

# PrPSeg: Universal Proposition Learning for Panoramic Renal Pathology Segmentation

Ruining Deng<sup>1</sup>, Quan Liu<sup>1</sup>, Can Cui<sup>1</sup>, Tianyuan Yao<sup>1</sup>, Jialin Yue<sup>1</sup>, Juming Xiong<sup>1</sup>, Lining Yu<sup>1</sup>, Yifei Wu<sup>1</sup>, Mengmeng Yin<sup>2</sup>, Yu Wang<sup>2</sup>, Shilin Zhao<sup>2</sup>, Yucheng Tang<sup>3</sup>, Haichun Yang<sup>2</sup>, Yuankai Huo<sup>1,2\*</sup>

1. Vanderbilt University, 2. Vanderbilt University Medical Center, 3. NVIDIA Corp.

\*contact author: yuankai.huo@vanderbilt.edu

## Abstract

*Understanding the anatomy of renal pathology is crucial for advancing disease diagnostics, treatment evaluation, and clinical research. The complex kidney system comprises various components across multiple levels, including regions (cortex, medulla), functional units (glomeruli, tubules), and cells (podocytes, mesangial cells in glomerulus). Prior studies have predominantly overlooked the intricate spatial interrelations among objects from clinical knowledge. In this research, we introduce a novel universal proposition learning approach, called panoramic renal pathology segmentation (PrPSeg), designed to segment comprehensively panoramic structures within kidney by integrating extensive knowledge of kidney anatomy.*

*In this paper, we propose (1) the design of a comprehensive universal proposition matrix for renal pathology, facilitating the incorporation of classification and spatial relationships into the segmentation process; (2) a token-based dynamic head single network architecture, with the improvement of the partial label image segmentation and capability for future data enlargement; and (3) an anatomy loss function, quantifying the inter-object relationships across the kidney.*

## 1. Introduction

Digital pathology has revolutionized the field of pathology [1], not only facilitating the transition from local microscopes to remote monitoring for pathologists but also providing a significant opportunity for large-scale computer-assisted quantification in pathology [5, 19, 41]. In the clinical anatomy of renal pathology, there are different levels of quantification necessary for disease diagnosis [42], severity recognition [34], and treatment effectiveness evaluation [28], ranging from region-level objects (like medulla and cortex) to functional units (glomerulus, tubules, vessels, etc.) and even to individual cells within these units.

These tasks, especially at the functional unit and cell levels, are prone to errors and variability in human examination and require labor-intensive efforts [26, 47, 53]. Therefore, achieving comprehensive quantification from regions to cells is necessary in renal pathology, but it remains an inevitably laborious task with manual human effort.

While many studies have developed pathological image segmentation techniques for pixel-level tissue characterization, particularly using deep learning methods [4, 12, 35, 44, 51], they still encounter three major limitations: (1) Current multi-network and multi-head designs [11, 25, 27, 37, 48, 52] focus only on single tissue structures or structures at similar scales, lacking a comprehensive approach to achieve all-encompassing segmentation across different levels from regions to cells. (2) These approaches require modifications to their architectures when new classes are introduced, preventing the reuse of existing backbones without significant alterations; (3) Comprehensive semantic (multi-label) segmentation and quantification on renal histopathological images remain challenging due to the intricate spatial relationships among different tissue structures. The spatial relationships between these different objects are illustrated in Figure 1. *Understanding these interrelations in renal pathology is crucial for achieving effective all-encompassing segmentation, yet recent advancements in deep learning have not fully incorporated this comprehensive modeling into the training process, nor have they achieved panoramic segmentation for the complete anatomy of the kidney.*

To address these challenges, we introduce a token-based dynamic head network designed to achieve panoramic renal pathology segmentation (PrPSeg) by modeling the spatial relationships among all objects. This approach allows for the reuse of the same architectural framework, even when the dataset size expands. A universal proposition matrix is established to translate anatomical relationships into computational modeling concepts. An anatomy loss is also proposed to integrate spatial relationships into the model train-

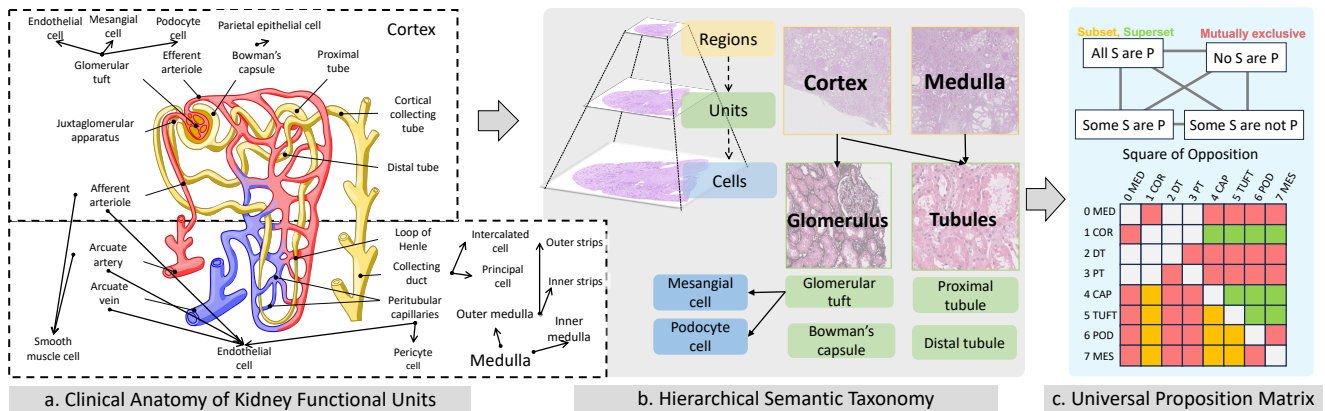


Figure 1. **Knowledge transformation from kidney anatomy to computational modeling** – This figure demonstrates the transformation of intricate clinical anatomical relationships within the kidney into a structured computational matrix. (a) Pathologists examine histopathology following the kidney anatomy. (b) This study revisits such kidney anatomy with hierarchical semantic taxonomy. (c) The proposed PrPSeg method further mathematically abstracts the semantic taxonomy as a universal proposition matrix. This matrix serves as a foundation for our computational model, reflecting the complex interplay of anatomical elements in the kidney.

ing as a form of semi-supervised learning. To our knowledge, this is the first deep learning algorithm to accomplish panoramic segmentation in renal pathology, demonstrating superior performance in all-encompassing segmentation.

The contributions of this paper are threefold:

- The design of a comprehensive universal proposition matrix for renal pathology, facilitating the incorporation of classification and spatial relationships into the segmentation process.
- The development of a token-based dynamic head in a single network architecture, improving partial label image segmentation.
- The formulation of an anatomy loss function, quantifying the inter-object relationships across the kidney.

## 2. Relative Work

### 2.1. Renal Pathology Segmentation

Recent advancements in deep learning have positioned Convolutional Neural Networks (CNNs) and Transformer-based networks as leading methods for image segmentation, particularly in renal pathology [15, 22]. Innovations in this field have ranged from CNN cascades for sparse tissue segmentation [16] to the deployment of AlexNet for pixel-wise classification and detection [17]. Notably, multi-class learning approaches using SegNet-VGG16 and DeepLab v2 have been implemented for detecting various glomerular structures and renal pathologies [7, 39]. In addition, instance segmentation and Vision Transformers (ViTs) have begun to find applications in this domain [18, 30, 43, 50].

However, most existing methods focus on segmenting single tissue types or multiple structures at similar levels, such as glomeruli and tubules [21, 32, 40]. Comprehensive approaches capable of spanning from tissue region level to

cell level remain unexplored. Moreover, some methods prioritize disease-positive region segmentation over a holistic understanding of kidney morphology [29, 38].

Recent methods utilize hierarchical information for semantic segmentation [33, 36] or classification and prediction [8]. However, these methods primarily focus on class-based relationships between objects. While all objects have a uniform resolution in natural images, this approach neglects the emphasis on pixel-wise anatomical and spatial relationships at multiple resolutions in kidney datasets.

Building upon these insights, our work introduces a token-based approach, leveraging class-specific and scale-specific tokens. This method is designed to capture heterogeneous features and employs semi-supervised learning to understand pixel-wise spatial relationships across multiple scales, achieving panoramic segmentation in renal pathology.

### 2.2. Dynamic Single Network

While multi-head single network designs have been proposed for multi-class renal pathology segmentation [6, 9, 14, 20], they often require dense multi-class annotations. Given the labor-intensive nature of such annotations, pathology data is frequently partially labeled. Additionally, forming spatial correlations, such as subset/superset relationships, remains a challenge in multi-class upsegmentation frameworks.

Recent developments in dynamic neural networks have paved the way for more comprehensive segmentation using single multi-label networks, even with partially labeled data [11, 52]. These networks dynamically generate neural network parameters, adapting to various imaging contexts. However, they predominantly rely on binary segmentation approaches and do not fully integrate spatial correlations in

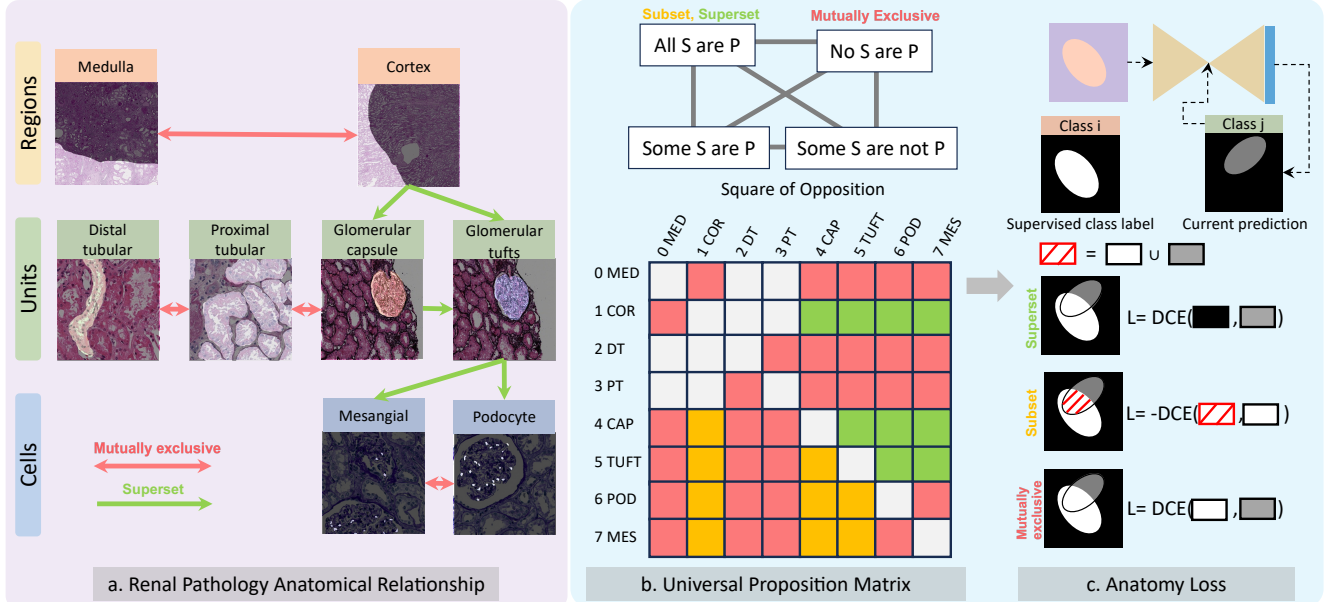


Figure 2. **Universal proposition matrix with anatomy loss** – This figure shows the key innovation of the proposed method. (a) Multi-scale (region-level, unit-level, and cell-level) hierarchical semantic taxonomy is presented. (b) The proposed PrPSeg mathematically models the semantic taxonomy as a universal proposition matrix, which delineates robust constraints and relationships between anatomical entities. (c) We further encode the universal proposition matrix as a novel anatomy loss function, designed to operationalize the affirmative and negatory relationships inherent in kidney anatomy.

their training processes.

Our work extends these concepts by translating spatial correlations from anatomy into a programming model, represented as a matrix coupled with an anatomy-based loss for semi-supervised learning. This strategy effectively harnesses the affirmative and negatory relationships in kidney anatomy, enabling detailed and comprehensive segmentation. It enhances the model’s ability to distinguish and classify the complex structures within kidney, representing a significant advancement in the field.

### 3. Method

#### 3.1. Problem Formulation

This study aims to segment an array of anatomical concepts in renal pathology, encapsulating three conceptual layers — regions ( $R_1, R_2$ ), functional units ( $F_1, F_2, F_3, F_4$ ), and cells ( $C_1, C_2$ ) — spanning 8 distinct objects.

Leveraging anatomical learning, our approach is tailored to achieve comprehensive segmentation in renal pathology by effectively interpreting both affirmative and negatory relationships within the anatomical relationship.

The pipeline is composed of three integral components: **Universal proposition matrix:** A meticulously crafted universal proposition matrix is employed to elucidate the anatomical relationships among various objects. This matrix aids in enhancing structural understanding from an engineering perspective, facilitating better cognition of the

complex renal architecture.

**Token-based dynamic head backbone:** We have developed a token-based dynamic head backbone architecture that is adept at interpreting class-aware and scale-aware knowledge pertinent to renal pathology. This component is pivotal in accommodating the extensibility requirements posed by the introduction of new data, ensuring the model’s adaptability and scalability.

**Anatomy loss function:** A novel anatomy loss function has been formulated, which operationalizes the affirmative and negatory relationships inherent in kidney anatomy. This function is a critical element in achieving nuanced, all-encompassing segmentation, bolstering the model’s ability to discern and categorize the intricate structures within the kidney.

#### 3.2. Universal proposition matrix

Renal pathology encompasses regions (the medulla and cortex), functional units (glomerulus, tubules, etc.), with corresponding cellular structures. Pathologists analyze the morphology of the kidney by examining the functional flow through these heterostructures. Traditionally, each heterostructure undergoes isolated examination and quantification to meet specific demands, often overlooking the homostructures within each unit, such as podocyte cells, mesangial cells in glomerular tufts. To enhance the understanding of the integrated kidney structure, we propose transforming kidney anatomy into an anatomical relation-

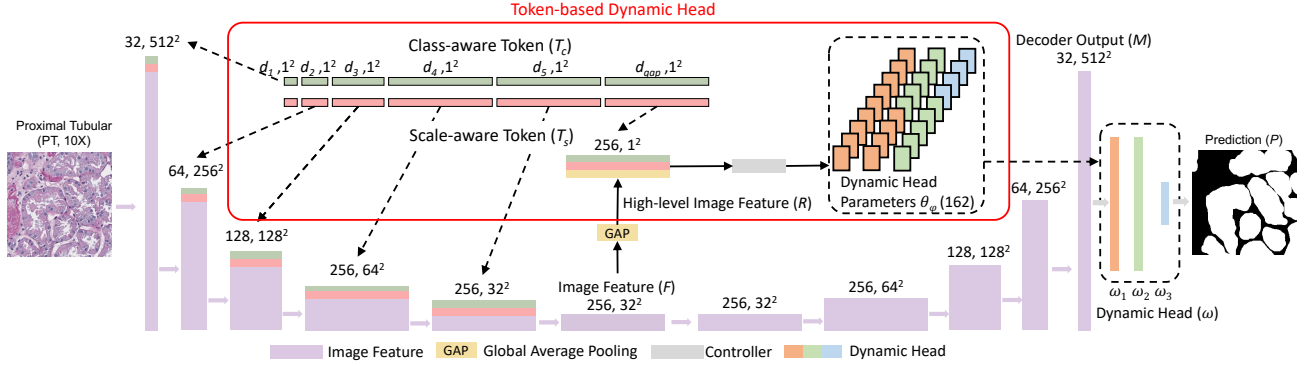


Figure 3. **Token-based dynamic head network architecture** – This figure illustrates the architecture of the proposed PrPSeg method. It incorporates a residual U-Net backbone, augmented with class-aware and scale-aware tokens. These tokens are integrated into each block of the encoder, as well as the Global Average Pooling (GAP) block, ensuring a comprehensive understanding of both class and scale features. Such features are aggregated by a fusion block to adaptively generate the parameters for a single dynamic segmentation head. The proposed method is able to segment all hierarchical semantic anatomies using a single network.

ship map, using principles of affirmation and negation from linguistic and grammatical concepts.

To translate these anatomy concepts into an engineering framework, we adopt Aristotle’s logic theory to develop an anatomical map for renal pathology. Aristotle’s theory examines the relationships between objects using four fundamental categorical propositions. Upon closer inspection, complex propositions reveal themselves as collections of simpler claims derived from these initial propositions. Specifically, we utilize two terms from Aristotle’s theory: (1) *Universal Affirmation*: “All  $S$  are  $P$ ,” and (2) *Universal Negation*: “No  $S$  are  $P$ ,” to universally assert properties for all group members, indicating strong constraints and relationships. For example, in the context of two kidney structures,  $A$  and  $B$ , if  $B$  is within  $A$ ,  $B$  is a subset of  $A$ , following the rule of Universal Affirmation. Conversely, if  $A$  and  $B$  have no inclusion relationship, they are mutually exclusive, aligning with Universal Negation. These propositions are employed to construct an anatomical relationship map representing the classification and spatial relationships among renal pathology objects, as illustrated in Figure 2.

This anatomical relationship is characterized by:

**Uniqueness:** Each pair of objects is linked by a single proposition. The expanding structure of the map, devoid of cycles, ensures stable inheritance relationships from regions down to cellular levels.

**Transmissibility:** Indirect relationships between objects can be deduced from direct relationships, as established by the two fundamental categorical propositions. Relationships between objects not directly connected can be determined by combining propositions along their connecting path.

Following the translation of this knowledge from clinical anatomy to an engineering paradigm, we introduce a universal proposition matrix,  $M_t \in \mathbb{R}^{n \times n}$ , to facilitate im-

plementation in computational models. Here,  $n$  represents the number of classes within the map. The matrix values are defined in Equation (1):

$$M_t(i, j) = \begin{cases} 1, & \text{if } I \subseteq J \\ -1, & \text{if } I \supseteq J \\ 2, & \text{if } I \cap J = \emptyset \\ 0, & \text{otherwise} \end{cases} \quad i, j = 1, 2, \dots, n \quad (1)$$

This matrix is designed to be extendable with the introduction of new data.

### 3.3. Token-based Dynamic Head Network

Pathological image segmentation faces three main challenges: (1) Heterogeneous object annotations are often partially labeled, with only one type of tissue annotated per pathological image; (2) It is challenging to form anatomical relationships (e.g. subset/superset relationships) in multi-class segmentation; for example, it is difficult to simultaneously segment glomerular capsule, tufts, and cells that share regions; (3) The annotation process on giga-pixel images is labor-intensive, leading to an ongoing data collection process with class extension. Therefore, a segmentation backbone optimized for binary segmentation of multiple classes with spatial overlap, and adaptable to data enlargement, is required. Previous designs, including multi-head and dynamic-head, are suboptimal for partially labeled learning and data extension, leading to changes in the backbone architecture and insufficient performance.

In this work, we propose a token-based dynamic head backbone designed to maintain consistent model architecture while accommodating a possibly increasing number of segmentation classes (In Figure 4). The backbone can be supervised with partially labeled data, understanding anatomical relationships effectively. The entire architec-

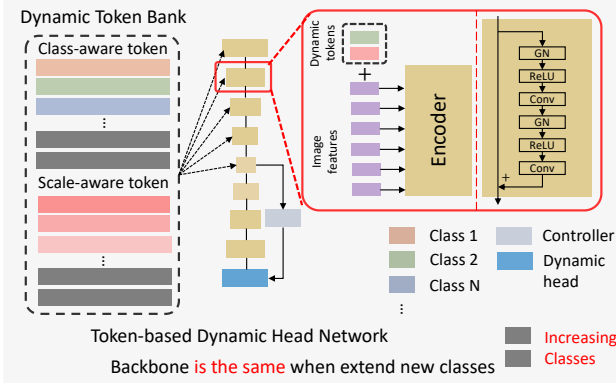


Figure 4. **Token-based dynamic head** – This figure visualizes the architecture of our proposed token-based dynamic head backbone. Central to our design is the ability to maintain a consistent model architecture while dynamically accommodating an increasing number of segmentation classes. This flexibility is achieved by extending the dimensions of the tokens, rather than altering the backbone structure. Key components include a dynamic token bank with class-aware and scale-aware tokens, an encoder, and a dynamic head network, all orchestrated to efficiently handle class expansion without necessitating changes to the backbone.

ture is demonstrated in Figure 3. The backbone of our proposed method is the Residual-U-Net from Omni-Seg [11], chosen for its superior segmentation performance in renal pathology. Instead of using dimensionally-changeable one-hot vectors for class-aware encoding and scale-aware encoding, we use dimensionally-stable class-aware tokens ( $T_c \in \mathbb{R}^{n \times d}$ ) and scale-aware tokens ( $T_s \in \mathbb{R}^{4 \times d}$ ) to comprehend the contextual information in the encoder ( $E$ ) of the model. Here,  $d$  is the sum of the channel numbers of blocks ( $d_1 + d_2 + \dots + d_b + d_{gap}$ ) in the encoder. Each interval of the channels represents level-specific features in each level of the encoder. Each class has a one-dimensional token  $t_c \in \mathbb{R}^{1 \times d}$  to store class-specific knowledge at the feature level among the whole dataset, while each magnification has a one-dimensional token  $t_s \in \mathbb{R}^{1 \times d}$  for providing scale-specific knowledge across four scales (5 $\times$ , 10 $\times$ , 20 $\times$ , and 40 $\times$ ). Inspired by the Vision Transformer [13], for an image  $I$  of class  $i$  with magnification  $m$ , at the  $b$ -th encoder block, the corresponding class token  $T_c(i)$  and  $T_s(m)$  are added to the current feature map ( $e_{b-1}$ ) before being fed into the current block ( $E_b$ ). This process is defined by the following equation:

$$\text{Ind}_{start} = \sum_{k=1}^{b-1} d_k, \text{Ind}_{end} = \sum_{k=1}^b d_k \quad (2)$$

$$e_b = E_b(T_c[i][\text{Ind}_{start} : \text{Ind}_{end}] + T_s[m][\text{Ind}_{start} : \text{Ind}_{end}] + e_{b-1}) \quad (3)$$

With the class token  $T_c$  and scale token  $T_s$ , the encoder captures domain-specific features in the image.

To further integrate class-specific and scale-specific information into the embedded features, we combine the last intervals of the class-token and scale token into the low-dimensional feature embedding at the bottom of the Residual-U-Net architecture. The image feature  $F$  is summarized by Global Average Pooling (GAP) and transformed into a feature vector in  $\mathbb{R}^{d_{gap}}$ .  $d_{gap}$  is the dimension of GAP feature. Differing from Omni-Seg [11], which implements a triple outer product to combine three vectors into a one-dimensional vector via a flatten function, we use a single 2D convolutional layer controller,  $\varphi$ , as a feature fusion block to refine the fusion vector as the final controller for dynamic head mapping:

$$\omega = \varphi(\text{GAP}(F) + T_c[i][:-d_{gap}] + T_s[i][:-d_{gap}]; \Theta_\varphi) \quad (4)$$

where  $\text{GAP}(F)$ ,  $T_c$ , and  $T_s$  are combined by the addition operation, and  $\Theta_\varphi$  represents the number of parameters in the dynamic head.

Following the approach of [11], a binary segmentation network is used to achieve multi-label segmentation via a dynamic filter. From the multi-label multi-scale modeling described above, we derive joint low-dimensional image feature vectors, class-specific tokens, and scale-specific tokens at an optimal segmentation magnification. These are then mapped to control a lightweight dynamic head, specifying (1) the target tissue type and (2) the corresponding pyramid scale.

The dynamic head consists of three layers. The first two layers contain eight channels each, while the final layer comprises two channels. We directly map parameters from the fusion-based feature controller to the kernels in the 162-parameter dynamic head to achieve precise segmentation from multi-modal features. The filtering process is expressed by the following equation:

$$P = (((M \cdot \omega_1) \cdot \omega_2) \cdot \omega_3) \quad (5)$$

where  $\cdot$  denotes convolution,  $P \in \mathbb{R}^{2 \times W \times H}$  is the final prediction, and  $W$  and  $H$  correspond to the width and height of the dataset, respectively.

The benefits of the dynamic-token design are twofold: (1) The backbone architecture remains stable and reusable as new classes are introduced, which is advantageous for incremental learning. (2) The binary segmentation scheme allows the model to predict multiple classes with spatial overlap, outperforming other multi-head and dynamic-head designs.

### 3.4. Anatomy Loss Function

With the introduction of the universal proposition matrix and token-based dynamic head architecture, we propose an online semi-supervised anatomy learning strategy to incorporate spatial correlation into the training process for com-

prehensive segmentation. For a given image  $I$  with a labeled class  $i$ , represented as  $Y_i$ , we generate predictions  $Y'_j$  for another class  $j$  on the same image. We then use the anatomical relationship to supervise the correlation between the supervised label  $Y_i$  and the semi-supervised prediction  $Y'_j$ : (1) If  $i$  is a superset of  $j$ ,  $Y'_j$  should not exceed the region of  $Y_i$ ; conversely, (2) if  $i$  is a subset of  $j$ ,  $Y'_j$  should cover  $Y_i$  as comprehensively as possible; and (3) if  $i$  is mutually exclusive with  $j$ , the overlap between  $Y_i$  and  $Y'_j$  should be minimized. The total anatomy loss is defined by the following equations:

$$L_{\text{upl}}(i, j) = \begin{cases} \text{DCE}(1 - Y_i, Y'_j), & \text{if } M_t(i, j) = 1 \\ -\text{DCE}(Y_i, Y_i \cup Y'_j), & \text{if } M_t(i, j) = -1 \\ \text{DCE}(Y_i, Y'_j), & \text{if } M_t(i, j) = 2 \\ 0, & \text{if } M_t(i, j) = 0 \end{cases} \quad (6)$$

where  $M_t$  is the anatomy matrix and DCE denotes the Dice Loss. The total loss function is an aggregate of supervised and semi-supervised losses, weighted by  $\lambda_{\text{upl}}$ .

$$L(i) = \text{DCE}(Y_i, Y'_i) + \text{BCE}(Y_i, Y'_i) + \lambda_{\text{upl}} \sum_{j=1}^n L_{\text{upl}}(i, j) \quad (j \neq i) \quad (7)$$

where BCE represents the Binary Cross-Entropy loss, and  $Y'_i$  is the prediction for class  $i$ .

## 4. Data and Experiment

### 4.1. Data

Our model leverages an 8-class, partially labelled dataset spanning various biological scales, from regions to cells. The dataset’s structure is detailed in Table 1. We sourced the human kidney dataset from three distinct resources:

**Regions:** Whole slide images of wedge kidney sections stained with periodic acid-Schiff (PAS, n=138) were obtained from non-cancerous regions of nephrectomy samples. The samples were categorized into several groups based on clinical data, including normal adults (n=27), patients with hypertension (HTN, n=31), patients with diabetes (DM, n=4), patients with both hypertension and diabetes (n=14), normal aging individuals (age>65y, n=10), individuals with aging and hypertension (n=36), and individuals with aging, hypertension, and diabetes (n=16). These tissues were scanned at 20× magnification and manually annotated in QuPath [2], delineating medulla and cortex contours. The WSIs were downsampled to 5× magnification and segmented into 1024×1024 pixel patches. Corresponding binary masks were derived from the contours.

**Functional Units:** Using 459 WSIs from NEPTUNE study [3], encompassing 125 patients with minimal change disease, we extracted 1,751 Regions of Interest (ROIs). These ROIs were manually segmented to identify four kinds

Table 1. Data collection

Class	Patch #	Size	Scale	Stain
Medulla	1619	1024 <sup>2</sup>	5 ×	P
Cortex	3055	1024 <sup>2</sup>	5 ×	P
DT	4615	256 <sup>2</sup>	10 ×	H,P,S,T
PT	4588	256 <sup>2</sup>	10 ×	H,P,S,T
Cap.	4559	256 <sup>2</sup>	5 ×	H,P,S,T
Tuft	4536	256 <sup>2</sup>	5 ×	H,P,S,T
Pod.	1147	512 <sup>2</sup>	20 ×	P
Mes.	789	512 <sup>2</sup>	20 ×	P

\*DT is distal tubular; PT is proximal tubular;  
 \*Cap. is glomerular capsule; Tuft is glomerular tuft;  
 \*Pod. is podocyte cell; Mes. is mesangial cell.  
 \*H is H&E; P is PAS; S is SIL; T is TRI.

of morphology objects with normal structure and methodology outlined in [27]. Each image, at a resolution of 3000×3000 pixels (40× magnification, 0.25 μm per pixel), represented one of four tissue types stained with Hematoxylin and Eosin Stain(H&E), PAS, Silver Stain (SIL), and Trichrome Stain (TRI). We treated these four staining methods as color augmentations and resized the images to 256×256 pixels, maintaining the original data splits from [27].

**Cells:** We acquired 11 PAS-stained WSIs from nephrectomy specimens with normal kidney function and morphology, scanned at 20× magnification. These pathological images were cropped into 512×512 pixel segments to facilitate cell labeling, following the annotation process described in [10].

The dataset was partitioned into training, validation, and testing sets at a 6:1:3 ratio across all classes, with splits conducted at the patient level to prevent data leakage.

### 4.2. Experiment Details

The training process of our model was divided into two distinct phases. In the initial phase, spanning the first 50 epochs, we employed a supervised learning strategy focused on minimizing binary dice loss and cross-entropy loss. Subsequently, for the remaining epochs, both supervised and semi-supervised learning strategies were utilized, incorporating anatomy loss to explore the spatial correlation among multiple objects. In our experiments,  $\lambda_{\text{upl}}$  is 0.1.

All images were either randomly cropped or padded to a uniform size of 512 × 512 pixels prior to being fed into the model in the training stage. We established 8 separate image pools, each designated for different tissue types, to organize training batches. This approach follows the image pooling strategy from Cycle-GAN [54]. The batch size was set to 4, while each image pool could accommodate up to 8 images. Once an image pool accumulated more than the batch size,

Table 2. Performance of deep learning based multi-class panoramic segmentation. Dice similarity coefficient scores (%) are reported.

Method	Backbone	Regions		Functional units				Cells		Average
		Medulla	Cortex	DT	PT	Cap.	Tufts	Pod.	Mes.	
U-Nets [20]	CNN	23.86	66.42	47.61	51.04	45.36	46.62	49.92	49.87	47.58
DeepLabV3 [39]	CNN	41.70	61.26	63.92	65.31	72.82	81.93	49.92	49.87	60.84
Residual-U-Net [45]	CNN	13.21	69.97	67.03	76.59	70.58	82.37	63.99	64.54	63.54
Multi-kidney [6]	CNN	13.13	69.77	61.58	62.13	82.27	62.03	<b>71.82</b>	65.13	60.98
Omni-Seg [11]	CNN	<b>72.96</b>	71.84	69.76	81.48	92.31	92.36	67.28	65.31	76.66
Segmenter [46]	Transformer	56.38	67.34	54.81	69.14	67.16	66.78	49.92	49.87	60.18
SegFormer [49]	Transformer	54.82	67.68	62.65	75.87	77.46	60.42	62.43	60.21	65.19
Unetr [24]	Transformer	16.22	69.70	61.99	69.35	72.48	58.10	58.14	56.32	57.79
Swin-Unetr [23]	Transformer	13.81	68.92	70.76	76.93	81.28	72.92	49.92	64.46	62.38
PrPSeg (Ours)	CNN	72.38	<b>72.64</b>	<b>72.45</b>	<b>85.27</b>	<b>94.23</b>	<b>94.40</b>	70.98	<b>66.96</b>	<b>78.66</b>

\*DT is distal tubular; PT is proximal tubular; Cap. is glomerular capsule

\*Pod. is glomerular podocyte cell; Mes. is glomerular mesangial cell

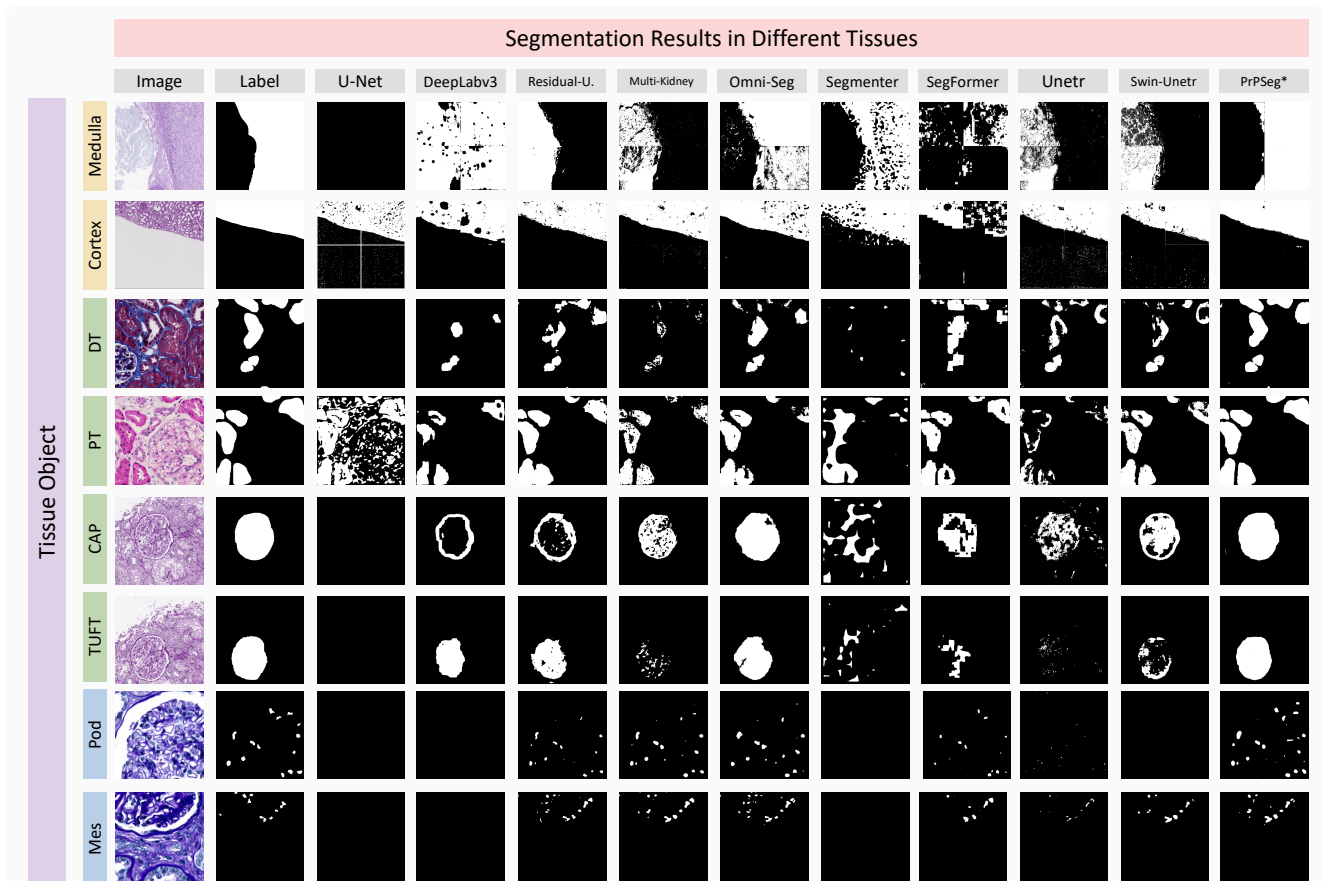


Figure 5. **Validation qualitative results** – This figure shows the qualitative results of different approaches. The proposed method achieved superior panoramic renal pathology segmentation on 8 classes range regions to cells with fewer false positives, false negatives, and morphological errors.

the images were retrieved from the pool and input into the network for processing. During each backpropagation step, Binary Dice Loss and Cross-entropy Loss were combined

as the loss function in the supervised learning phase.

For weight optimization, we employed the Adam optimizer with an initial learning rate of 0.001 and a de-

causal factor of 0.99. We also implemented general data augmentation techniques, including Affine transformations, Flip, Contrast adjustment, Brightness adjustment, Coarse Dropout, Gaussian Blur, and Gaussian Noise. These augmentations, sourced from the `imgaug` package [31], were applied to the entire training dataset with a probability of 0.5.

Model selection was based on the mean Dice coefficient score across 8 classes, evaluated on the validation dataset. The best-performing models within the first 100 epochs were then assessed on the testing dataset. Testing images were initially processed using either center-cropping or non-overlapping tiling to attain the same uniform size  $512 \times 512$  pixels. For evaluation, these images were subsequently either center-cropped or re-aggregated to their original dimensions. All experiments were conducted on a uniform platform, specifically a workstation equipped with an NVIDIA RTX A6000 GPU.

## 5. Result

We conducted a comparative analysis of our proposed Universal Proposition Learning approach against various baseline models. These models include multi-class segmentation architectures such as (1) U-Nets [20], (2) DeepLabV3 [39], (3) Residual-U-Net [45], (4) a CNN-based multi-class kidney pathology model [6], (5) Omni-Seg [11], (6) Segmenter-ViT/16 [46], (7) SegFormer [49], (8) Unetr [24], and (9) Swin-Unetr [23].

### 5.1. Panoramic Segmentation Performance

Table 2 and Figure 5 showcase the results from an 8-class segmentation evaluation. Table 2 demonstrates that our proposed method, PrPSeg, surpasses baseline models in most evaluated metrics. Figure 5 further highlights the qualitative superiority of our approach, evidenced by reduced instances of false positives, false negatives, and morphological errors. The Dice similarity coefficient (Dice: %, the higher, the better) was employed as the primary metric for quantitative performance assessment. The results indicate that, while multi-head designs struggle with managing spatial relationships between objects (e.g., subset/superset relationships between the capsule and tuft), the dynamic-head paradigm exhibits superior performance compared to other methods.

### 5.2. Ablation Study

Table 3 showcases the enhancements brought about by our proposed token-based design and learning strategy across two different backbone architectures. The results indicate that the token-based dynamic head design boosts the model’s performance in segmenting all levels of objects. With the integration of the token-based dynamic head architecture and universal proposition learning, the proposed

Table 3. Ablation study of different design. Dice similarity coefficient scores (%) are reported.

Backbone	TDH	UPL	Regions	Units	Cells	Average
Swin-Unetr [23]			41.37	75.47	57.19	62.38
Swin-Unetr [23]	✓		68.55	82.70	49.90	70.97
Omni-Seg [11]			72.40	83.98	66.29	76.66
Omni-Seg [11]	✓		72.43	86.39	66.49	77.89
PrPSeg (Ours)	✓	✓	<b>72.51</b>	<b>86.58</b>	<b>68.97</b>	<b>78.66</b>

\*TDH is Token-based Dynamic Head

\*UPL is Universal Proposition Learning

method exhibited superior performance across all considered metrics.

We also provide an ablation study for two data extension scenarios in the Appendix (A.1). The proposed PrPSeg method is flexible to extend to new classes by merely updating tokens and the adaptable proposition matrix, without changing the backbone network, while demonstrating superior performance compared to baseline methods across all seven new classes.

## 6. Conclusion

In this work, we have developed PrPSeg, a token-based dynamic segmentation network, specifically crafted to facilitate panoramic renal pathology segmentation by effectively modeling the spatial interconnections among diverse anatomical structures. This innovative approach enables the consistent use of the same architectural framework amidst dataset expansions and introduces a universal proposition matrix. This matrix adeptly transforms intricate anatomical relationships into computational modeling paradigms. Furthermore, we have introduced a novel anatomical loss function, integrating these spatial relationships into our model’s training regimen through semi-supervised learning. To the best of our knowledge, our algorithm is the first to achieve comprehensive panoramic segmentation in the domain of renal pathology. The integration of token-based dynamic head design alongside our universal proposition learning strategy, which meticulously maps anatomical relationships into the realm of engineering programming, enhances the model’s efficacy in all-encompassing segmentation.

## 7. Acknowledgement

This research was supported by NIH R01DK135597(Huo), DoD HT9425-23-1-0003(HCY), NIH NIDDK DK56942(ABF). This work was also supported by Vanderbilt Seed Success Grant, Vanderbilt Discovery Grant, and VISE Seed Grant. This research was also supported by NIH grants R01EB033385, R01DK132338, REB017230, R01MH125931, and NSF 2040462. We extend gratitude to NVIDIA for their support by means of the NVIDIA hardware grant.



## References

- [1] Jathin Bandari, Thomas W Fuller, Robert M Turner, and Louis A D'Agostino. Renal biopsy for medical renal disease: indications and contraindications. *Can J Urol*, 23(1): 8121–6, 2016. **1**
- [2] Peter Bankhead, Maurice B Loughrey, José A Fernández, Yvonne Dombrowski, Darragh G McArt, Philip D Dunne, Stephen McQuaid, Ronan T Gray, Liam J Murray, Helen G Coleman, et al. Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7, 2017. **6**
- [3] 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. **6**
- [4] T de Bel, Meyke Hermsen, Geert Litjens, and J Laak. Structure instance segmentation in renal tissue: a case study on tubular immune cell detection. In *Computational Pathology and Ophthalmic Medical Image Analysis*, pages 112–119. Springer, 2018. **1**
- [5] Ewert Bengtsson, Håvard Danielsen, Darren Treanor, Metin N Gurcan, Calum MacAulay, and Béla Molnár. Computer-aided diagnostics in digital pathology, 2017. **1**
- [6] 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. **2, 7, 8**
- [7] 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*, 184: 105273, 2020. **2**
- [8] Richard J. Chen, Chengkuan Chen, Yicong Li, Tiffany Y. Chen, Andrew D. Trister, Rahul G. Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16144–16155, 2022. **2**
- [9] Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019. **2**
- [10] Ruining Deng, Yanwei Li, Peize Li, Jiacheng Wang, Lucas W Remedios, Saydolimkhon Agzamkhodjaev, Zuhayr Asad, Quan Liu, Can Cui, Yaohong Wang, et al. Democratizing pathological image segmentation with lay annotators via molecular-empowered learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 497–507. Springer, 2023. **6**
- [11] Ruining Deng, Quan Liu, Can Cui, Tianyuan Yao, Jun Long, Zuhayr Asad, R Michael Womick, Zheyu Zhu, Agnes B Fogo, Shilin Zhao, et al. Omni-seg: A scale-aware dynamic network for renal pathological image segmentation. *IEEE Transactions on Biomedical Engineering*, 2023. **1, 2, 5, 7, 8**
- [12] Huijun Ding, Zhanpeng Pan, Qian Cen, Yang Li, and Shifeng Chen. Multi-scale fully convolutional network for gland segmentation using three-class classification. *Neurocomputing*, 380:150–161, 2020. **1**
- [13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. **5**
- [14] Xi Fang and Pingkun Yan. Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction. *IEEE Transactions on Medical Imaging*, 39(11):3619–3629, 2020. **2**
- [15] Chunyue Feng and Fei Liu. Artificial intelligence in renal pathology: Current status and future. *Bosnian Journal of Basic Medical Sciences*, 2022. **2**
- [16] M Gadermayr, AK Dombrowski, BM Klinkhammer, P Boor, and D Merhof. Cnn cascades for segmenting whole slide images of the kidney. arxiv 2017. *arXiv preprint arXiv:1708.00251*. **2**
- [17] Jaime Gallego, Anibal Pedraza, Samuel Lopez, Georg Steiner, Lucia Gonzalez, Arvydas Laurinavicius, and Gloria Bueno. Glomerulus classification and detection based on convolutional neural networks. *Journal of Imaging*, 4(1):20, 2018. **2**
- [18] Zeyu Gao, Bangyang Hong, Xianli Zhang, Yang Li, Chang Jia, Jialun Wu, Chunbao Wang, Deyu Meng, and Chen Li. Instance-based vision transformer for subtyping of papillary renal cell carcinoma in histopathological image. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pages 299–308. Springer, 2021. **2**
- [19] Jeremias Gomes, Jun Kong, Tahsin Kurc, Alba CMA Melo, Renato Ferreira, Joel H Saltz, and George Teodoro. Building robust pathology image analyses with uncertainty quantification. *Computer Methods and Programs in Biomedicine*, 208: 106291, 2021. **1**
- [20] Germán González, George R Washko, and Raúl San José Estépar. Multi-structure segmentation from partially labeled datasets. application to body composition measurements on ct scans. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 215–224. Springer, 2018. **2, 7, 8**
- [21] Laxmi Gupta, Barbara Mara Klinkhammer, Peter Boor, Dorit Merhof, and Michael Gadermayr. Iterative learning to make the most of unlabeled and quickly obtained labeled data in histology. In *International Conference on Medical Imaging with Deep Learning—Full Paper Track*, 2018. **2**
- [22] Satoshi Hara, Emi Haneda, Masaki Kawakami, Kento Morita, Ryo Nishioka, Takeshi Zoshima, Mitsuhiro Kometani, Takashi Yoneda, Mitsuhiro Kawano, Shigehiro Karashima, et al. Evaluating tubulointerstitial compartments in renal biopsy specimens using a deep learning-based ap-

- proach for classifying normal and abnormal tubules. *PLoS one*, 17(7):e0271161, 2022. 2
- [23] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI Brainlesion Workshop*, pages 272–284. Springer, 2021. 7, 8
- [24] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584, 2022. 7, 8
- [25] Meyke Hermsen, Thomas de Bel, Marjolijn Den Boer, Eric J Steenbergen, Jesper Kers, Sandrine Florquin, Joris JTH Roelofs, Mark D Stegall, Mariam P Alexander, Byron H Smith, et al. Deep learning–based histopathologic assessment of kidney tissue. *Journal of the American Society of Nephrology*, 30(10):1968–1979, 2019. 1
- [26] John D Imig, Xueying Zhao, Ahmed A Elmarakby, and Tengis Pavlov. Interactions between podocytes, mesangial cells, and glomerular endothelial cells in glomerular diseases. *Frontiers in Physiology*, page 488, 2022. 1
- [27] Catherine P Jayapandian, Yijiang Chen, Andrew R Janowczyk, Matthew B Palmer, Clarissa A Cassol, Mirosław Sekulic, Jeffrey B Hodgins, Jarcy Zee, Stephen M Hewitt, John O’Toole, et al. Development and evaluation of deep learning–based segmentation of histologic structures in the kidney cortex with multiple histologic stains. *Kidney international*, 99(1):86–101, 2021. 1, 6
- [28] JoséA Jiménez-Heffernan, M Auxiliadora Bajo, Cristian Perna, Gloria del Peso, Juan R Larrubia, Carlos Gamallo, JoséA Sánchez-Tomero, Manuel López-Cabrera, and Rafael Selgas. Mast cell quantification in normal peritoneum and during peritoneal dialysis treatment. *Archives of pathology & laboratory medicine*, 130(8):1188–1192, 2006. 1
- [29] Tan Yee Jing, Nazahah Mustafa, Haniza Yazid, and Khairul Shakir Ab Rahman. Segmentation of tumour regions for tubule formation assessment on breast cancer histopathology images. In *Proceedings of the 11th International Conference on Robotics, Vision, Signal Processing and Power Applications*, pages 170–176. Springer, 2022. 2
- [30] Jeremiah W Johnson. Automatic nucleus segmentation with mask-rcnn. In *Science and Information Conference*, pages 399–407. Springer, 2019. 2
- [31] Alexander B. Jung, Kentaro Wada, Jon Crall, Satoshi Tanaka, Jake Graving, Christoph Reinders, Sarthak Yadav, Joy Banerjee, Gábor Vecsei, Adam Kraft, Zheng Rui, Jirka Borovec, Christian Vallentin, Semen Zhydenko, Kilian Pfeiffer, Ben Cook, Ismael Fernández, François-Michel De Rainville, Chi-Hung Weng, Abner Ayala-Acevedo, Raphael Meudec, Matias Laporte, et al. imgaug. <https://github.com/aleju/imgaug>, 2020. Online; accessed 01-Feb-2020. 8
- [32] Shruti Kannan, Laura A Morgan, Benjamin Liang, McKenzie G Cheung, Christopher Q Lin, Dan Mun, Ralph G Nader, Mostafa E Belghasem, Joel M Henderson, Jean M Francis, et al. Segmentation of glomeruli within trichrome images using deep learning. *Kidney international reports*, 4(7):955–962, 2019. 2
- [33] Tsung-Wei Ke, Jyh-Jing Hwang, Yunhui Guo, Xudong Wang, and Stella X. Yu. Unsupervised hierarchical semantic segmentation with multiview cosegmentation and clustering transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2
- [34] John A Kellum. Acute kidney injury. *Critical care medicine*, 36(4):S141–S145, 2008. 1
- [35] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36(7):1550–1560, 2017. 1
- [36] Liulei Li, Tianfei Zhou, Wenguan Wang, Jianwu Li, and Yi Yang. Deep hierarchical semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1246–1257, 2022. 2
- [37] Yilong Li, Xingru Huang, Yaqi Wang, Zhaoyang Xu, Yibao Sun, and Qianni Zhang. U-net ensemble model for segmentation in histopathology images. 2019. 1
- [38] Zhenrong Lin, Jidong Li, Qipeng Yao, Haocheng Shen, and Lihang Wan. Adversarial learning with data selection for cross-domain histopathological breast cancer segmentation. *Multimedia Tools and Applications*, 81(4):5989–6008, 2022. 2
- [39] 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. 2, 7, 8
- [40] Elise Marechal, Adrien Jaugey, Georges Tarris, Michel Paindavoine, Jean Seibel, Laurent Martin, Mathilde Funes de la Vega, Thomas Crepin, Didier Ducloux, Gilbert Zanetta, et al. Automatic evaluation of histological prognostic factors using two consecutive convolutional neural networks on kidney samples. *Clinical Journal of the American Society of Nephrology*, 17(2):260–270, 2022. 2
- [41] David Marti-Aguado, Alejandro Rodríguez-Ortega, Claudia Mestre-Alagarda, Mónica Bauza, Elena Valero-Pérez, Clara Alfaro-Cervello, Salvador Benlloch, Judith Pérez-Rojas, Antonio Ferrández, Pilar Alemany-Monraval, et al. Digital pathology: accurate technique for quantitative assessment of histological features in metabolic-associated fatty liver disease. *Alimentary Pharmacology & Therapeutics*, 53(1):160–171, 2021. 1
- [42] Claire Mounier-Vehier, Christophe Lions, Patrick Devos, Olivier Jaboureck, Serge Willoteaux, Alain Carre, and Jean-Paul Beregi. Cortical thickness: an early morphological marker of atherosclerotic renal disease. *Kidney international*, 61(2):591–598, 2002. 1
- [43] Cam Nguyen, Zuhayr Asad, and Yuankai Huo. Evaluating transformer-based semantic segmentation networks for pathological image segmentation. *arXiv preprint arXiv:2108.11993*, 2021. 2

- [44] Jian Ren, Evita Sadimin, David J Foran, and Xin Qi. Computer aided analysis of prostate histopathology images to support a refined gleason grading system. In *Medical Imaging 2017: Image Processing*, page 101331V. International Society for Optics and Photonics, 2017. [1](#)
- [45] Massimo Salvi, Alessandro Mogetta, Alessandro Gambella, Luca Molinaro, Antonella Barreca, Mauro Papotti, and Filippo Molinari. Automated assessment of glomerulosclerosis and tubular atrophy using deep learning. *Computerized Medical Imaging and Graphics*, 90:101930, 2021. [7](#), [8](#)
- [46] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7262–7272, 2021. [7](#), [8](#)
- [47] Julia Wijkström, Channa Jayasumana, Rajeewa Dasanayake, Nalin Priyawardane, Nimali Godakanda, Sisira Siribaddana, Anneli Ring, Kjell Hultenby, Magnus Söderberg, Carl-Gustaf Elinder, et al. Morphological and clinical findings in sri lankan patients with chronic kidney disease of unknown cause (ckdu): Similarities and differences with mesoamerican nephropathy. *PloS one*, 13(3):e0193056, 2018. [1](#)
- [48] Hao Wu, Shuchao Pang, and Arcot Sowmya. Tgnet: A task-guided network architecture for multi-organ and tumour segmentation from partially labelled datasets. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022. [1](#)
- [49] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021. [7](#), [8](#)
- [50] Shaoshuai Yan, Xiangsheng Huang, Wei Lian, and Caifang Song. Self reinforcing multi-class transformer for kidney glomerular basement membrane segmentation. *IEEE Access*, 2023. [2](#)
- [51] Caihong Zeng, Yang Nan, Feng Xu, Qunjuan Lei, Fengyi Li, Tingyu Chen, Shaoshan Liang, Xiaoshuai Hou, Bin Lv, Dandan Liang, et al. Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning. *The Journal of pathology*, 252(1):53–64, 2020. [1](#)
- [52] Jianpeng Zhang, Yutong Xie, Yong Xia, and Chunhua Shen. Dodnet: Learning to segment multi-organ and tumors from multiple partially labeled datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1195–1204, 2021. [1](#), [2](#)
- [53] Yi Zheng, Clarissa A Cassol, Saemi Jung, Divya Veerapaneni, Vipul C Chitalia, Kevin YM Ren, Shubha S Bellur, Peter Boor, Laura M Barisoni, Sushrut S Waikar, et al. Deep-learning-driven quantification of interstitial fibrosis in digitized kidney biopsies. *The American journal of pathology*, 191(8):1442–1453, 2021. [1](#)
- [54] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. [6](#)