2024 IEEE International Symposium on Biomedical Imaging (ISBI) | 979-8-3503-1333-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISBI56570.2024.10635729

GLOMNET: A HOVER DEEP LEARNING MODEL FOR GLOMERULUS INSTANCE SEGMENTATION

Noémie Moreau[1,2] *Michelle Shabani*[3,4] *Christoph Schell*[3,4] *Katarzyna Bozek*[1,2,5]

[1] Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany
[2] Center for Molecular Medicine Cologne, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany
[3] Institute of Surgical Pathology, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[4] Core Facility for Histopathology and Digital Pathology, Medical Center, University of Freiburg, Freiburg, Germany
[5] Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany

ABSTRACT

Glomeruli are essential kidney structures for blood filtration. Their damage can impact the filtering capability of the kidney, leading to its failure. Hence, glomerulus detection and evaluation are crucial for kidney disease diagnosis. Most deep learning methods for glomerulus segmentation focus on semantic segmentation and do not allow instances' separation. In this paper, we present GlomNet, a network for glomerulus instance segmentation. Our network is composed of one EfficientNet encoder and two decoders: one for binary output prediction and the other for the prediction of Horizontal and Vertical distances of object pixels to their centers of mass (HoVer Maps). During post-processing, these HoVer Maps are used to determine object boundaries and separate glomerular instances. Our method was trained and tested on 176 images from two clinical centers. Our network outperformed two state-of-the-art methods with a Dice score of 0.79 and a Panoptic Quality of 0.61.

Index Terms— Kidney tissue, Glomerulus segmentation, Instance segmentation

1. INTRODUCTION

Glomeruli are essential kidney structures that take part in the blood-filtering process. Their damage can impact the filtering capability of the kidney and therefore cause kidney failure in patients. Glomeruli detection and evaluation are crucial for kidney disease diagnosis. Patients presenting kidney failure symptoms undergo a kidney biopsy, then, pathologists manually assess each glomerulus to evaluate the kidney damage. This task is very tedious and time-consuming, yet essential to propose the most suitable treatment for each patient.

In recent years, several studies used deep learning methods for glomeruli detection and evaluation, but most authors focused only on semantic segmentation. For example, a SegNet-VGG19 model was used by [1], a standard U-Net by [2] and a DeepLab V2 model by [3]. Singh Samant et al. [4] and Gu et al. [5] also compared several network combinations including FCN-Resnet, Deeplab V3, EfficientUNet, and LinkNet. All of these networks are very performant and reached Dice scores around 0.90. However, in cases where glomeruli are close to each other as shown in Figure 1, these methods do not allow the separation of the glomerular instances. Consequently, semantic segmentation alone does not allow to correctly count the number of glomeruli in a single image.

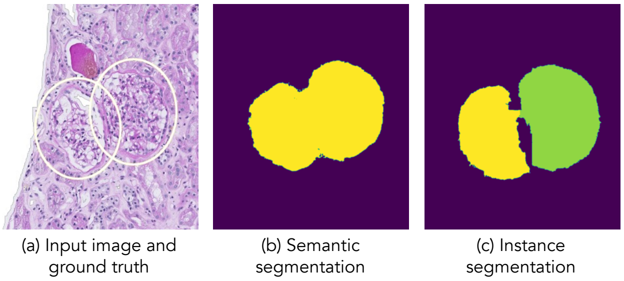
(a) Input image and ground truth
(b) Semantic segmentation
(c) Instance segmentation

Fig. 1. Comparison of result for two glomeruli joined together. (a) Input image with the ground truth annotations (Cologne dataset: precise contours are not available). (b) Semantic segmentation with an EfficientUNet: no separation between the two instances. (c) Instance segmentation with GlomNet: separation between the two instances.

To the best of our knowledge, only Jiang et al. [6] study focused on instance segmentation. They used a Cascade Mask R-CNN architecture that simultaneously predicts the instance's bounding boxes and does the segmentation of each instance. This network is a common architecture for natural image instance segmentation. A major limitation of this method is the difficulty of stitching the instances' predictions between the different tiles of the image, as Whole Slide Images (WSIs) are large images that do not fit the network input size.

In this paper, we present GlomNet, a network for glomerulus instance segmentation. Inspired by Graham et al. [7], we used the prediction of horizontal and vertical distance maps to compute glomerulus boundaries and separate the different instances. We demonstrate the superiority of our network in comparison with a standard EfficientUNet for semantic segmentation and with the Cascade Mask R-CNN developed by Jiang et al. [6].

2. METHODS

2.1. Dataset

A total of 176 WSIs coming from two centers were used in this study.

44 WSIs came from 26 patients from the Freiburg University Hospital. One patient had Autosomal dominant polycystic kidney disease, 10 had renal cancer, and 15 had diverse nephrotic glomerular diseases. Each sample was embedded in formaldehyde-fixed paraffin and stained using Periodic Acid Schiff. Images were acquired using the VENTANA DP 200 software and a VENTANA DP 200, Roche Diagnostics scanner with a 20x objective. Glomerulus contour annotations were performed using QuPath by one annotator and were reviewed by an expert pathologist.

132 WSIs came from 13 patients from the Cologne University Hospital part of the ForMe registry cohort (NCT03949972). All patients were presenting a nephrotic syndrome: either Minimal Change Disease (MCD) or Focal and Segmental GlomeruloSclerosis (FSGS). Each sample was fixed with 4% unbuffered formalin and stained using Periodic Acid Schiff. Images were acquired with a 40x magnification using NZacquire software and a Hamamatsu NanoZoomer S360 scanner with a 20x objective. Annotations were performed by one expert pathologist on the Omero platform by drawing an ellipse around each glomerulus. Precise contours were not available for this part of the dataset.

In total, 1431 glomeruli were annotated, 534 on Freiburg WSIs and 897 on Cologne WSIs.

2.2. Network Architecture

The base architecture of our network is a standard UNet with an EfficientNet encoder. UNet is an encoder-decoder network first developed by Ronneberger et al. [8] in 2015 before being improved by several authors over the years [9]. It is the most common architecture used for semantic segmentation of medical images and was already used in several articles for glomerulus semantic segmentation [2, 5, 4]. EfficientNet is a convolutional neural network developed by Tan and Le [10] that uses a compound coefficient technique, which uniformly adjusts the network depth, width, and resolution to scale up the model. The authors developed several EfficientNet variants of different dimensions, which surpassed in accuracy most state-of-the-art Convolutional Neural Networks (CNNs) with better efficiency. The UNet and EfficientNet architectures can be combined into an EfficientUNet: a UNet with an EfficientNet encoder. This architecture was already used for the semantic segmentation of glomerulus in two papers and reached the highest scores [5, 4]. We chose the deeper variant of the EfficientNet, EfficientNet B7, as a deeper network usually shows a stronger representative capability.

This architecture is only suitable for semantic segmentation, therefore, we introduced modifications to expand its functionality to instance segmentation. In natural images, most instance segmentation methods use two-stage instance segmentation models like Mask R-CNN models that detect and then segment each instance individually [11]. For large images like WSIs, a major limitation of the two-stage approach is the difficulty of stitching the instance predictions between the different tiles of the image a posteriori [6]. To overcome this problem, we were inspired by the work of Graham et al. [7] on nuclear instance segmentation. They developed a UNet with a second decoder that predicts the Horizontal and Vertical distance of object pixels to their centers of mass (”HoVer Maps”). After inference on tiles, these HoVer Maps can be easily stitched by taking the prediction's average of each overlapping tile and then used to compute object boundaries and separate instances during postprocessing. Following this idea, we added a second decoder to our EfficientUNet to predict the horizontal and vertical distance of glomerulus pixels from their center of mass. The final architecture of GlomNet can be visualized in Figure 2.

Our architecture organization is similar to HoVerNet [7] but is adapted to the specific glomerulus' characteristics with a bigger input size and a deeper encoder. Moreover, unlike cells, glomerulus can exceed the size of a single patch, therefore our ground truth HoVer Maps were computed for the entire WSI and not at a patch level, as done by HoVerNet.

2.3. Loss function

As in [7], we used a loss function with two distinct parts, one for each branch: $L = L_{Bin} + L_{HoVer}$

For the binary segmentation branch loss (L_{Bin}), a combination of the cross-entropy loss (L_{CE}) and the Dice loss (L_{Dice}) was used: $L_{Bin} = 0.2 \times L_{Dice} + 0.8 \times L_{CE}$

For the HoVer Maps branch, we used the same loss (L_{HoVer}) as [7]: a combination of the mean square error loss (L_{MSE}) of the predicted HoVer Maps and the ground truth HoVer Maps and the mean square error loss ($L_{GradMSE}$) of their respective gradient: $L_{HoVer} = L_{MSE} + L_{GradMSE}$

2.4. Postprocessing

Within the Horizontal and Vertical distances, pixel values between two instances have a high difference. This property is used to separate glomeruli.

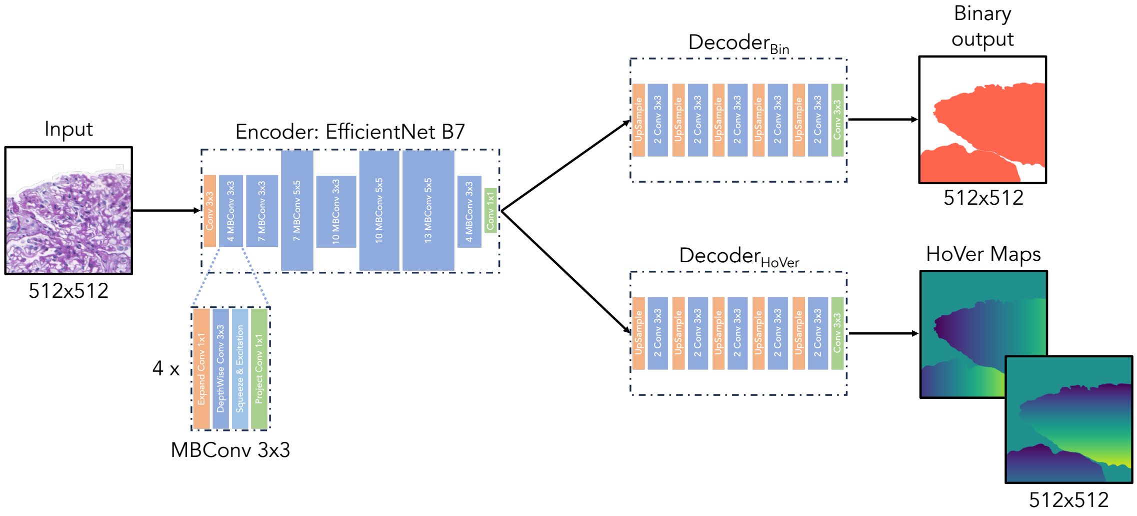

Fig. 2. Network architecture overview: UNet with an EfficientNet encoder and two decoders.

Specifically, we computed the horizontal and vertical derivatives using a Sobel operator to extract glomeruli boundaries. These boundaries were, then, subtracted from the binary mask, and the obtained result was used to compute the Euclidean Distance Transformed (EDT) map. Finally, we extracted the local peaks in the EDT map and used them as markers for the watershed algorithm to determine how to split the mask into instances, given the inverse of the EDT map as the energy landscape.

2.5. Evaluation

2.5.1. *Semantic segmentation metrics*

The Dice Similarity Coefficient (DSC) or F1-score and the Intersection over Union (IoU) or Jaccard Index are the most common semantic segmentation metrics [12]. We used these scores to evaluate the overlap between the ground truth segmentation (GT) and the prediction ($Pred$).

2.5.2. *Instance segmentation metrics*

Instance segmentation is composed of two tasks: object detection and semantic segmentation. The Panoptic Quality (PQ) is well suited for its evaluation, as it combines the assessment of overall detection performance, the Detection Quality (DQ) and the Segmentation Quality (SQ) of True Positive (TP) instances in a single score [12]:

$$PQ = DQ \times SQ$$

The Detection Quality is a F1-score where TP are overlapping GT and $Pred$ instances with an $IoU \geq 0.5$, False Positives (FP) are unmatched $Pred$ instances, and False Negatives (FN) are unmatched GT instances.

The Segmentation Quality is the mean IoU of correctly detected instances.

3. EXPERIMENTS AND RESULTS

3.1. Training process

The dataset was divided into three subsets: a training set, a validation set, and a testing set. As some slides were consecutive, slides belonging to the same patient were grouped in the same set to avoid any bias. From the Freiburg dataset, 24 WSIs were used for training, 11 for validation, and five for testing. From the Cologne dataset, 49 WSIs were used for training, 24 for validation, and 59 for testing.

Slides were divided into 512×512 patches at 20x magnification, with 75% overlap. To reduce the false segmentation of small structures that can look like glomerulus, patches with a tissue pixel volume $< 25\%$ or with a glomerular pixel volume $< 10\%$ were discarded. Moreover, to have a better-balanced training set between the background and glomerulus classes, we randomly selected only two background patches for one foreground patch. In the end, 40 920 patches were used for training.

The GlomNet was trained on a Tesla V100 SXM2 32 GB GPU for 1 epoch with a batch size of 8 (5 115 iterations). The AdamW optimizer was used with a learning rate of 0.0001 and a weight decay of 0.001. The training was performed during only one epoch, as after more epochs, results on the Freiburg dataset were degraded because of the lack of precise contours in the Cologne dataset.

Inference was performed using a sliding window method with a 25% overlap, and values in the overlapping part were obtained by the average of each prediction.

Table 1. Quantitative results on the entire testing set (59 Cologne WSIs + 5 Freiburg WSIs). DSC = Dice Similarity Coefficient, IoU = Intersection over Union, DQ = Detection Quality, SQ = Segmentation Quality, PQ = Panoptic Quality. Between parenthesis are the results for the Freiburg subpart of the testing set, composed of five images. Significantly better results are in **bold** (p-value < 0.005 calculated with paired t-test).

Methods	DSC	IoU	DQ	SQ	PQ	Mean inference + postprocessing time
EfficientUNet	0.69 (0.86)	0.58 (0.76)	0.66 (0.76)	0.64 (0.89)	0.49 (0.68)	32s
Cascade Mask R-CNN	**0.80** (0.86)	**0.68** (0.76)	0.63 (0.74)	**0.82** (0.89)	0.52 (0.66)	257s
GlomNet	**0.79** (0.89)	**0.67** (0.80)	**0.78** (0.81)	0.77 (0.89)	**0.61** (0.72)	104s

The implementation was done using Monai framework [13]. All models and source-code are available online: `https://github.com/bozeklab/GlomNet`

3.2. Results

We compared our proposed model with a standard EfficientUNet for semantic segmentation with the same parameters described in section 3.1. To extract the different instances, we separated the binary segmentation into connected components and removed the smallest ones for a better result.

We also did a comparison with the method proposed in [6] for instance segmentation. We reproduced their method on our dataset by using the same parameters as in their article but the newest version of the MMDetection ToolBox (`https://github.com/open-mmlab/mmdetection`, last accessed October 24, 2023). After inference, image tiles were stitched back to the WSI format: if two instances from two overlapping patches had an $IoU \geq 0.5$, they were considered the same instance.

Quantitative results of the three methods on the entire testing set are presented in Table 1. As the absence of precise contours for the Cologne dataset degraded the segmentation results, we also did the evaluation on only the Freiburg subpart of the testing set, composed of five images. These results are presented between parenthesis for each method and metric.

For semantic segmentation (DSC and IoU), GlomNet obtained significantly better results than the plain EfficientUNet and comparable results with the Cascade Mask R-CNN on the complete test set. Results on the Freiburg subset confirmed its comparability with the two other networks.

For instance segmentation, GlomNet obtained significantly better results than the two other methods for the Detection Quality (DQ) and for the final Panoptic Quality (PQ). However, the Segmentation Quality (SQ) was significantly better for the Cascade Mask R-CNN. Results on the Freiburg subset confirm its better results than the two other methods, with a better DQ and PQ and a similar SQ.

Example results for one slide are presented in Figure 3. GlomNet successfully separated two close glomerular instances, but the Cascade Mask R-CNN failed.

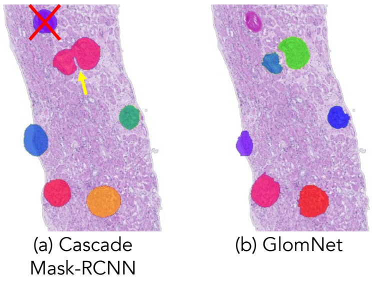

Fig. 3. Glomerulus instance segmentation results. (a) Cascade Mask R-CNN result with a false positive segmentation (red cross) and a joined segmentation of two instances (yellow arrow). (b) GlomNet result.

4. DISCUSSION AND CONCLUSION

In this paper, we presented GlomNet, a network for glomerulus instance segmentation. Inspired by Graham et al. [7], we added a second decoder to the EfficientUNet architecture for the prediction of horizontal and vertical distance maps to compute glomerulus boundaries and separate the different instances.

Our results showed that adding a second decoder for the prediction of horizontal and vertical distance is a suitable architecture for instance segmentation on WSI images. Indeed, our network reached the best DQ and PQ which indicates a better instance separation than the two other methods, while obtaining comparable results for semantic segmentation.

Moreover, the post-processing process of our model is less intricate than with the Cascade Mask R-CNN used by Jiang et al. [6]. The tile stitching process of the Cascade Mask R-CNN required an extensive calculation of instances' overlaps between the tiles when, for GlomNet, only an average of each tile prediction was necessary. Hence, total inference and post-processing time was reduced by a factor of two.

In summary, we demonstrated the superiority of our network for the glomerulus instance segmentation in comparison with a standard EfficientUNet for semantic segmentation and a Cascade Mask R-CNN for instance segmentation.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. Analysis of patient samples from Freiburg was approved by the Institutional Ethics Committee of the University Medical Center Freiburg (EK 21/1288; 18/512). WSIs from Cologne University Hospital are part of the ForMe registry cohort (NCT03949972) which was approved by the Ethics Committee of Cologne University. Written informed consent was obtained from all patients.

6. ACKNOWLEDGMENTS

This work was supported by the Clinical Research Unit 329. Both K.B. and N.M. were hosted by the Center for Molecular Medicine Cologne throughout this research. K.B. was supported by the BMBF program Junior Group Consortia in Systems Medicine (01ZX1917B), BMBF program for Female Junior Researchers in Artificial Intelligence (01IS20054) and the NRW return program. M.S has no relevant financial or non-financial interests to disclose. C.S. was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-IDs 241702976, 438496892, 501370692 and project-ID 386793560 – CRU 329 to C.S. WSIs from Cologne University Hospital were collected thanks to the FOrMe registry/KFO329 funding.

7. REFERENCES

[1] Gloria Bueno, M Milagro Fernandez-Carrobles, Lucia Gonzalez-Lopez, and Oscar Deniz, "Glomerulosclerosis identification in whole slide images using semantic segmentation," *Computer methods and programs in biomedicine*, vol. 184, pp. 105273, 2020.

[2] Xiang Li, Richard C Davis, Yuemei Xu, Zehan Wang, Nao Souma, Gina Sotolongo, Jonathan Bell, Matthew Ellis, David Howell, Xiling Shen, et al., "Deep learning segmentation of glomeruli on kidney donor frozen sections," *Journal of Medical Imaging*, vol. 8, no. 6, pp. 067501–067501, 2021.

[3] Giovanna Maria Dimitri, Paolo Andreini, Simone Bonechi, Monica Bianchini, Alessandro Mecocci, Franco Scarselli, Alberto Zacchi, Guido Garosi, Thomas Marcuzzo, and Sergio Antonio Tripodi, "Deep learning approaches for the segmentation of glomeruli in kidney histopathological images," *Mathematics*, vol. 10, no. 11, pp. 1934, 2022.

[4] Surender Singh Samant, Arun Chauhan, Jagadish Dn, and Vijay Singh, "Glomerulus detection using segmentation neural networks," *Journal of Digital Imaging*, pp. 1–10, 2023.

[5] Ye Gu, Ruyun Ruan, Yan Yan, Jian Zhao, Weihua Sheng, Lixin Liang, and Bingding Huang, "Glomerulus semantic segmentation using ensemble of deep learning models," *Arabian Journal for Science and Engineering*, vol. 47, no. 11, pp. 14013–14024, 2022.

[6] Lei Jiang, Wenkai Chen, Bao Dong, Ke Mei, Chuang Zhu, Jun Liu, Meishun Cai, Yu Yan, Gongwei Wang, Li Zuo, et al., "A deep learning-based approach for glomeruli instance segmentation from multistained renal biopsy pathologic images," *The American Journal of Pathology*, vol. 191, no. 8, pp. 1431–1441, 2021.

[7] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical image analysis*, vol. 58, pp. 101563, 2019.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas 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.

[9] Nahian Siddique, Paheding Sidike, Colin Elkin, and Vijay Devabhaktuni, "U-net and its variants for medical image segmentation: theory and applications," *arXiv preprint arXiv:2011.01118*, 2020.

[10] Mingxing Tan and Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[11] Wenchao Gu, Shuang Bai, and Lingxing Kong, "A review on 2d instance segmentation based on deep neural networks," *Image and Vision Computing*, vol. 120, pp. 104401, 2022.

[12] L Maier-Hein, A Reinke, P Godau, M Tizabi, F Büttner, E Christodoulou, B Glocker, F Isensee, J Kleesiek, M Kozubek, et al., "Metrics reloaded: Recommendations for image analysis validation. arxiv 2023," *arXiv preprint arXiv:2206.01653*, 2023.

[13] M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al., "Monai: An open-source framework for deep learning in healthcare," *arXiv preprint arXiv:2211.02701*, 2022.