

Automated Annotations enable Large-Scale Morphological Analysis in Kidney Histopathology Using Deep Learning

Katharina Kösseler^{a,*}

^aInstitute of Pathology, University Clinic Aachen, RWTH Aachen University

Abstract. Deep learning in digital pathology relies on high-quality annotations, which are expensive to produce. We propose an automated pipeline that uses multiplex immunofluorescence (mIF) to generate ground truth annotations for peritubular capillaries (PTCs) in periodic acid–Schiff (PAS) stained kidney biopsies. Image registration between mIF and PAS, combined with thresholding endothelial cell markers and a deep learning-based structural segmentation, yields robust PTC annotations, optionally refined via the Segment Anything Model (SAM). Since mIF processing alters PAS appearance, we propose the employment of a cycle-consistent GAN to align the post-mIF and normal PAS stains, ensuring consistent segmentation performance across stains. This scalable annotation strategy will lay the groundwork for advanced morphological analyses tied to clinical outcomes.

1 Introduction

In recent years, artificial intelligence has become increasingly important in digital pathology. Deep learning models have been applied for automating diagnosis as well as predicting patient outcome and deriving new morphological biomarkers [18].

One particularly promising application is pathomics [5, 10]. While end-to-end outcome or diagnosis prediction models are essentially black boxes and do not give a direct explanation, pathomics quantitatively assesses patient tissue. First, the WSI (whole-slide image) of the tissue scan is segmented utilizing a deep learning model. Subsequently, quantitative features are extracted from the instances in the segmentation masks such as diameter or distance to closest instance. This yields a large dataframe containing information about all segmented instances which can be statistically related to clinical outcomes.

However, training robust deep learning segmentation models remains challenging, as they require high-quality pixel-level annotations. Manual annotation of vast numbers of WSIs is time-consuming, prone to inter-observer variability, and requires expert knowledge.

A solution for generating ground truth annotations for routine tissue scans is utilizing registered immunofluorescence or immunohistochemistry images for automatically generating a segmentation mask [20, 13]. In our project, we employ multiplex immunofluorescence (mIF) to automatically generate ground truth annotations

for periodic acid Schiff (PAS) stained WSIs of the same kidney tissue. PAS staining was chosen because it is a common procedure in nephropathology. The use of mIF permits the incorporation of different tissue markers for the differentiation of the renal microanatomic structure and immune cells. This enables the creation of an extensive ground truth annotation dataset for a variety of classes. We initially focused on annotating peritubular capillaries (PTCs) which are small blood vessels. Primarily because our dataset originates from transplant cases, where PTCs are important for transplant rejection classification according to the Banff criteria [16, 1]. In the future, we will extend these annotations to include additional classes, such as immune cell nuclei.

2 Method

2.1 Data

Our dataset consists of 42 mIF and PAS image pairs, each showing a slice from a kidney transplant biopsy. The cases include a range of diagnoses: normal (n=10), T-cell-mediated rejection (TCMR) (n=8), antibody-mediated rejection (ABMR) (n=9), TCMR and ABMR (n=6) and polyomavirus nephropathy (n=9).

The mIF images incorporate 24 markers that enable detailed differentiation of immune cells and renal microanatomic structures. These images were acquired using the Lunaphore COMET system at a resolution of 0.28 μm per pixel. Subsequently, the same tissue was stained with PAS and scanned at a resolution of 0.2521 μm per pixel using the Aperio AT2 whole-slide scanner by Leica Biosystems.

2.2 Annotation Generation

To generate the annotations, we began by performing background subtraction on all mIF scans with the Horizon image viewer to remove autofluorescence. Furthermore, the Warpy workflow [8] was utilized to register the corresponding PAS and mIF images.

PTC annotations were generated from three mIF channels: PLVAP, CD31 and DAPI. Both PLVAP and CD31 target endothelial cells, while DAPI stains cell nuclei. Given that endothelial cells line all blood vessels as well as the lymphatic system [9], these markers are well-suited for identifying PTCs. First, Otsu thresholding [15] was applied separately to the PLVAP and CD31 channels to produce initial masks of positively stained cells, which were then combined. Holes in the mask were filled, and partially captured nuclei were fully included. For this adjustment, nuclei were initially segmented from

* Corresponding Author. Email: kkoesseler@ukaachen.de.

the DAPI channel using the StarDist extension in QuPath [17, 3]. After these corrections, any remaining holes in the annotations were closed.

To refine the PTC annotations further, instances that were clearly not PTCs were filtered out. A hand-annotated tissue mask was used to exclude areas outside of the tissue. Additionally, a structural segmentation mask generated using the FLASH deep learning model from [10] was employed to filter the annotations. The mask segments different structures in the kidney: glomeruli, tubuli, arteries, and veins. This step is necessary, as arteries are also lined by endothelial cells as well as the capillaries inside of glomeruli. Finally, all instances that are smaller than a threshold of 510 pixels were removed. The threshold was determined by a medical student through visual inspection.

We also tested the Segment Anything Model (SAM) [14] for annotation refinement and plan to assess the performance of MicroSAM [2], optimized for microscopy data. In future work, the annotation creation workflow will be extended to include additional classes such as nuclei from various cell types, which can be distinguished by their marker intensities with the support of the existing FLASH segmentation mask.

2.3 Segmentation Pipeline

For the training of a segmentation model, the PAS WSI will be divided into patches, as gigapixel size prevents direct processing. We will benchmark various segmentation network architectures such as nnUNet [12], UNI [6] and CellViT++ [11] for nuclei segmentation to determine the most suitable approach. To ensure reliable benchmarking, the automatically generated annotations will be primarily reviewed by a medical student, with selective validation by a supervising pathologist to minimize variability while keeping human intervention limited.

The initial segmentation model will be trained and evaluated on our relatively small dataset of 42 whole-slide images. As a next step, we plan to assess the segmentation performance on a larger cohort, with the expectation that the model's capabilities will transfer. However, we have observed that PAS WSIs stained after applying the mIF workflow appear different from those where PAS staining was directly applied using the same protocol. A similar issue was noted by Wiedenmann et al. [20] for Hematoxylin and Eosin staining. In their work, a cycle GAN was used for stain transfer from the normal stain to the post-mIF H&E stain, which improved the segmentation quality of a model trained on post-mIF H&E WSIs when applied to normal H&E WSIs. Based on this observation, we also plan to explore synthetic image generation methods to improve transferability between the two stain modalities: normal PAS and post-mIF PAS.

2.4 Morphological Features: Engineering & Analysis

It has been shown that structural segmentation masks can be used to extract meaningful and interpretable morphological features, such as the diameter or the distance to the closest instance, which in turn can be correlated with clinical outcomes [10, 7]. With the guidance of a pathologist, we will engineer such features from the segmentation outputs. The segmentation model and subsequent feature extraction pipeline will be applied to a large patient cohort. Finally, statistical analyses will be conducted to discover the relationships between these extracted features and clinical outcomes.

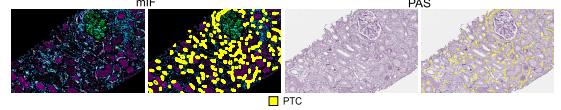


Figure 1. Example output from the automatic annotation pipeline.

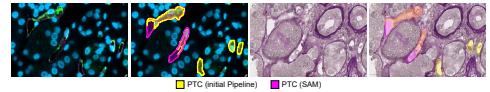


Figure 2. Example of the SAM improvement.

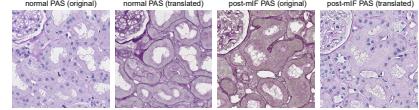


Figure 3. Stain translation for one patient's normal and post-mIF PAS patch.

3 Results

The pipeline for generating the PTC annotations has been implemented in Python and applied to 15 of the 42 image pairs. In these cases, a medical student has already corrected the FLASH structural segmentation mask. Figure 1 shows an example annotation. The annotations appear promising, though with room for improvement.

To enhance the results, we experimented with the QuPath extension for SAM [19] to refine the generated annotations. An example of these refined annotations is presented in Figure 2. Currently, we are exploring the integration of SAM or Micro-SAM directly into the Python pipeline for annotation creation.

In parallel, we trained a cycle-consistent generative adversarial network (CycleGAN) for stain translation between post-mIF and normal PAS, using the code from Bouteldja et al. [4]. The goal was to improve the segmentation quality of the FLASH model on the post-mIF PAS WSI. Figure 3 illustrates two translated patches, highlighting the potential benefits of this approach. The model was initially trained during the data acquisition phase when only a subset of the dataset was available. We will soon retrain the model using the complete dataset.

4 Outlook

As a next step, we will implement a pipeline for training and evaluating the segmentation model on PAS stained kidney biopsies using the automatically generated annotations. Over time, additional annotations will be assembled and the pipeline extended to other classes. Although this work primarily targets peritubular capillaries, the modular design also enables application to other renal structures, such as glomerular capillaries. Once the annotations have been reviewed, we will benchmark various segmentation models and select the best-performing one to apply to a comprehensive multi-site transplant cohort of more than 6000 cases, potentially incorporating a generative model to enhance transferability. Since no public paired PAS–mIF kidney datasets are available, generality of the annotation pipeline will be assessed within this large cohort. Finally, we will extract pathomics features and analyze their association with clinical parameters. While the pipeline builds on established methods, its integration into a scalable framework tailored to nephropathology represents a novel contribution with high clinical relevance.

Acknowledgements

I would like to thank my supervisor Roman David Bülow and my colleague Eva Marie Zeine for their contributions to this work.

References

- [1] Reference guide to the banff classification, banff classification for renal allograft pathology, 2022. <https://banffffoundation.org/central-repository-for-banff-2019-resources-3/>. Accessed: 2025-07-01.
- [2] A. Archit, L. Freckmann, S. Nair, N. Khalid, P. Hilt, V. Rajashekhar, M. Freitag, C. Teuber, G. Buckley, S. von Haaren, et al. Segment anything for microscopy. *Nature Methods*, pages 1–13, 2025.
- [3] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, et al. Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7, 2017.
- [4] N. Bouteleja, D. L. Hölscher, B. M. Klinkhammer, R. D. Buelow, J. Lotz, N. Weiss, C. Daniel, K. Amann, and P. Boor. Stain-independent deep learning-based analysis of digital kidney histopathology. *The American journal of pathology*, 193(1):73–83, 2023.
- [5] R. D. Bülow, D. L. Hölscher, I. G. Costa, and P. Boor. Extending the landscape of omics technologies by pathomics. *npj Systems Biology and Applications*, 9(1):38, 2023.
- [6] R. J. Chen, T. Ding, M. Y. Lu, D. F. Williamson, G. Jaume, A. H. Song, B. Chen, A. Zhang, D. Shao, M. Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024.
- [7] Y. Chen, J. Zee, A. R. Janowczyk, J. Rubin, P. Toro, K. J. Lafata, L. H. Mariani, L. B. Holzman, J. B. Hodgin, A. Madabhushi, et al. Clinical relevance of computationally derived attributes of peritubular capillaries from kidney biopsies. *Kidney360*, 4(5):648–658, 2023.
- [8] N. Chiaruttini, O. Burri, P. Haub, R. Guiet, J. Sordet-Dessimoz, and A. Seitz. An open-source whole slide image registration workflow at cellular precision using fiji, qupath and elastix. *Frontiers in Computer Science*, 3:780026, 2022.
- [9] M. Féletal. The endothelium, part i: Multiple functions of the endothelial cells—focus on endothelium-derived vasoactive mediators. 2011.
- [10] D. L. Hölscher, N. Bouteleja, M. Joodaki, M. L. Russo, Y.-C. Lan, A. V. Sadr, M. Cheng, V. Tesar, S. V. Stillfried, B. M. Klinkhammer, et al. Next-generation morphometry for pathomics-data mining in histopathology. *Nature communications*, 14(1):470, 2023.
- [11] F. Hörst, M. Rempe, H. Becker, L. Heine, J. Keyl, and J. Kleesiek. Celavit++: Energy-efficient and adaptive cell segmentation and classification using foundation models. *arXiv preprint arXiv:2501.05269*, 2025.
- [12] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
- [13] T. Kataria, S. Rajamani, A. B. Ayubi, M. Bronner, J. Jedrziewicz, B. S. Knudsen, and S. Y. Elhabian. Automating ground truth annotations for gland segmentation through immunohistochemistry. *Modern Pathology*, 36(12):100331, 2023.
- [14] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.
- [15] N. Otsu et al. A threshold selection method from gray-level histograms. *Automatica*, 11(285–296):23–27, 1975.
- [16] C. Roufosse, N. Simmonds, M. Clahsen-van Groningen, M. Haas, K. J. Henriksen, C. Horsfield, A. Loupy, M. Mengel, A. Perkowska-Ptasinska, M. Rabant, et al. A 2018 reference guide to the banff classification of renal allograft pathology. *Transplantation*, 102(11):1795–1814, 2018.
- [17] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II*, pages 265–273, 2018. doi: 10.1007/978-3-030-00934-2_30.
- [18] A. H. Song, G. Jaume, D. F. Williamson, M. Y. Lu, A. Vaidya, T. R. Miller, and F. Mahmood. Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering*, 1(12):930–949, 2023.
- [19] K. Sugawara. Training deep learning models for cell image segmentation with sparse annotations. *bioRxiv*, 2023. doi: 10.1101/2023.06.13.544786. URL <https://www.biorxiv.org/content/early/2023/06/13/2023.06.13.544786>.
- [20] M. Wiedenmann, M. Barch, P. S. Chang, J. Giltnane, T. Risom, and A. Zijlstra. An immunofluorescence-guided segmentation model in hematoxylin and eosin images is enabled by tissue artifact correction using a cycle-consistent generative adversarial network. *Modern Pathology*, 37(11):100591, 2024.