How feasible is it to segment human glomerulus with a model trained on mouse histology images?

Luiz Souza[1,6], **Jefferson Silva**[1,5], **Paulo Chagas**[1],
Angelo Duarte[3], **Washington LC dos-Santos**[2,4], **Luciano Oliveira**[1]

[1]IVISION Lab, Universidade Federal da Bahia, Bahia, Brazil

[2]Universidade Federal da Bahia, Bahia, Brazil

[3]Universidade Estadual de Feira de Santana, Bahia, Brazil

[4]Fundação Oswaldo Cruz – Instituto Gonçalo Moniz, Bahia, Brazil

[5]Universidade Federal do Maranhão, Maranhão, Brazil

[6]Instituto Federal do Maranhão, Maranhão, Brazil

{luiz.otavio, paulo.chagas, lrebouca}@ufba.br

jefferson.fs@ufma.br

angeloduarte@uefs.br

wluis@bahia.fiocruz.br

Abstract. *Many genetic, physiological and structural characteristics of internal organs are shared by mice and humans. Hence, mice are frequently used in experimental model of human diseases. Although this is an indisputable truth in medicine, there is an avenue to go in computational pathology, where digital images are the main objects of investigation. Considering the lack of study about knowledge transfer between mice and humans concerning machine learning models, we propose investigating if it is possible to segment glomeruli in human WSIs by training deep learning models on mouse data only. Three different semantic segmenters were evaluated, which had their performance assessed on two data sets comprised of 18 mouse WSIs and 30 human WSIs. The results found corroborate our hypothesis validation.*

1. Introduction

A common approach in scientific medical studies is to perform animal experiments before clinical tests in humans. Some of these studies includes histological analysis of representative tissue samples. Mice are frequently selected for these animal studies, not only for their genetic and physiological similarities with humans [Kim et al., 2007], but also because the rodents are inexpensive to breed and their tiny size takes less space in the laboratory facilities [Smith and Corrow, 2005]. That preemptive mouse experimentation also benefits the nephropathology, which is the domain of this work. Kidney biopsies may be required for diagnosing renal diseases. The renal samples are processed and scanned into whole-slide images (WSI), which are easier to share and analyze [Barisoni et al., 2013, Farris et al., 2017, Santos et al., 2019].

We define the glomerulus as a network formed by tiny blood capillaries responsible for blood filtration. Given its primary function, the location and segmentation of glomeruli are valuable information that pathologists often extract from a kidney WSI. As locating glomeruli is time-consuming and error-prone, a promising alternative arises with developing an automatic glomerular segmentation approach, providing fast and reliable supportive information for the pathologists' decision-making pipeline. In this context, a challenge comes out in gathering a large amount of annotated data, which can be even harder to obtain if one considers human biopsies. Bearing all that in mind, a question naturally arises: *how feasible is it to segment human glomerulus with a model trained on mouse histology images?*

When addressing automatic segmentation, deep learning techniques have stood out in the literature, achieving state-of-the-art results in several domains, including medical imaging and, more specifically, WSI segmentation task [de Bel et al., 2018, Gadermayr et al., 2019, Ginley et al., 2020, Jiang et al., 2021]. Nonetheless, deep learning techniques require a reasonable amount of data to train the model from scratch or to have it fine-tuned to a specific domain. This limitation motivated our study.

Although there is some works using mouse and human species outside the domain of histology [Chater et al., 2021, Hossain et al., 2021], only a few deep learning studies on histological images have been exploring samples of these species. Bouteldja et al. [2021] developed a custom U-Net network for automated multi-class segmentation of glomerular images of different mammalian species, not only mice and humans. In another study, Simon et al. [2018] modified local binary patterns (LBP) feature extractor to train a support vector machine (SVM) model for glomerulus detection in WSIs. For classification and segmentation, Ginley et al. [2019] proposed a deep-learning-based approach to quantify some kidney structures (nuclei, capillary lumina, and Bowman spaces); the authors defined a set of features that describe the structural progression of diabetic nephropathy, feeding a recurrent neural network for classification; finally, a DeepLabV2 network is used for glomerular segmentation on WSIs. In another approach, [Lutnick et al., 2019] applied a DeepLabV2 network to segment glomerulus and internal glomerular structures from human and mouse renal tissue slides.

From the histology perspective, it is worth noting that there is no obstacle to mixing mouse and human samples. The renal histological structures are similar across these species, despite the difference in size. So far related studies have addressed human and mouse data sets in isolated (same specie for training and testing) or combined evaluation (mix the species for both training and testing). Considering all that, the main contribution of our paper is the investigation of the cross-species compatibility of human and mouse data for glomerulus segmentation. In words, our main goal is to investigate if a deep learning model, training on mouse data, is able to segment human glomerulus.

2. Materials and methods

The analytical protocol depicted in Fig. 1 shows the outline of the proposed study split as follows: (i) glomerulus annotation, (ii) patch generation, (iii) training of the selected architectures, (iv) model prediction, and (v) stitching the predicted patches. **The first step** consists in the extraction of kidney biopsy sections with $40\times$ magnification and stained with hematoxylin and eosin (HE), periodic acid-Schiff (PAS), and periodic acid-

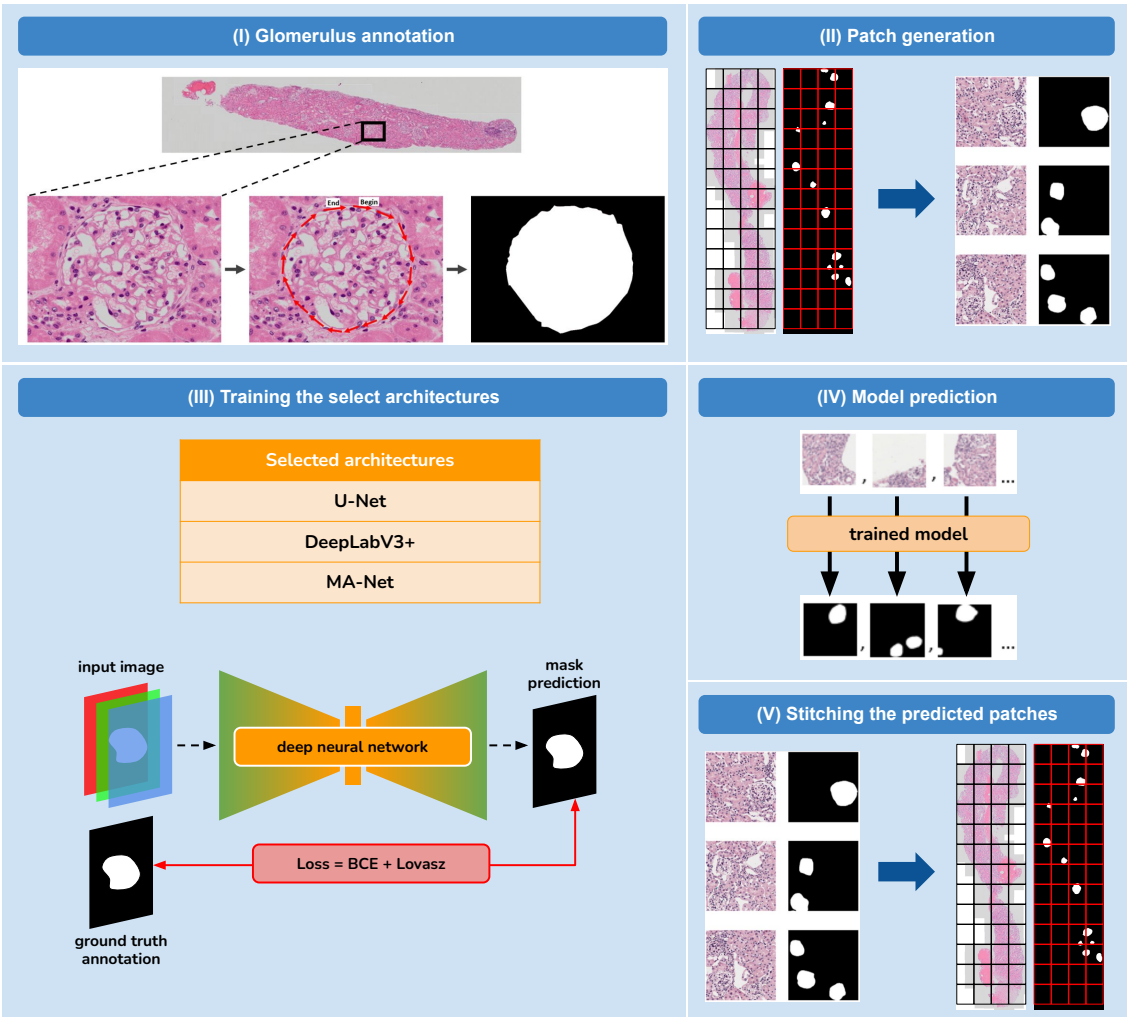

Figure 1. Protocol proposed for this study. From (i) to (v): glomeruli are manually annotated from sections of kidney biopsies; each section is divided into patches; three different types of architectures are selected for training; the best model of each architecture is selected to predict glomerulus or non-glomerulus pixel in each patch; finally, the resulting masks are stitched to compose the final segmented WSI mask.

methenamine silver (PAMS), subsequently the Cytomine[1] software was used to perform manual annotation of the glomeruli to support the training of the network models. **In the second step**, the WSIs were divided into patches of 1024×1024 pixels with padding size of 256 pixels; each patch is resized to 320×320 pixels due to high-memory footprint with the goal of increasing training speed. **In the third step**, the generated patches are used to train the following networks: U-Net [Ronneberger et al., 2015], DeepLabV3+ [Chen et al., 2018], and MA-Net [Fan et al., 2020], all using the EfficientNet-b1 [Tan and Le, 2019] encoder. **In the fourth step**, the best prediction model was selected for each architecture, and the segmentation performance was assessed by means of the dice coefficient (Dice) metric. Finally, **in the fifth step**, the resulting patches are simply stitched to compound the final semantic segmentation mask.

[1]https://cytomine.com/

Table 1. Summary of the data sets. Lutnick's data set comprises only mouse data, while WSI_Fiocruz 2's only human data.

Data set	Stain	# WSI		# Glom.		Average dimension (pixels)
		Train	Test	Train	Test	
Lutnick's	HE	14	4	634	171	19,511 × 20,044
WSI_Fiocruz 2's	HE	6	4	78	72	23,216 × 10,356
	PAS	6	4	95	40	22,566 × 10,972
	PAMS	6	4	86	53	22,706 × 11,691

2.1. Data sets

The collection of WSIs used in this study originated from two data sets: Lutnick's and WSI_Fiocruz 2's. Lutnick's data set contains 18 kidney sections from mice, while WSI_Fiocruz 2's contains 30 kidney sections from humans. Table 1 summarize the details of the two data sets.

Lutnick's data set. According to [Lutnick et al., 2019], all histological sections were collected by following protocols approved by the Institutional Animal Care and Use Committee at the University at Buffalo, obeying the guidelines and specifications of the American Veterinary Medical Association guidelines on euthanasia. C57BL/6J mice were euthanized, and their kidneys were perfused, extracted, and embedded in paraffin. Mice were either treated with streptozotocin (STZ) to induce diabetic nephropathy or with an STZ vehicle for control. The renal biopsies were sliced from the paraffin-embedded kidney. Sections at 2 μm thick were stained with HE and bright-field imaged at 0.25 μm per pixel resolution using an Aperio ScanScope slide scanner (Leica Biosystems) in 40$\times$ magnification. This data set consists of 18 HE-stained images.

WSI_Fiocruz 2's data set. The human data set consists of histopathology images of kidney WSIs, 30 gigapixels in size. Each WSI contains 3 to 6 sections of kidney biopsies. This digital image library was built by Dr. Washington LC dos-Santos at Gonçalo Moniz Institute (FIOCRUZ), for the diagnosis of glomerular diseases in referral nephrology services of public hospitals of Bahia State, Brazil. The renal biopsies were fixed in formalin-acetic acid-alcohol to preserve their histological structure, later included in paraffin. Sections of 2 μm were stained by HE, PAS, and PAMS. The WSIs were captured using a VSI-500 Olympus scanner with 40$\times$ magnification. The selected images contain 10 WSIs stained with HE, 10 WSIs stained with PAS, and 10 WSIs stained with PAMS.

2.2. Semantic segmentation networks

We selected three types of segmentation networks: U-Net [Ronneberger et al., 2015], DeepLabV3+ [Chen et al., 2018], and MA-Net [Fan et al., 2020]. Each network relies on a different deep learning segmentation paradigm. The U-Net is a typical encoder-decoder architecture, in which the encoder extracts features of different spatial resolutions, and the decoder receives these features as input to define the segmentation mask; each layer of the encoder is connected with each layer of the decoder, forming a U shape.

Table 2. Intra-dataset comparative results, considering three types of segmentation networks over the train and test data sets described on Table 1.

Data set	Stain	Network	μDice
Lutnick's	HE	U-Net	0.847($\pm$0.078)
		DeepLabV3+	**0.875**($\pm$**0.042**)
		MA-Net	0.773($\pm$0.098)
WSI_Fiocruz 2's	HE	U-Net	0.791($\pm$0.046)
		DeepLabV3+	**0.818**($\pm$**0.053**)
		MA-Net	0.014($\pm$0.004)
	PAS	U-Net	0.581($\pm$0.176)
		DeepLabV3+	**0.789**($\pm$**0.033**)
		MA-Net	0.050($\pm$0.019)
	PAMS	**U-Net**	**0.718**($\pm$**(0.130)**
		DeepLabV3+	0.660($\pm$0.162)
		MA-Net	0.334($\pm$0.109)

The DeepLabV3+ is a deep learning architecture designed for semantic image segmentation, created from improvements on DeepLabV3. Even though the encoder can process multi-scale contextual information by applying dilated convolution at multiple scales, DeepLabV3+ was not properly designed to obtain long-range dependencies in the deep learning process. The multi-scale attention net (MA-Net) is a network architecture that consists of a self-attention mechanism for adaptive feature extraction using two stages: (i) position-wise attention block (PAB), which covers feature inter-dependencies between pixels in spatial dimensions and (ii) a multi-scale fusion attention block (MFAB), which captures the channel dependencies between any feature map by multi-scale semantic feature fusion.

2.3. Implementation details of the network architectures

U-Net, DeepLabV3+, and MA-Net architectures were implemented using the Pytorch framework [Paszke et al., 2019] version 1.9.1, initially loading all networks with weights pre-trained on the Imagenet-1k data set [Russakovsky et al., 2015]. The models were trained across 50 epochs with a batch size of 16, a warm-up learning rate scheduler with a maximum of 0.0001, and weight decay of 0.00001, using a loss that combines binary cross-entropy (BCE) [Yi-de et al., 2004] and Lovasz [Berman et al., 2018] loss function. All experiments were run in a computer with AMD EPYC 7742 64-Core Processor, 1TB RAM, and an A100-SXM4 NVIDIA GPU containing 40GB of memory. In addition, the time spent took about 3 hours to train and test sets for three selected architectures. The pre-trained weights were fine-tuned on the top layers only, aiming to keep the rich features learned previously. In order to improve the variability of the input data, we used the following online training data augmentation techniques: resizing, vertical and horizontal flip, random rotation in intervals of 90 degrees, shift scale rotation, Gaussian noise, Gaussian blur, random brightness contrast, random hue saturation value, optical distortion, grid distortion, and piece-wise affine.

3. Experimental results

The sizes of the train and test sets were thought to keep the test set size similar for both human and mouse evaluations, thus assessing the different domains on the same quantity

of images (check Table 1). As the human data set contains two sections per patient out of 5 patients, we decided to group 3 patients (6 WSIs) in the training set and 2 patients (4 WSIs) in the test set. We also adopted 4 WSIs for Lutnick's test set, thus leaving 14 WSIs for training. The selection of the WSI splits on both data sets was adequately randomized to avoid selection bias.

Our first analysis was based on an intra-dataset evaluation, which consisted of training and testing the models on the same data set. The goal was to provide a baseline to compare when training on mouse data while testing on human data. Table 2 summarizes our findings considering the three selected semantic segmentation networks. It is noteworthy that most networks achieved satisfactory results on the mouse and human data sets, independently, especially when evaluating with the DeepLabV3+, which returned the highest marks on the Lutnick's test set (0.875), and on HE- and PAS-stained WSI_Fiocruz 2's test set (0.818). The U-Net architecture was the very best model on PAMS-stained WSI_Fiocruz 2's test set (0.718), although it reached results close to DeepLabV3+ on the other stains. As attention-based models require a large amount of high-variability data, MA-Net only performed well on Lutnick's test set (0.773) because the mouse training data is the largest and more diverse one.

Table 3. Comparative results considering training the network models on *Lutnick's train set* (TL), which contains only HE-based images, and on *WSI_Fiocruz 2's train set* (TA), which contains HE, PAS and PAMS. Prediction was performed on WSI_Fiocruz 2's test set (see Table 1 for data set split information).

Stain	Network	μDice	
		(**TL**)	(**TA**)
HE	U-Net	0.785(±0.050)	**0.791**(±**0.046**)
	DeepLabV3+	0.773(±0.069)	**0.818**(±**0.053**)
	MA-Net	**0.804**(±**0.037**)	0.014(±0.004)
PAS	U-Net	0.362(±0.415)	**0.581**(±**0.176**)
	DeepLabV3+	0.606(±0.142)	**0.789**(±**0.033**)
	MA-Net	**0.370**(±**0.358**)	0.050(±0.019)
PAMS	U-Net	0.664(±0.113)	**0.718**(±**(0.130)**
	DeepLabV3+	**0.687**(±**0.168**)	0.660(±0.162)
	MA-Net	**0.548**(±**0.193**)	0.334(±0.109)

Given the results found in the intra-dataset evaluation, we could make sure that the selected models performed the glomerular segmentation from our current data sets regardless of the numerical differences. As a matter of fact, if we had more data for training, surely it would guarantee higher scores as the data trend has shown in Table 2. Then we moved forward to the main focus of this work: the cross-species experiments.

As we ultimately desire to segment human glomeruli, the following evaluation was to compare the results of the deep learning models, separately trained on human and mouse training sets, testing these models on the WSI_Fiocruz 2's test set. Table 3 summarizes the evaluation on the WSI_Fiocruz 2's test set, considering models trained on Lutnick's (L) and WSI_Fiocruz 2's train (TA) sets. The overall best result was achieved by a TA-trained DeepLabV3+ evaluated on HE-stained WSI_Fiocruz 2's test set (0.818). This achievement is justifiable due to the domain maintained across training and test sets and DeepLabV3+'s robustness. However, one good point to highlight is that the second-best

Table 4. Comparative results considering training the network models with Lutnick's train set (stained on HE). Prediction was performed on *WSI_Fiocruz 2's test set* (A_{test})[1] and *WSI_Fiocruz 2's train and test sets* (A_{entire}) (see Table 1 for data set split information).

Stain	Network	μDice	
		(A_{test})	(A_{entire})
HE	U-Net	0.785(±0.050)	**0.810**(±**0.047**)
	DeepLabV3+	**0.773**(±**0.069**)	0.760(±0.039)
	MA-Net	**0.804**(±**0.037**)	0.707(±0.068)
PAS	U-Net	0.362(±0.415)	**0.625**(±**0.149**)
	DeepLabV3+	**0.606**(±**0.142**)	0.549(±0.174)
	MA-Net	0.370(±0.358)	**0.522**(±**0.154**)
PAMS	U-Net	**0.664**(±**0.113**)	0.656(±0.186)
	DeepLabV3+	0.687(±0.168)	**0.694**(±**0.186**)
	MA-Net	0.548(±0.193)	**0.564**(±**0.197**)

[1]Note that this column comes from Table 3.

result was achieved by an L-trained MA-Net evaluated on HE-stained WSI_Fiocruz 2's test set (0.804). Considering the color stains, Lutnick's training set containing only HE-stained images explains the stable and competitive results on HE-stained WSI_Fiocruz 2's test set. We can also observe other reasonable results in different stains and training sets. Still, the HE-stained samples presented the high scores between humans and mice across all architectures for glomerular segmentation. This makes us validate our hypothesis to segment human glomerulus with a model trained on mouse histology images, at least considering a reasonable amount of mouse data stained on HE. Overall, the TL-trained models outperformed the TA-trained models on 4 out of 9 tests, which is a favorable result but not enough to conclude that our hypothesis applies to all stains.

As we have a limited human data set, we opted to use the TL-trained models for evaluating both WSI_Fiocruz 2's test set and the entire (train and test) sets. Our goal with this analysis is to test our hypothesis in a more significant group, verifying whether the earlier results were biased by the test set random choice. Table 4 summarizes the test versus entire set comparison where, as expected, the networks obtained the best results on HE-stained samples. It is interesting to note that the results on A_{entire} were greater or very close to the A_{test} results for most cases. The one exception was the combination of MA-Net assessing HE-stained samples, where the μDice decreased but also remained competitive (dropped from 0.804 on A_{test} to 0.707 on A_{entire}). These results corroborate our hypothesis validation, strengthening the assumption that human-mice compatibility occurs in HE-stained samples. In addition, the A_{test} and A_{entire} proximity shows that the experiments are not biased on the aleatoric train/test set creation. It is important to note that the columns TL in Table 3 and A_{test} in Table 4 are the same.

The quantitative results should be read under the perspective of the results of Table 4 being compared against the results summarized in Table 2. In this regard and taking as an assumption that each Dice score of a segmenter expresses its "opinion", if one calculates the mean of these "opinions", it is possible to generate baselines to be cross compared. From Table 2, we have the following means plus standard deviations of the Dice values on WSI_Fiocruz 2's test set (across all networks for each stain), respectively for HE, PAS, and PAMS stains: 0.541 (±0.457), 0.473 (±0.381), and 0.571 (±0,207).

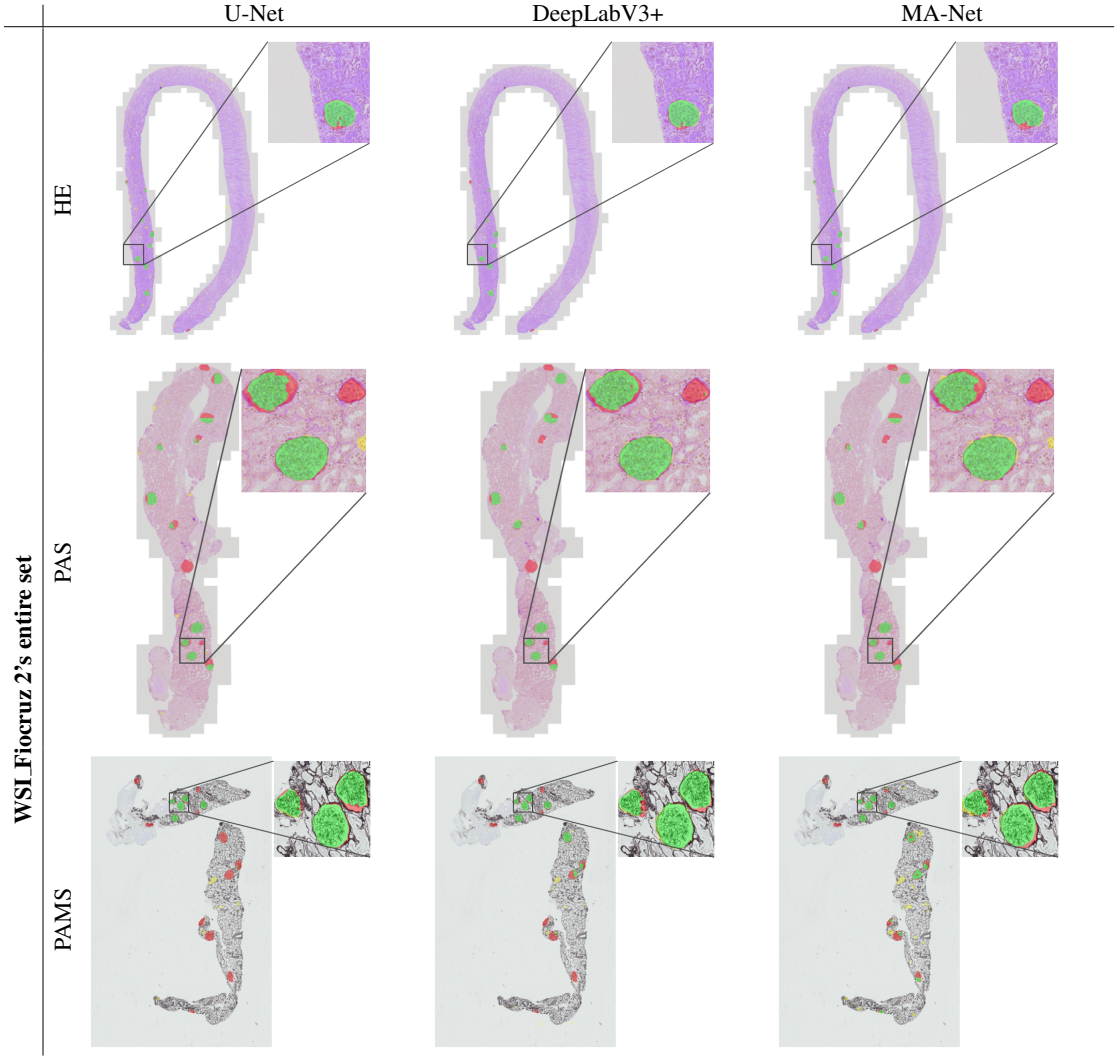

Figure 2. Samples of expected results from U-Net, DeepLabV3+, MA-Net, when trained on Lutnick's train set, predicting on WSI_Fiocruz 2's entire data set. Glomeruli in yellow are false positives (very few examples), in red are false negatives, and in green are true positive.

Contrasting these values with the means of the "opinions" from Table 4 in each stain, we have: 0.787 ($\pm$0.016), 0.446 ($\pm$0.139), and 0.633 ($\pm$0.075), respectively. Disregarding the difference in the standard deviations between the two sets of means and the very low Dice score of MA-Net in the HE-stained images of Table 2, one might firmly state that the segmenters follow the same trend in both sets of results. It is worth noting that the differences presented in the standard deviations comes from the problem found with MA-Net when trained on WSI_Fiocruz 2's train set (containing a small amount of data). As a matter of fact, all results keep almost the same or were improved in Table 4 when considering a larger set of human data, which finally gives a hint that we achieved favorable results even for different stains.

3.1. Qualitative analysis

Figure 2 illustrates a visual comparison of the segmentation results over some samples of the WSI_Fiocruz 2's data set when trained on Lutnick's train set. All the segmenters show

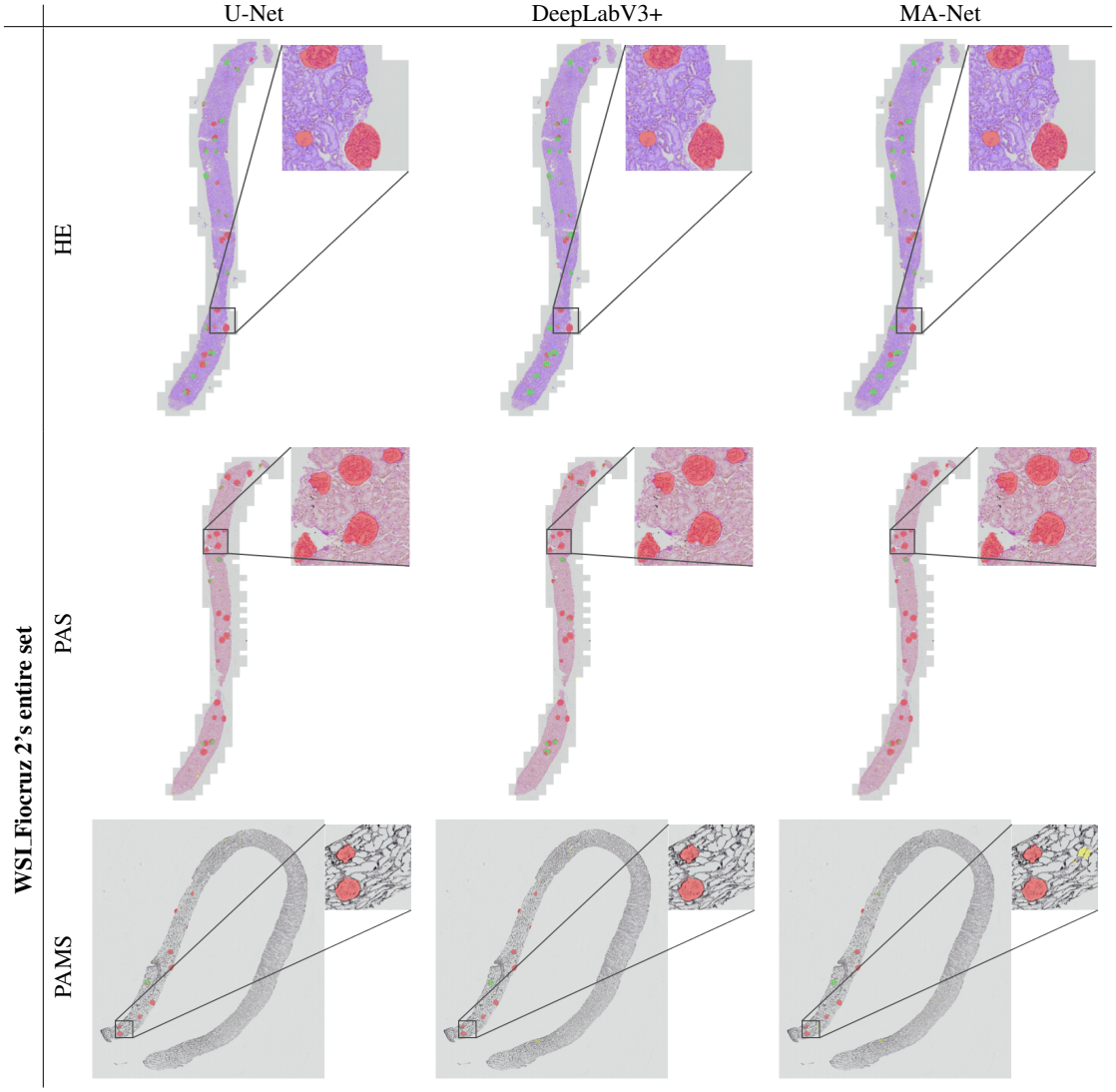

Figure 3. Samples of missing detection from U-Net, DeepLabV3+, MA-Net when trained on Lutnick's train set, predicting on WSI_Fiocruz 2's entire data set. Note that there is only one small false-positive region on PAMS stain predicted with MA-Net.

similar visual results on the samples shown in the figure. As expected, on HE-stained images, all segmenters present good visual results with only true-positive segmentation (in green), as they were also trained over HE-stained images. Also on PAS, one can notice a reasonable generalization of all segmenters. However, MA-Net still presents a false positive (in yellow) on this stain while all segmenters fail to detect one glomerulus (in red). DeepLabV3+ slightly showed more pixels in the true-positive zone on this stain. On PAMS, the three segmenters perform almost the same, but again with DeepLabV3+ presenting marginally better results.

In Figure 3, some of the visual worst results are depicted. Most results are missing detection (false negatives), but MA-Net produced a false positive on the PAMS-stained image. In fact, very few false positives can be found in the segmentation of the U-Net and DeepLabV3+ over the entire WSI_Fiocruz 2's data set.

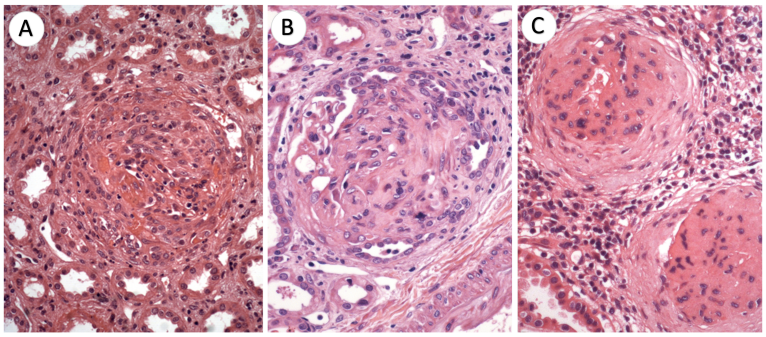

Figure 4. Glomeruli substantially changed with bruised boundaries by a disease with: Ⓐ - a cellular crescent filling the Bowman space, Ⓑ - a fibrous crescent and sclerosis of glomerular tufts, and Ⓒ - periglomerular fibrosis.

Within the qualitative analysis, one can also highlight the kind of glomerulus used to assess the performance of the segmenters. Both data sets are indeed comprised of WSIs containing normal glomerulus or those with minimal change disease (MCD), which is a type of lesion that it is not possible to verify a change in the morphological aspect of the glomerulus. Therefore, the anatomy of each glomerulus is preserved, facilitating the segmentation work. The segmentation results would surely be very different in the case of a data set formed, for instance, by glomerulus depicted in Fig. 4. All glomeruli in the figure present bruised boundaries due to three different types of lesion: Ⓐ - a cellular crescent filling the Bowman space, Ⓑ - a fibrous crescent and sclerosis of glomerular tufts, and Ⓒ - periglomerular fibrosis. Common to all samples, there is the lack of the Bowman's capsule, which delimits the normal or MCD glomeruli in the data sets we used for performance evaluation. In summary, it is reasonable to state it is not expected that a very well-tunned, off-the-shelf segmentation method would present satisfactory results over the challenging glomeruli illustrated in Fig. 4. It is a limitation of our work and an interesting room for investigation in the future.

4. Concluding remarks

The segmentation of human glomerulus carried out from training on mouse data is a challenging task, which can bring significant benefits in the field. Although we have shown in our work that it is feasible to use mouse data to train models to segment human data, achieving promising results with considerably few data, experiments were performed over data sets with well-shaped glomeruli. It is not difficult to infer that if we evaluated the selected networks on a data set containing challenging shaped glomeruli of the type found in Fig. 4, certainly we would not reach satisfactory results. This limitation opens up an exciting guess to future proposals of new methods to tackle the expected generalization that allows the practical application of these systems.

Acknowledgments

PathoSpotter is partially sponsored by Fundação de Apoio à Pesquisa do Estado da Bahia (FAPESB), grant No. TO P0008/15 and TO-SUS0031/2018 and Inova Fiocruz - Innovative ideas. Luiz Souza and Paulo Chagas have scholarships from FAPESB, grants TO-BOL0660/2018 and TO-BOL0344/2018, respectively. Angelo Duarte has the support from Universidade Estadual de Feira de Santana (UEFS) by Financiamento interno

da UEFS (FINAPESQ), grant No. TO 074/2021. Washington LC dos-Santos and Luciano Oliveira have research scholarships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grants 306779/2017 and 308580/2021-4, respectively.

References

Laura Barisoni, Cynthia C Nast, J Charles Jennette, Jeffrey B Hodgin, Andrew M Herzenberg, Kevin V Lemley, Catherine M Conway, Jeffrey B Kopp, Matthias Kretzler, Christa Lienczewski, et al. Digital pathology evaluation in the multicenter nephrotic syndrome study network (neptune). *Clinical Journal of the American Society of Nephrology*, 8(8):1449–1459, 2013.

Maxim Berman, Amal Rannen Triki, and Matthew B Blaschko. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4413–4421, 2018.

Nassim Bouteldja, Barbara M Klinkhammer, Roman D Bülow, Patrick Droste, Simon W Otten, Saskia Freifrau von Stillfried, Julia Moellmann, Susan M Sheehan, Ron Korstanje, Sylvia Menzel, et al. Deep learning–based segmentation and quantification in experimental kidney histopathology. *Journal of the American Society of Nephrology*, 32(1):52–68, 2021.

Sara Chater, Nathan Lauzeral, Anass Nouri, Youssef El Merabet, and Florent Autrusseau. Learning from mouse ct-scan brain images to detect mra-tof human vasculatures. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2830–2834. IEEE, 2021.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

Thomas de Bel, Meyke Hermsen, Bart Smeets, Luuk Hilbrands, Jeroen van der Laak, and Geert Litjens. Automatic segmentation of histopathological slides of renal tissue using deep learning. In *Medical Imaging 2018: Digital Pathology*, volume 10581, page 1058112. International Society for Optics and Photonics, 2018.

Tongle Fan, Guanglei Wang, Yan Li, and Hongrui Wang. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020.

Alton Brad Farris, Cynthia Cohen, Thomas E Rogers, and Geoffrey H Smith. Whole slide imaging for analytical anatomic pathology and telepathology: practical applications today, promises, and perils. *Archives of pathology & laboratory medicine*, 141(4): 542–550, 2017.

Michael Gadermayr, Ann-Kathrin Dombrowski, Barbara Mara Klinkhammer, Peter Boor, and Dorit Merhof. Cnn cascades for segmenting sparse objects in gigapixel whole slide images. *Computerized Medical Imaging and Graphics*, 71:40–48, 2019.

Brandon Ginley, Brendon Lutnick, Kuang-Yu Jen, Agnes B Fogo, Sanjay Jain, Avi Rosenberg, Vighnesh Walavalkar, Gregory Wilding, John E Tomaszewski, Rabi Yacoub, et al. Computational segmentation and classification of diabetic glomerulosclerosis. *Journal of the American Society of Nephrology*, 30(10):1953–1967, 2019.

Brandon Ginley, Kuang-Yu Jen, Avi Rosenberg, Felicia Yen, Sanjay Jain, Agnes Fogo, and Pinaki Sarder. Neural network segmentation of interstitial fibrosis, tubular atrophy, and glomerulosclerosis in renal biopsies. *arXiv preprint arXiv:2002.12868*, 2020.

Md Murad Hossain, Niloufar Saharkhiz, and Elisa E Konofagou. Feasibility of harmonic motion imaging using a single transducer: In vivo imaging of breast cancer in a mouse model and human subjects. *IEEE Transactions on Medical Imaging*, 40(5):1390–1404, 2021.

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*, 191(8):1431–1441, 2021.

Hye-Ryoung Kim, Sung-Won Park, Hee-Jung Cho, Kyung-Ae Chae, Ji-Min Sung, Jin-Suk Kim, Christopher P Landowski, Duxin Sun, AM Abd El-Aty, Gordon L Amidon, et al. Comparative gene expression profiles of intestinal transporters in mice, rats and humans. *Pharmacological research*, 56(3):224–236, 2007.

Brendon Lutnick, Brandon Ginley, Darshana Govind, Sean D McGarry, Peter S LaViolette, Rabi Yacoub, Sanjay Jain, John E Tomaszewski, Kuang-Yu Jen, and Pinaki Sarder. An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nature machine intelligence*, 1(2):112–119, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3): 211–252, 2015.

Justino Duarte Santos, Rodrigo de Melo Sousa Veras, Romuere Rodrigues Veloso, Nayze Lucena Sangreman Aldeman, Kelson Romulo Teixeira Aires, Andrea Gomes Campos Bianchi, et al. Classificação de imagens de biópsias renais com glomeruloesclerose segmentar e focal ou com lesões mınimas utilizando transfer learning em cnn. In *Anais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 82–93. SBC, 2019.

Olivier Simon, Rabi Yacoub, Sanjay Jain, John E Tomaszewski, and Pinaki Sarder. Multi-radial lbp features as a tool for rapid glomerular detection and assessment in whole slide histopathology images. *Scientific reports*, 8(1):1–11, 2018.

Abigail L Smith and Dorcas J Corrow. Modifications to husbandry and housing conditions of laboratory rodents for improved well-being. *ILAR journal*, 46(2):140–147, 2005.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

Ma Yi-de, Liu Qing, and Qian Zhi-Bai. Automated image segmentation using improved pcnn model based on cross-entropy. In *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.*, pages 743–746. IEEE, 2004.