. Zheng, and F. Yang, "Deep learning-based optical-resolution photoacoustic microscopy for in vivo 3d microvasculature imaging and segmentation," *Adv. Intell. Syst.*, vol. 4, no. 9, 2022. [Online]. Available: https://doi.org/10.1002/aisy.202200004

[14] J. Lo, M. Heisler, V. Vanzan, S. Karst, I. Z. Matovinovic, S. Loncaric, E. V. Navajas, M. F. Beg, and M. V. Sarunic, "Microvasculature segmentation and inter-capillary area quantification of the deep vascular complex using transfer learning," *CoRR*, vol. abs/2003.09033, 2020. [Online]. Available: https://arxiv.org/abs/2003.09033

[15] M. Chen, C. Zhao, X. Tian, Y. Liu, T. Wang, and B. Lei, "Placental super micro-vessels segmentation based on resnext with convolutional block attention and u-net," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 4015–4018.

[16] X. Lu, J. Zhang, T. Zhang, X. Zhang, J. Lu, X. Miao, W. Chen, J. Jiang, D. Ding, and S. Du, "Using pre-operative radiomics to predict microvascular invasion of hepatocellular carcinoma based on gd-eob-dtpa enhanced MRI," *BMC Medical Imaging*, vol. 22, no. 1, p. 157, 2022. [Online]. Available: https://doi.org/10.1186/s12880-022-00855-w

[17] S. Wu, S. Wu, H. Feng, Z. Hu, Y. Xie, Y. Su, T. Feng, and L. Li, "An optimized segmentation and quantification approach in microvascular imaging for octa-based neovascular regression monitoring," *BMC Medical Imaging*, vol. 21, no. 1, p. 13, 2021. [Online]. Available: https://doi.org/10.1186/s12880-021-00546-y

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[19] H. Chen, K. Sun, Z. Tian, C. Shen, Y. Huang, and Y. Yan, "Blendmask: Top-down meets bottom-up for instance segmentation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. Computer Vision Foundation / IEEE, 2020, pp. 8570–8578.

[20] E. Xie, P. Sun, X. Song, W. Wang, X. Liu, D. Liang, C. Shen, and P. Luo, "Polarmask: Single shot instance segmentation with polar representation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. Computer Vision Foundation / IEEE, 2020, pp. 12 190–12 199.

[21] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "Solov2: Dynamic and fast instance segmentation," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., 2020.

[22] Y. Lee and J. Park, "Centermask: Real-time anchor-free instance segmentation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. Computer Vision Foundation / IEEE, 2020, pp. 13 903–13 912.