Conditional image synthesis for improved segmentation of glomeruli in renal histopathological images

Florian Allender, Rémi Allègre, Cédric Wemmert, Jean-Michel Dischler

Conditional image synthesis for improved segmentation of glomeruli in renal histopathological images

1st Florian Allender
ICube, CNRS
Université de Strasbourg
Strasbourg, France
fallender@unistra.fr

2nd Rémi Allègre
ICube, CNRS
Université de Strasbourg
Strasbourg, France
remi.allegre@unistra.fr

3rd Cédric Wemmert
ICube, CNRS
Université de Strasbourg
Strasbourg, France
wemmert@unistra.fr

4th Jean-Michel Dischler
ICube, CNRS
Université de Strasbourg
Strasbourg, France
dischler@unistra.fr

***Abstract*—In a context of limited data availability, we consider the supervised segmentation of glomerular structures in patches of renal histopathological whole slide images. These structures are complex, include multiple substructures, and exhibit great variability in their shape, making their robust segmentation challenging. In this context, using appropriate data augmentation techniques is crucial to obtain more robust results. We investigate data augmentation based on random spatial deformations and conditional image synthesis for the training of a U-Net model. We rely on a SPADE model to perform the synthesis, using label maps built from the real patches available for training as input. Synthesis from ground truth masks only results in noisy patches, where substructures are absent, whereas additional structure information yield more realistic patches. We show that the best improvements of the segmentation performances are obtained by mixing real patches with synthetic patches generated from ground truth masks only, which yields an increase of up to 0.76 of average dice score w.r.t. augmentation based on spatial deformations only. We conclude that, using conditional image synthesis, patches synthesized with no additional structure information better contribute to the robustness of glomeruli segmentation than patches synthesized with structure information extracted from available real patches.**

***Index Terms*—Digital Histopathology, Glomeruli segmentation, Data augmentation, Conditional image synthesis**

I. Introduction

In the field of digital histopathology, Deep Learning aims at bringing support to pathologists for faster and more accurate diagnoses [1], [2]. Deep models based on Convolutional Neural Networks (CNN), like e.g. U-Net for segmentation [3], have the ability to learn patterns in images, which necessitates large amounts of annotated data to prevent overfitting. This requirement is a challenge, because manually segmenting Whole Slide Images (WSIs) requires expert knowledge and is extremely time-consuming and costly.

To circumvent the limitations related to small or imbalanced datasets, a standard approach in Deep Learning is to perform *Data augmentation* [4]. It consists in expanding the variety of training data in an artificial manner, which may involve *image manipulations* (spatial and colorimetric transformations, noise injection, image mixing) or *synthetic image generation*. The goal of synthetic image generation is to increase the diversity of training data by producing *novel* realistic-looking images. State-of-the-art approaches rely on Generative Adversarial Networks (GANs) [5], highly popular for their ability to learn data distributions. While the image manipulation approach raises the questions of the appropriate set of transformations and parameter tuning (see e.g. [6]), synthetic image generation raises the issue of the whole synthesis pipeline, and is itself concerned by the size and balance of the available data [7]–[9]. GAN-based image synthesis requires a large amount of data for training [7]. To cope with limited data, augmentations based on image manipulations [8] or regularization techniques [9] have been proposed.

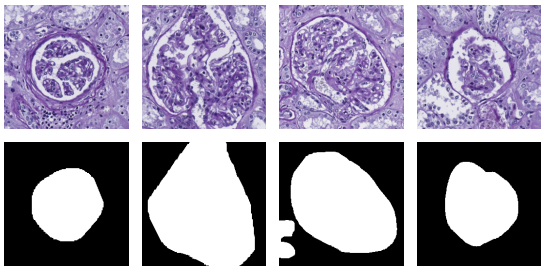

Fig. 1. Top row: Examples of glomeruli in patches of WSIs, with variable shapes (Periodic acid-Schiff staining); Bottom row: glomeruli ground truth binary masks.

In the field of digital histopathology, GANs have been successfully used for stain transfer or stain-to-stain translation [10], and also for the generation of tissue patches [11]. Previous works have addressed the synthesis of tissue patches to improve the training of classifiers in the context of cancer diagnosis [12]–[14], or improve cell nuclei segmentation [15], with demonstrated benefits.

Renal histopathology is crucial in the study of kidney diseases, and especially transplant rejection, that occurs with an incidence of 7.9% in the first year [16]. In a context of limited data, we consider the supervised segmentation of glomerular structures in patches of renal WSIs. These complex ball-shaped structures are responsible for blood filtration and con-

tain multiple substructures: membranes, capillaries, mesangial and endothelial cells, podocytes. The appearance of glomeruli thus exhibits great variability (Fig. 1), making their robust segmentation challenging. The state-of-the-art approaches rely on deep segmentation models such as U-Net [3], [17]–[20] or SegNet [21], [22], possibly with a pre-trained backbone [23], [24].

For improved generalizability, standard augmentation techniques include geometric transformations as well as transformations related to staining variations [10], [19], [25]. Murali et al. [26] explored the usage of Deep Convolutional GANs [27] to generate glomeruli images, accompanied by a visual evaluation by experts. However, no evaluation in the context of glomeruli segmentation has been performed. The complexity of the glomerular structures makes their synthesis a difficult task and classic methods either fail to reproduce fine details or suffer from mode collapse.

In this paper, we study the mixing of random spatial deformations and GAN-based conditional image synthesis as a data augmentation technique in the context of glomeruli segmentation in patches of renal WSI using a standard U-Net. We use random spatial deformations to inject *geometric* diversity in the training dataset, and conditional image synthesis to introduce *texture* variations. We describe our method and validation protocols in Section II and discuss the results in Section III.

II. Baseline and proposed method

The segmentation task consists in separating glomerular structures from surrounding tissue at pixel level in input patches. Our baseline for training the U-Net model consists of real glomeruli patches with ground truth binary segmentation masks (see Fig. 1 for examples), such that each patch contains either exactly one complete glomerulus, and possibly some parts of other glomeruli. The training, validation and test datasets are respectively composed of 660, 400 and 431 patches of size 256×256 The validation and test datasets will remain the same for all experiments. The training dataset is augmented live during training with rotations and flips with a probability of 0.5. We propose to augment the training dataset offline with random spatial deformations and synthetic patches, and evaluate the impact on segmentation.

A. Spatial deformations

For random spatial deformations, we use grid-based random deformations as suggested in the original U-Net paper [3], i.e. a 3×3 regular grid of control points, random displacements following a Gaussian distribution at control points, and cubic spline interpolation of displacements between control points. For our patches, we set the standard deviation of the displacements to 20, which amounts to a local magnitude of up to 60 pixels. We use symmetric padding along the edges of the input images for deformed locations that fall outside their bounds.

B. Patch synthesis

Architecture: We consider an image-to-image translation approach – label map to image in our case, which is a type of conditional synthesis, and employ the state-of-the-art GAN-based SPADE model [28]. The SPADE design draws its name from the "Spatially-Adaptive Normalization". While the commonly used normalization layers tend to wash away useful information from the inputs label maps [29], SPADE redesigns the generator in such a way that the normalization of the input is learned through convolutional layers (Fig. 2 top-left). The label maps are injected into SPADE "residual blocks" (Fig. 2 top-right) inserted between upsampling layers, which allows to skip the contracting path used in previous image-to-image translation architectures. The generator can then take a random vector as input, which enables multi-modal synthesis and reduces mode collapse (Fig. 2 bottom row).

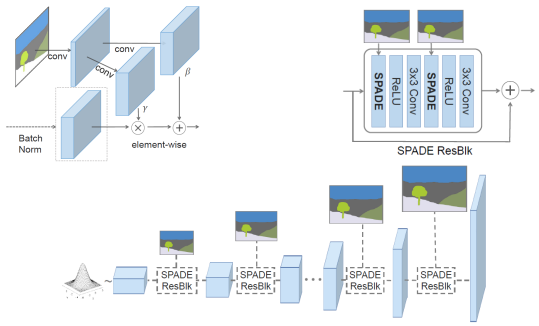

Fig. 2. Architecture of SPADE [28]. Top-left: The Spatially-Adaptive Normalization layer; Top-right: The SPADE residual block; Bottom row: The SPADE generator (images taken from [28]).

Training and synthesis: We train the SPADE network with adaptive discriminator augmentation, as introduced in StyleGAN2-ADA [8]. The probability to use augmentations evolves w.r.t. to the performance of the discriminator, preventing it to overfit and thus allowing for better results in the case of small datasets. Our set of augmentations includes rotations, flips and spatial deformations applied in the way described in Section II-A. These transformations have the most significant impact according to the authors [8]. The model is trained on 660 glomeruli patches, different from the training set of the segmentation model.

We first train the model using ground truth binary segmentation masks as label maps. To generate new glomeruli patches, we apply random spatial deformations (using the technique described in Section II-A) to the segmentation training set masks and use the results as inputs to SPADE. The generated images lack visual fidelity, with absent substructures, as highlighted by Fig. 3. Lumina of vessels are lost, as well as relevant placement of cell nuclei. This led us to consider the construction of enriched label maps.

Label maps enrichment: We propose to enrich glomeruli ground truth binary masks with additional *structure* information extracted from glomeruli images to improve the guidance of the SPADE generator. No tissue feature nor internal components of glomeruli reflecting substructures can reliably be segmented, since no corresponding ground truth is available. We therefore chose to separate the non-textured areas (i.e. lumina of vessels, urinary space, slide outer space) from textured areas

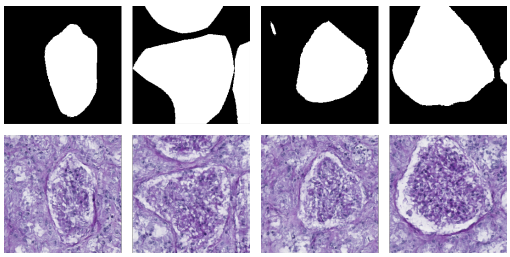

Fig. 3. Results obtained with the SPADE model trained with ground truth binary masks only. Top row: Input glomeruli binary masks; Bottom row: Synthesis results. Many details are lost and the images are noisy.

(glomeruli and surrounding tissue) in order to get structure information reliably. To do so, we assume that non-textured areas are not affected by stainings, nor important artifacts, and correspond to the color of the microscope light, which is close to white [30]. We convert glomeruli patches in HSV color space and then perform thresholding on the S component to obtain binary masks for textured areas. Fig. 4 shows examples of binary texture masks for some real glomeruli patches. For our datasets, we set the threshold to the absolute value of 0.1. Next, we combine these texture masks with corresponding glomeruli masks to produce 4-class label maps that we call "structure label maps".

Fig. 4. Binary masks separating textured and non-textured areas for the glomeruli patches presented in Fig. 1.

Training and synthesis: We performed a new training of the SPADE model using structured label maps obtained as described above. We generate synthetic glomeruli images by passing structure label maps extracted from the baseline training dataset and spatially deformed with the method described in Section II-A, as inputs to the generator. Fig. 5 shows some artificial glomeruli patches synthesized with our method. The enriched label maps enforce spatial constraints resulting in generated glomeruli visually closer to real ones. However, some fine texture details, like cell nuclei, do not appear as sharp as in real patches.

C. Validation protocol

Our objective is to evaluate the impact of training a U-Net segmentation network [3] by mixing augmentations based on random spatial deformations and synthetic images obtained via SPADE. As described earlier, we apply spatial deformations to label maps provided as inputs to the SPADE generator, and we evaluate our synthetic data against deformed real images. We therefore created 5 datasets offline:

- "Real": our baseline dataset, with 10 random deformations applied to each images (6600 images).

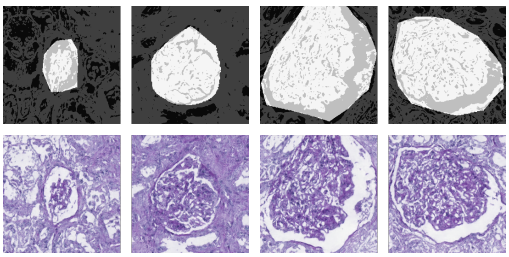

Fig. 5. Results obtained with the SPADE model trained with 4-class "structure" label maps. Top row: Input structure label maps; Bottom row: Synthesis results. The synthesized images are visually closer to real images, although the finest texture details are not reproduced.

- "SPADE-GTMask": SPADE-generated dataset, with 10 times randomly deformed ground truth binary masks from the baseline as input (6600 images).
- "SPADE-Structure": SPADE-generated dataset, with 10 times randomly deformed structure label maps from the baseline glomeruli patches as input (6600 images).
- "Real + SPADE-GTMask": baseline dataset, with 5 random deformations applied to each images (3300 images) + SPADE-generated dataset, with 5 times deformed ground truth binary masks from the baseline as input (3300 images).
- "Real + SPADE-Structure": baseline dataset, with 5 random deformations applied to each images (3300 images) + SPADE-generated dataset, with 5 times deformed structure label maps from the baseline glomeruli patches as input (3300 images).

We train a U-Net network on each dataset 5 times and average the results. A schematic view of our method workflow can be found in Fig. 6.

III. Results and Discussion

The glomeruli segmentation performance results of the U-Net model are summarized in Table I for the training datasets presented in Section II-C. The best average dice score is obtained with the dataset that mixes real and synthetic images generated with SPADE and the ground truth masks as label maps. This score is 0.76 above the score obtained with the Real dataset, and 0.85 above the score obtained with SPADE and structure label maps. Our interpretation is that the patches of SPADE-Structure remain too close from the real patches of the training dataset, despite the texture variations. The noisy aspect of the patches of SPADE-GTMask better contributes to make the U-Net model more robust. We note that when training on synthetic images alone, the SPADE-Structure dataset obtains a score 2.47 above SPADE-GTMask, which supports our claim that the SPADE-Structure patches are more realistic than the SPADE-GTMask patches. It is also worth mentioning that the lower standard deviations on the dice scores for the datasets mixing real and synthetic patches support improved segmentation robustness.

These results show that the visual fidelity of the synthetic images is not necessarily crucial when a segmentation model

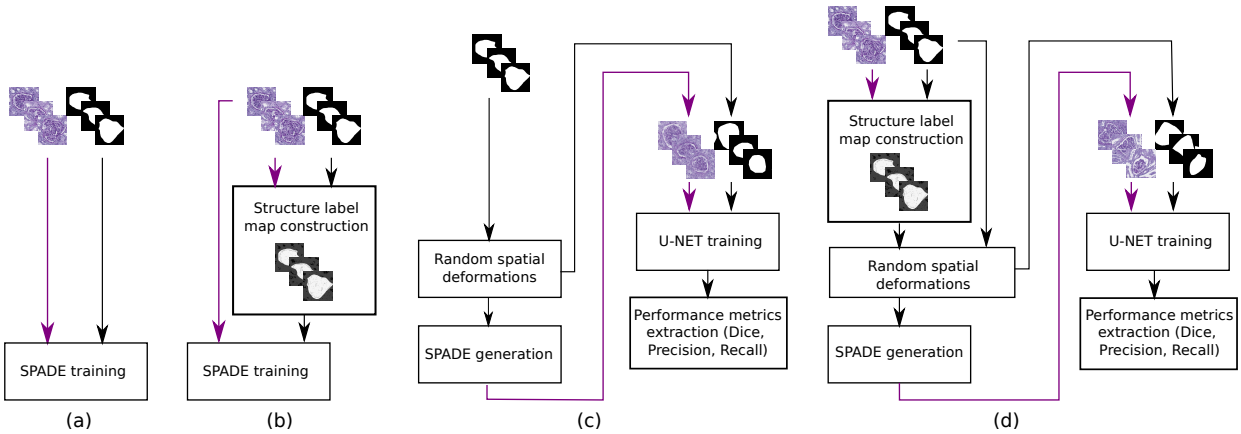

Fig. 6. Workflow process. Step 1: A SPADE model is trained with pairs of glomeruli patches and label maps. Label maps are either ground-truth binary masks (a) or structure label maps obtained by thresholding the S component of the glomeruli patches in HSV color space (b). Step 2: The U-Net segmentation model is trained. First, random spatial deformations are applied to the real training dataset. The resulting images are fed to the SPADE models trained in Step 1, thus creating the datasets SPADE-GTMask (c) and SPADE-Structure (d), that we use to train the U-Net segmentation model.

TABLE I
AVERAGE DICE, PRECISION AND RECALL W.R.T. TRAINING DATASET FOR GLOMERULI SEGMENTATION.

Training dataset	**Dice**	**Precision**	**Recall**
Real	93.31 ± 0.46	95.64 ± 0.42	91.10 ± 0.95
SPADE-GTMask	85.86 ± 1.67	$\mathbf{96.77 \pm 0.39}$	77.20 ± 2.84
SPADE-Structure	88.33 ± 1.46	93.70 ± 2.16	82.41 ± 5.40
Real + SPADE-GTMask	$\mathbf{94.07 \pm 0.17}$	94.85 ± 0.85	$\mathbf{93.30 \pm 0.79}$
Real + SPADE-Structure	93.22 ± 0.14	94.86 ± 0.44	91.64 ± 0.45

is trained with a mix of real and synthetic data. The small obtained performance improvement also shows that the chosen generative approach has a limited ability to expand the diversity of the training dataset. The amount of beneficial randomness in texture and geometry, and the question of how to enable the emergence of *novelty*, deserve more investigations.

Future lines of research will include the exploration of alternative ways to further increase the diversity of the training dataset. This could be achieved by enabling and controlling *structure* variations w.r.t. the baseline training set, rather than relying on already seen structures. This would require to devise a generative model with appropriate design to synthesize new structure label maps to feed the conditional patch synthesis model. Alternatives to the U-Net model, such as SegNet, or pre-trained versions of such models [23], [24], will be considered. Another line of research would be to explore unconditional patch synthesis, using e.g. StyleGAN2-ADA [8], that we did not consider in this study, and perform comparisons with conditional synthesis.

IV. ACKNOWLEDGEMENTS

The glomeruli patches used in our experiments are courtesy of the ERACoSysMed project "SysMIFTA", co-funded by EU H2020 and the national funding agencies German Ministry of Education and Research (BMBF) project management PTJ (FKZ: 031L-0085A), and Agence National de la Recherche (ANR), project number ANR-15-CMED-0004.

The authors acknowledge the High Performance Computing Center ROMEO hosted by the University of Reims Champagne-Ardenne for access to computing resources.

REFERENCES

[1] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-van de Kaa, P. Bult, B. van Ginneken, and J. van der Laak, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Scientific reports*, vol. 6, pp. 26286, 2016.

[2] C. Wemmert, J. Weber, F. Feuerhake, and G. Forestier, *Deep Learning for Biomedical Data Analysis: Techniques, Approaches, and Applications*, chapter Deep Learning for Histopathological Image Analysis, pp. 153–169, 2021.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.

[4] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 60, 2019.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2014, vol. 27.

[6] K. Faryna, J. van der Laak, and G. Litjens, "Tailoring automated data augmentation to h&e-stained histopathology," in *Proc. Conference on Medical Imaging with Deep Learning*, 2021, vol. 143 of *Proc. Machine Learning Research*, pp. 168–178.

[7] P. Shamsolmoali, M. Zareapoor, E. Granger, H. Zhou, R. Wang, M. E. Celebi, and J. Yang, "Image synthesis with adversarial networks: A comprehensive survey and case studies," *Information Fusion*, vol. 72, pp. 126–146, 2021.

[8] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," in *Proc. Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[9] H. Tseng, L. Jiang, C. Liu, M. Yang, and W. Yang, "Regularizing generative adversarial networks under limited data," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 7921–7931.

[10] J. Vasiljević, F. Feuerhake, C. Wemmert, and T. Lampert, "Towards histopathological stain invariance by unsupervised domain augmentation using generative adversarial networks," *Neurocomputing*, vol. 460, pp. 277–291, 2021.

[11] M. E. Tschuchnig, G. J. Oostingh, and M. Gadermayr, "Generative adversarial networks in digital pathology: A survey on trends and future potential," *Patterns*, vol. 1, no. 6, pp. 100089, 2020.

[12] J. Wei, A. Suriawinata, L. Vaickus, B. Ren, X. Liu, J. Wei, and S. Hassanpour, "Generative Image Translation for Data Augmentation in Colorectal Histopathology Images," in *Proc. Machine Learning for Health NeurIPS Workshop*, 2020, vol. 116 of *Proc. Machine Learning Research*, pp. 10–24.

[13] A. B. Levine, J. Peng, D. Farnell, M. Nursey, Y. Wang, J. R. Naso, H. Ren, H. Farahani, C. Chen, D. Chiu, A. Talhouk, B. Sheffield, M. Riazy, P. P. Ip, C. Parra-Herran, A. Mills, N. Singh, B. Tessier-Cloutier, T. Salisbury, J. Lee, T. Salcudean, S. J. Jones, D. G. Huntsman, C. B. Gilks, S. Yip, and A. Bashashati, "Synthesis of diagnostic quality cancer pathology images by generative adversarial networks," *The Journal of Pathology*, vol. 252, no. 2, pp. 178–188, 2020.

[14] A. C. Quiros, R. Murray-Smith, and K. Yuan, "Pathologygan: Learning deep representations of cancer tissue," in *Proc. Conference on Medical Imaging with Deep Learning*, 2020, vol. 121 of *Proc. Machine Learning Research*, pp. 669–695.

[15] F. Mahmood, D. Borders, R. Chen, G. McKay, K. J. Salimian, A. Baras, and N. Durr, "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE Transactions on Medical Imaging*, 2019.

[16] R. H. Naik and S. H. Shawar, "Renal transplantation rejection," *StatPearls [Internet]*, 2021, PMID: 31971715.

[17] T. de Bel, M. Hermsen, B. Smeets, L. Hilbrands, J. van der Laak, and G. Litjens, "Automatic segmentation of histopathological slides of renal tissue using deep learning," in *Proc. Medical Imaging 2018: Digital Pathology*. International Society for Optics and Photonics, 2018, vol. 10581, pp. 285 – 290.

[18] T. Lampert, O. Merveille, J. Schmitz, G. Forestier, F. Feuerhake, and C. Wemmert, "Strategies for training stain invariant cnns," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 905–909.

[19] O. Merveille, T. Lampert, J. Schmitz, G. Forestier, F. Feuerhake, and C. Wemmert, "An automatic framework for fusing information from differently stained consecutive digital whole slide images: A case study in renal histology," *Computer Methods and Programs in Biomedicine*, vol. 208, pp. 106157, 2021.

[20] R. C. Davis, X. Li, Y. Xu, Z. Wang, N. Souma, G. Sotolongo, J. Bell, M. Ellis, D. Howell, X. Shen, K. Lafata, and L. Barisoni, "Deep learning segmentation of glomeruli on kidney donor frozen sections," *medRxiv*, 2021.

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[22] N. Altini, G. D. Cascarano, A. Brunetti, F. Marino, M. T. Rocchetti, S. Matino, U. Venere, M. Rossini, F. Pesce, L. Gesualdo, and V. Bevilacqua, "Semantic segmentation framework for glomeruli detection and classification in kidney histological sections," *Electronics*, vol. 9, no. 3, 2020.

[23] G. Bueno, M. M. Fernandez-Carrobles, L. Gonzalez-Lopez, and O. Deniz, "Glomerulosclerosis identification in whole slide images using semantic segmentation," *Computer Methods and Programs in Biomedicine*, vol. 184, pp. 105273, 2020.

[24] M. Salvi, A. Mogetta, A. Gambella, L. Molinaro, A. Barreca, M. Papotti, and F. Molinari, "Automated assessment of glomerulosclerosis and tubular atrophy using deep learning," *Computerized Medical Imaging and Graphics*, vol. 90, pp. 101930, 2021.

[25] X. Li, R. C. Davis, Y. Xu, Z. Wang, N. Souma, G. Sotolongo, J. Bell, M. Ellis, D. Howell, X. Shen, K. J. Lafata, and L. Barisoni, "Deep learning segmentation of glomeruli on kidney donor frozen sections," *Journal of Medical Imaging*, vol. 8, no. 6, pp. 1 – 18, 2021.

[26] L. K. Murali, B. Lutnick, B. Ginley, J. E. Tomaszewski, and P. Sarder, "Generative modeling for renal microanatomy," *Proc. SPIE–the International Society for Optical Engineering*, vol. 11320, pp. 113200F, 2020.

[27] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. International Conference on Learning Representations*, 2016.

[28] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2332–2341.

[29] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[30] A. Korzynska, L. Roszkowiak, D. Pijanowska, W. Kozlowski, and T. Markiewicz, "The influence of the microscope lamp filament colour temperature on the process of digital images of histological slides acquisition standardization," *Diagnostic Pathology*, vol. 9, no. 1, pp. S13, Dec 2014.