An automatic framework to study the tissue micro-environment of renal glomeruli in differently stained consecutive digital whole slide images

Odyssée Merveille, Thomas Lampert, Jessica Schmitz, Germain Forestier, Friedrich Feuerhake, Cédric Wemmert

An automatic framework to study the tissue micro-environment of renal glomeruli in differently stained consecutive digital whole slide images

Odyssee Merveille, Thomas Lampert, Jessica Schmitz, Germain Forestier, Friedrich Feuerhake, Cédric Wemmert

***Abstract*— *Objective:* This article presents an automatic image processing framework to extract quantitative high-level information describing the micro-environment of glomeruli in consecutive whole slide images (WSIs) processed with different staining modalities of patients with chronic kidney rejection after kidney transplantation. *Methods:* This three step framework consists of: 1) cell and anatomical structure segmentation based on colour deconvolution and deep learning 2) fusion of information from different stainings using a newly developed registration algorithm 3) feature extraction. *Results:* Each step of the framework is validated independently both quantitatively and qualitatively by pathologists. An illustration of the different types of features that can be extracted is presented.**

***Conclusion:* The proposed generic framework allows for the analysis of the micro-environment surrounding large structures that can be segmented (either manually or automatically). It is independent of the segmentation approach and is therefore applicable to a variety of biomedical research questions. *Significance:* Chronic tissue remodelling processes after kidney transplantation can result in interstitial fibrosis and tubular atrophy (IFTA) and glomerulosclerosis. This pipeline provides tools to quantitatively analyse, in the same spatial context, information from different consecutive WSIs and help researchers understand the complex underlying mechanisms leading to IFTA and glomerulosclerosis.**

Manuscript received June ??, 2020. This work was supported by 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 the the e:Med project SYSIMIT, project management DLR (FKZ: 01ZX1608B), and Agence National de la Recherche (ANR), project number ANR-15-CMED-0004.

O. Merveille is with ICube, University of Strasbourg, CNRS (UMR 7357), Strasbourg, France and Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69XXX, LYON, France (e-mail: odyssee.merveille@creatis.insa-lyon.fr).

T. Lampert is with ICube, University of Strasbourg, CNRS (UMR 7357), Strasbourg, France (e-mail: lampert@unistra.fr).

J. Schmitz is with the Institute of Pathology, Hannover Medical School, Germany (e-mail: Schmitz.Jessica@mh-hannover.de).

G. Forestier is with IRIMAS, Université de Haute Alsace, Mulhouse, France (e-mail: germain.forestier@uha.fr).

F. Feuerhake is with Institute of Pathology, Hannover Medical School, Germany and University Clinic, Freiburg, Germany (e-mail: Feuerhake.Friedrich@mh-hannover.de).

C. Wemmert is with ICube, University of Strasbourg, CNRS (UMR 7357), Strasbourg, France (e-mail: wemmert@unistra.fr).

***Index Terms*— digital pathology, brightfield images, chromogenic duplex immunohistochemistry, digital whole slide image, glomeruli segmentation, glomeruli matching**

I. Introduction

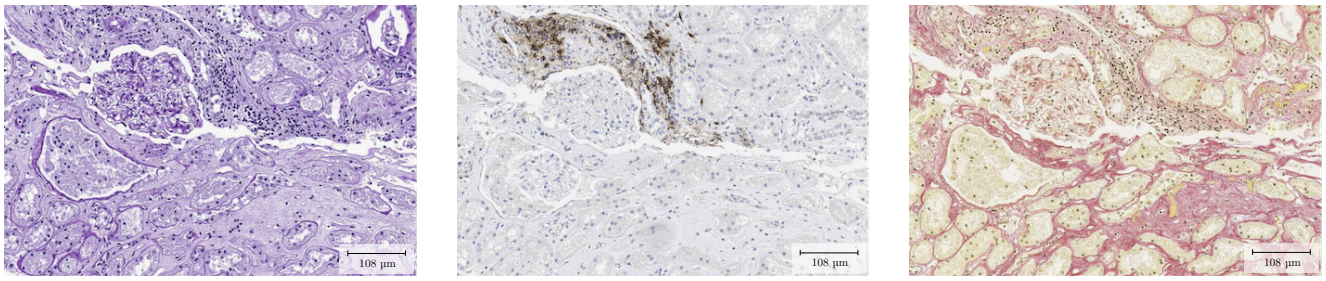
Kidney transplantation is performed at an annual rate of more than 90 000 world-wide[1]. Kidney replacement therapy after renal failure can restore renal function for many years, thereby reducing the burden for individual patients and for health systems that are associated with hemodialyis. In the past decades, successful therapy strategies were developed to avoid acute rejection, and substantially reduce the risk of chronic rejection. This shifted the attention towards slowly progressing fibrotic changes that can contribute to the decline of graft function.

Chronic tissue remodeling is histologically characterised by the appearance of Interstitial Fibrosis and Tubular Atrophy (IFTA) and glomerulosclerosis. In recent years, works studying the mechanisms leading to these pathologies have been carried out [1], [2]. In particular, macrophages have recently been identified as a key player in the inflammation and fibrosis process [3]. Depending on their phenotype ("M1-like" or "M2-like"), macrophages can be pro or anti-inflammatory and they also play a role in the activation of fibroblasts inducing IFTA and glomerulosclerosis.

A common approach in histopathology is the visual evaluation of consecutive sections of a biopsy by trained pathologists who integrate the information from several stainings, each one providing specific information on the tissue (see Fig. 1), into a written report.

With the emergence of system biomedicine, there has been an increasing trend to study complex mechanisms based on quantitative data such as inflammation [4], [5], cancer clonal evolution [6], or immune reactions [7]. In this context, Whole Slide Images (WSI) with different stainings are studied separately, and the fusion of information from these different stainings is required to obtain a comprehensive data set. Pathologists mentally perform this fusion while analysing a piece of tissue. This trivial task for trained pathologists

[1] www.transplant-observatory.org/download/2017-activity-data-report



(a) Periodic Acid Schiff (PAS) (b) CD3 (c) Sirius Red

Fig. 1: Example of three consecutives WSI of a kidney nephrectomy sample with three common stainings. Each staining provide different information on the tissue: general structural information in PAS, distribution of T lymphocytes in CD3, specific structures such as collagen or muscular fibres in Sirius Red.

is highly complex for computers and requires specifically designed algorithms commonly referred to as slide registration.

WSI registration algorithms should indeed take into account several specificities intrinsic to histopathology: the tissue shape and orientation of two consecutive slides may vary because of the sample preparation (fixation, embedding, sectioning, etc.); the composition of the tissue between two slides can vary significantly as the cells and structures may appear, disappear, or have different structure depending on the sectioning level; finally, the set of stainings used may highlight different structures or cells which results in slides that look quite different (see Fig. 1).

Algorithms used for WSI registration in the literature usually apply non-rigid deformations resulting in visually pleasing registration. Most methods use the mutual information similarity metric to register two WSIs with different stainings [8]–[10] as it relies on statistical relations between the intensities of two stainings instead of direct correlations. Nevertheless, these methods may fail for stainings with very different appearances as they are only based on raw intensities. To overcome this, Cooper et al. [11] proposed to rely on purely geometric features for the registration. More recently, Song et al. [12] developed an unsupervised content classification algorithm that computes more complex features describing the structures of each image. Even though these non-rigid methods yield good visual registration, they introduce spatial deformations that induce significant bias in geometric features extraction.

II. Proposed approach

This article presents an automatic pipeline to analyse histology slides from patients with chronic renal graft rejection. Many features quantifying the inflammation can be extracted from this pipeline and used by pathologists for diagnosis purpose. More complex features, like spatial correlations between cell populations, can also be extracted to describe the tissue state and help researchers, clinicians and pathologists to better understand the mechanisms leading to IFTA and glomerulosclerosis.

Instead of applying non-rigid registration, we propose to merge the information from consecutive slides by finding common landmarks across stainings and locally superimpose the regions around these landmarks from each staining. Glomeruli are spherical structures with a diameter of around $150\,\mu\mathrm{m}$ and are one of the key functional units of the kidney, responsible for the filtration of primary urine from the blood. They are thereby good candidates for landmarks as they have a high probability to be present in several consecutive slides (the average slice of tissue is $3\,\mu\mathrm{m}$ thick) and they present an isotropic structure making them easily detected in each slide whatever the cutting direction.

In order to merge the information from several consecutive slides, we propose to match the glomeruli across slices and then locally superimpose each glomerulus neighbourhood to perform the multi-stain analysis. Thus, our framework is threefold: 1) cell and structures segmentation, 2) glomeruli matching, 3) feature extraction from different stainings in the same spatial reference. An overview of this framework is presented in Fig. 2.

Several contributions are presented in this article:

- A new staining registration strategy is proposed to avoid tissue deformation based on glomeruli matching. This matching algorithm is validated on real data showing robust performance.
- An automatic pipeline able to extract quantitative features from consecutive WSI with different stainings is proposed. Combined with the matching algorithm, this pipeline allow for the registration of features from different stainings in the same spatial context, leading to a global multi-stain analysis pipeline.

The remainder of this article is organised as follows: Section III presents the glomeruli matching algorithm and Section IV presents the complete analysis pipeline. Several experiments are conducted in Section V: The validation of the proposed matching algorithm both independently and in the context of the pipeline; the application of the complete pipeline to four consecutive nephrectomy WSIs; an illustration of several interesting features that can be computed from such an analysis framework.

III. Glomeruli Matching

Instead of applying a non-rigid registration that would deform the shape of the structures and the statistical properties of the neighbouring area, this section presents a novel glomeruli matching algorithm in order to locally superimpose glomeruli neighbourhoods between slices.

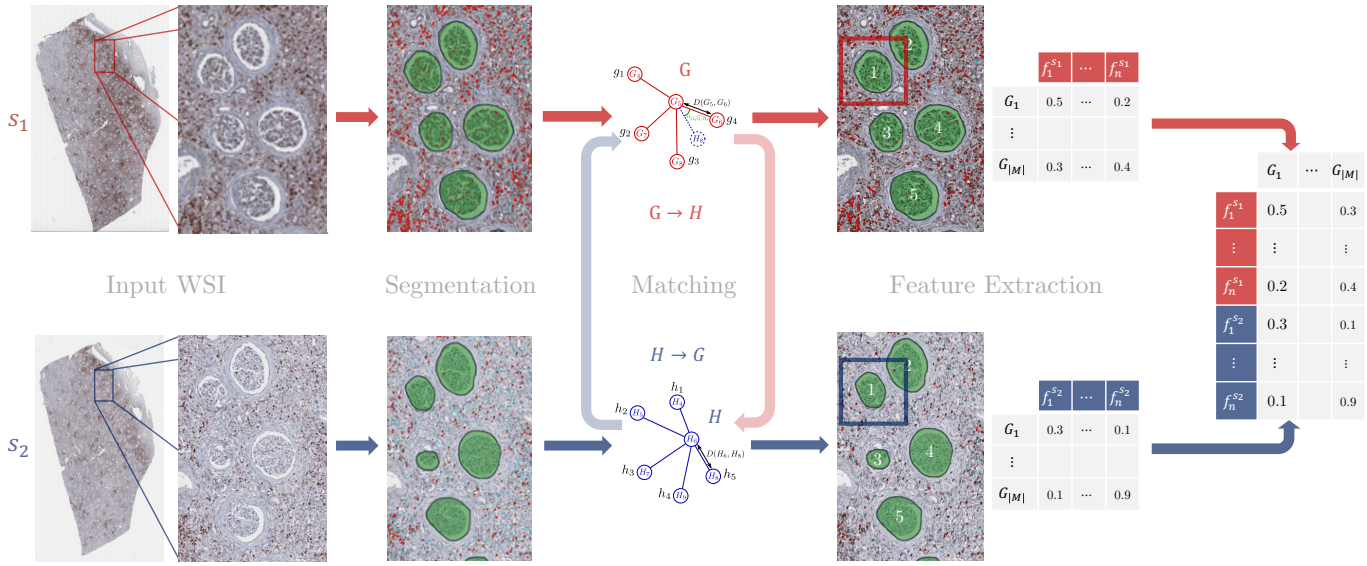


Fig. 2: Overview of the proposed analysis framework. Glomeruli segmentation is performed on two consecutive WSI with staining s_1 and s_2. Glomeruli matching is then performed from s_1 to s_2 ($G \longrightarrow H$) and from s_2 to s_1 ($H \longrightarrow G$). Only matchings that are found in both directions are kept. Patches around each matched glomeruli are extracted and for each of them several features are computed. Feature matrices from both stainings are concatenated so that in each column (i.e. for each matched glomeruli) we have features extracted from the two stainings s_1 (in red) and s_2 (in blue).

A. Method

Let G be the set of glomeruli in a WSI and H be the set of glomeruli in a WSI consecutive to it. The cardinality of a set A is denoted $|A|$, such that $|G|$ and $|H|$ are the number of glomeruli respectively in G and H.

Matching G to H can be seen as an inexact graph matching problem. Let $\mathcal{G} = (G, E_G)$ and $\mathcal{H} = (H, E_H)$ be two graphs where E_G (resp. E_H) is a set of edges between the glomeruli G (resp. H). The inexact matching problem is defined as

$$\hat{x} = \arg\min_{x \in X} \sum_{k} E(G_k, H_{x_k}), \tag{1}$$

where $X \in \mathbb{N}^{|G|}$ is the set of all possible matchings from $\mathcal{G}$ to $\mathcal{H}$ and $E : G \times H \rightarrow \mathbb{R}$ is a matching energy function.

Inexact graph matching is an NP-complete problem that is usually solved by finding an approximate solution using heuristic search strategies. In this work, the complexity of the global inexact graph matching problem is reduced by incorporating prior knowledge regarding the solution and adopting a subgraph assignment splitting strategy inspired by the work of Raveaux et al. [13].

The largest contribution to the complexity of general graph matching comes from the combinatorial problem of matching the vertices in G and H, which is not constrained by relative spatial locations. The position of the same glomerulus in two consecutive slides, however, should be similar relative to the surrounding tissue, assuming that the tissues in both slides have been approximately rigidly registered.

Based on this observation, the global inexact graph matching problem of $\mathcal{G}$ to $\mathcal{H}$ is transformed into $|G|$ sub-graph assignment problems. In the following, the general matching strategy is first developed, then the assignment energy used to match two glomeruli is presented.

From Global to Local Matching: Let G and H be embedded in $\mathbb{R}^2$. We define the set of edges of both graphs such that $E_G = \{(x, y) \in G^2, D(x, y) \leq d_{sb}\}$ and $E_H = \{(x, y) \in H^2, D(x, y) \leq d_{sb}\}$ with $d_{sb} \in \mathbb{R}$ and $D : \mathrm{R} \times \mathrm{R} \rightarrow \mathrm{R}$ a function returning the Euclidean distance between two points (vertices).

Instead of finding a global matching (Equation (1)), that could lead to the matching of glomeruli far from each other in consecutive WSIs, the problem is reduced to $|G|$ sub-problems defined by finding for each vertex $g \in G$, its best match $\hat{h} \in H$, among all vertices of H that are close to g, such that

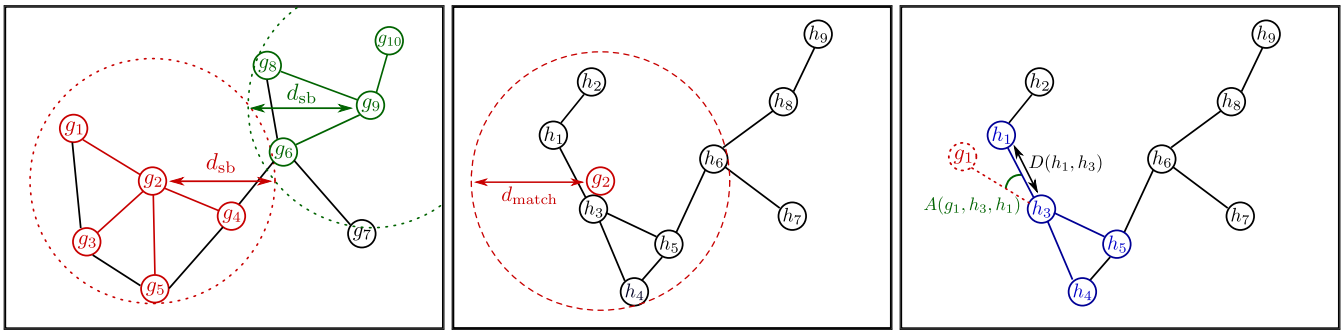
$$\hat{h} = \arg\min_{h \in \mathcal{N}_H^{d_{\text{match}}}(g)} E_{\text{match}}(g, h), \tag{2}$$

where $\mathcal{N}_H^{d_{\text{match}}}(g) = \{h \in H, D(h, g) \leq d_{\text{match}}\}$ and $E_{\text{match}}(g, h)$ is the matching energy of h to g that will be defined in the next subsection. An illustration of $\mathcal{N}_H^{d_{\text{match}}}(g)$ is presented in Fig. 3.

Since a glomerulus of H can only be matched to one glomerulus of G, the matching with the lowest energy E_{match} is retained for each g.

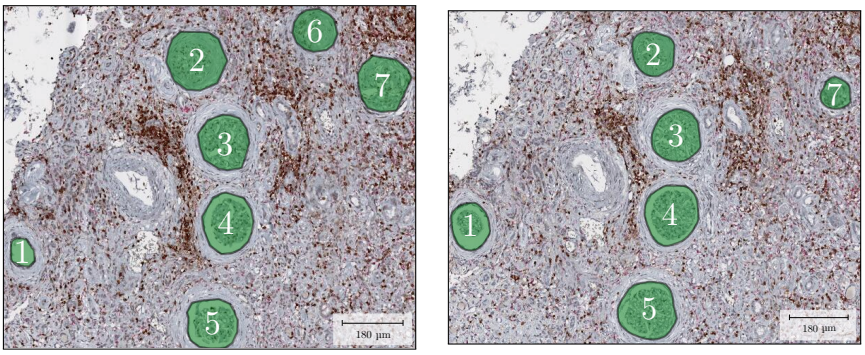
Assignment of Glomeruli Neighbourhood: It can be observed in WSIs that although the shape and size of a glomerulus slice may significantly vary between consecutive slides, its position relative to neighbouring glomeruli is relatively constant (see Fig. 4). To constrain the matching strategy with this observation, the matching energy of two glomeruli slices g and h is defined to be the minimal assignment energy of their respective neighbourhoods. More formally, let $\mathcal{G}_i^{\text{sb}} = (G_i^{\text{sb}}, E_{G_i}^{\text{sb}})$ be a subgraph centered on $g_i \in G$ such that

$$G_i^{\text{sb}} = \{g_k \in G, (g_i, g_k) \in E_G\} \cup g_i,$$
$$E_{G_i}^{\text{sb}} = \{(x, g_i) \in E_G,\ x \in G\}.$$



(a) (b) (c)

Fig. 3: Illustration of the glomeruli matching steps of two graphs $\mathcal{G}$ (a) and $\mathcal{H}$ (b). (a) Example of two subgraphs $\mathcal{G}_2^{\text{sb}}$ in red and $\mathcal{G}_9^{\text{sb}}$ in green of $\mathcal{G}$. (b) The set of vertices of H that can be matched to g_2: $\mathcal{N}_H^{d_{\text{match}}}(g_2) = \{h_1, h_2, h_3, h_4, h_5, h_6\}$. (c) The energy $E_{\text{nb}}(g_2, g_1, h_3, h_1)$ is computed based on the angle $A(g_1, h_3, h_1)$ and the distances $D(h_1, h_3)$ and $D(g_1, g_2)$. In the example of matching $\mathcal{G}_2^{\text{sb}}$ to $\mathcal{H}_3^{\text{sb}}$, the matching energy for $N = 3$ would be $E_{\text{match}}(g_2, h_3) = E_{\text{nb}}^{g_1 h_1} + E_{\text{nb}}^{g_5 h_4} + E_{\text{nb}}^{g_4 h_5}$.



(a) (b)

Fig. 4: Illustration of glomeruli matching on two consecutive WSIs. Consecutive glomerulus cuts may present different shapes and sizes (cf. glomeruli number 1 and 7) but their position in the tissue relative to other glomeruli is similar.

Fig. 4 presents examples of such sub-graphs.

Let $\tilde{G}_i^{\text{sb}} = G_i^{\text{sb}} \setminus g_i$ be the set of vertices connected to g_i. The assignment of $\tilde{G}_i^{\text{sb}}$ to $\tilde{H}_j^{\text{sb}}$ is define as follows:

$$E_{\text{match}}(g_i, h_j) = \min_{f \in F^{ij}} \sum_{i=1}^{N} RF_i \left(E_{\text{nb}}^{g_i h_j}(g, f(g)),\ g \in \tilde{G}_i^{\text{sb}} \right), \quad (3)$$

where F^{ij} is the set of all possible mappings of vertex of $\tilde{G}_i^{\text{sb}}$ to vertex of $\tilde{H}_j^{\text{sb}}$, and RF_i is the rank filter of order i such that $RF_1\left((a_i)_{i\in[1,n]}\right)$ is the minimum and RF_n is the maximum. The term $E_{\text{nb}}^{g_i h_j}$ is the energy of assigning two vertices, such that

$$E_{\text{nb}}^{g_i h_j}(g, h) = \frac{1}{90} A(g, g_i, h) + \frac{|D(g_i, g) - D(h_j, h)|}{D(g_i, g)}, \quad (4)$$

where $A(g, g_i, h) \in [0, 180]$ is the angle $\widehat{g g_i h}$ in degrees (see Fig. 3).

As such, the energy $E_{\text{match}}(g_i, h_j)$ is the sum of the N neighbour associations with the lowest E_{nb} with $N \leq |\tilde{G}_i^{\text{sb}}|$. The parameter N allows flexibility in the neighbourhood pattern matching, which is necessary as some neighbours can appear or disappear between two slides. The higher N the less flexibility allowed. This matching strategy is performed bidirectionally, i.e. from G to H and H to G, to increase its robustness. The matchings that are consistent between the two are kept to form the set of matched glomeruli M.

B. Parameter values

The proposed algorithm has two parameters: the maximum distance defining where a match can be found, d_{match}, and the number of neighbour associations to compute the assignment energy, N; and a hidden parameter d_{sb}, which is the distance defining the subgraphs.

In practice, d_{sb} is defined based on the glomeruli distribution and the number of associations N required to compute the assignment energy. As the assignment energy is defined based on N associations, most glomeruli should have at least N neighbours. In practice d_{sb} is defined such that most of the glomeruli in the image have at least $N+1$ neighbours to take into account the appearance and disappearance of glomeruli between consecutive slides.

The robustness of this algorithm was experimentally assessed on synthetic data (see supplementary materials).

IV. Automatic Analysis Pipeline

This section presents each step of the proposed analysis pipeline. Section IV-A describes the dataset used in this work.

TABLE I: Number of glomeruli per patient and stain. The numbers in brackets for patient 2 are the number of glomeruli before removing those situated in areas affected by staining irregularities, please refer to the text for more information.

Patient	CD3-CD68	CD3-CD163	CD3-CD206	CD3-MS4A4A
1	445	482	480	470
2	185 (271)	173 (255)	180 (267)	176 (253)
3	135	128	130	122
4	285	254	274	244
Total	1050	1037	1064	1012

The following sections present each step of the pipeline, namely the glomeruli and cell segmentation (Section IV-B), the registration (Section IV-C) and the feature extraction (Section IV-D).

A. Dataset

Tissue samples were collected from four patients who underwent allograft nephrectomy for various reasons. Each paraffin-embedded sample was cut into four consecutive $3\,\mu\mathrm{m}$ thick sections, each being stained with one of the following combination of immunohistochemistry markers using an automated staining instrument (Ventana Benchmark Ultra): CD3-CD68 (T cells & macrophage lineage marker), CD3-CD163 (T cells & M2-like macrophages), CD3-CD206 (T cells & M2-like macrophages), or CD3-MS4A4A (T cells & M2-like macrophages). The 4 M2-like macrophage stainings detect different subsets of M2 macrophages polarised along the large spectrum of alternatively activated ("M2-like") macrophages. Whole slide images were acquired using an Aperio AT2 scanner at $40\times$ magnification (a resolution of $0.253\,\mu\mathrm{m/pixel}$). All the healthy and sclerotic glomeruli in each WSI were annotated by outlining them using Cytomine [14] and validated by pathology experts. The number of glomeruli for each patient and in each staining is summarised in Table I. For technical reasons (most likely due to uneven tissue fixation), staining artefacts occurred in patient 2, that resulted in the need for manual removal of some areas. As the affected tissue was removed from the evaluation, Table I reports both the number of glomeruli including the ignored tissue (in parentheses) and the final corrected results. The WSIs and annotations of patient 1 are shown in Fig. 5 and larger scale crops in Fig. 6.

To validate the matching algorithm[2], approximately 270 glomeruli were manually associated with each other between the four slides of patient 1 (including 220 that exist within all four slides), and approximately 185 glomeruli in patient 2 (including 169 that exist within all four slides).

B. Structure and Cell Segmentation

In order to compute high-level quantitative features describing the kidney tissue (shape, distances, correlations etc.), we first need to analyse the tissue composition by detecting the units of interest on each WSI. In this article we focus on the glomeruli, which are important functional units of the kidney and our landmark for the fusion of information, and the cells highlighted by the stainings in each WSI. We propose a segmentation strategy for both that we refer to as glomeruli segmentation and cell segmentation.

1) Glomeruli Segmentation: Two approaches can be taken to segment the glomeruli slices in all stainings: develop a segmentation model for each staining [15]–[21], or a stain invariant/multi-stain segmentation model [22].

Computer vision approaches such as perceptual organisation [15], histogram of gradients [16], colour profiles [16], local binary patterns [17], [23], and combinations of approaches [18] integrate background knowledge into the task. Nevertheless, there is no general consensus on the type of features to extract and so data driven approaches have gained in popularity. Most recently, deep learning approaches [19]–[21] have become the de-facto standard for segmentation due to their state-of-the-art performance, however, being data driven they require a large amount of training data. To overcome this, pretrained networks such as GoogleNet and AlexNet can be used [19].

The proposed matching framework is agnostic to the segmentation algorithm used. In the demonstrated application, segmentation is performed using a U-Net [24] as it has been proven to be successful in biomedical imaging [25], in particular in glomeruli detection [20].

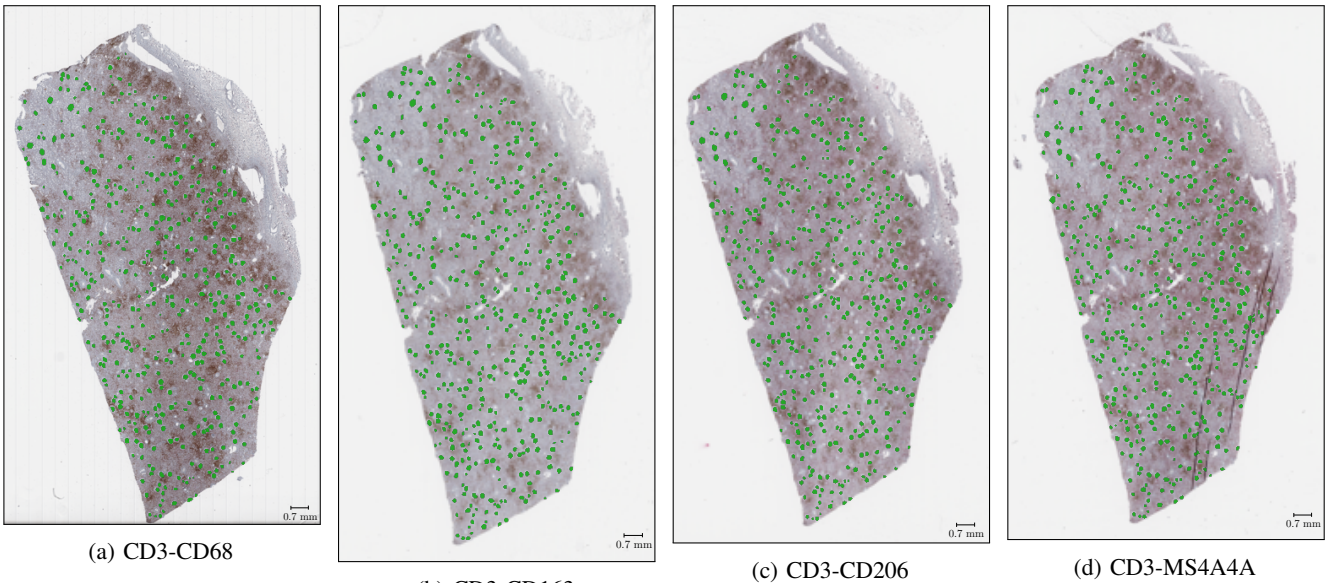
Glomeruli segmentation is framed as a two classes problem: *glomeruli* and *tissue*. The slide background (non-tissue) is manually removed from consideration. The input to the network are patches centred on a glomerulus, and those that do not contain a glomerulus, randomly sampled.

The U-Net was implemented as described in the original article [24] using cross entropy loss. The following parameter values were used: batch size of 8, learning rate of 0.0001, 60 epochs, and the network that achieves the lowest validation loss is kept. The input patch size is 508×508 pixels, which is sufficient to contain a glomerulus at a resolution of $0.506\,\mu\mathrm{m/pixel}$.

The following data augmentation is performed with an independent probability of 0.5:
elastic deformation: using the parameters $\sigma = 10$, $\alpha = 100$;
affine: random rotation sampled from the interval $[0°, 180°]$, random shift sampled from $[-205, 205]$ pixels, random magnification sampled from $[0.8, 1.2]$, and horizontal/vertical flip;
noise: additive Gaussian noise with $\sigma \in [0, 2.55]$;
blur: Gaussian filter with $\sigma \in [0, 1]$;
brightness enhance with a factor sampled from $[0.9, 1.1]$;
colour enhance with a factor sampled from $[0.9, 1.1]$;
contrast enhance with a factor sampled from $[0.9, 1.1]$.
These values were chosen to produce realistic images. All samples are standardised to $[0, 1]$ and normalised by the mean and standard deviation of the training set.

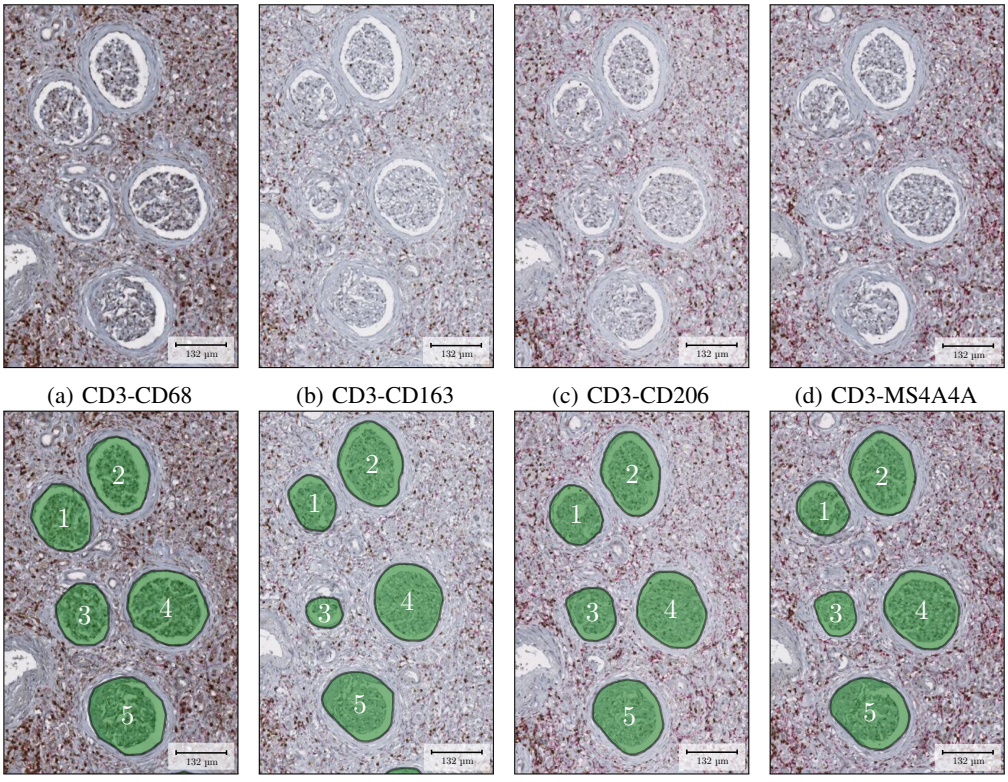
Because of the relatively small amount of training data in the experiements presented in Section V, and the large variance observed between the stainings and characteristics of each patient, the U-Net used upsampling instead of transposed convolution to reduce the number of learnable parameters. Furthermore, the output of the U-Net was postprocessed by removing the smallest connected components and closing small holes.

[2]The code will be made available upon acceptance.



(a) CD3-CD68 (b) CD3-CD163 (c) CD3-CD206 (d) CD3-MS4A4A

Fig. 5: Fully annotated consecutive kidney nephrectomy WSIs used in this study (patient 1, see Table I). Each green disk is an individual glomerulus.



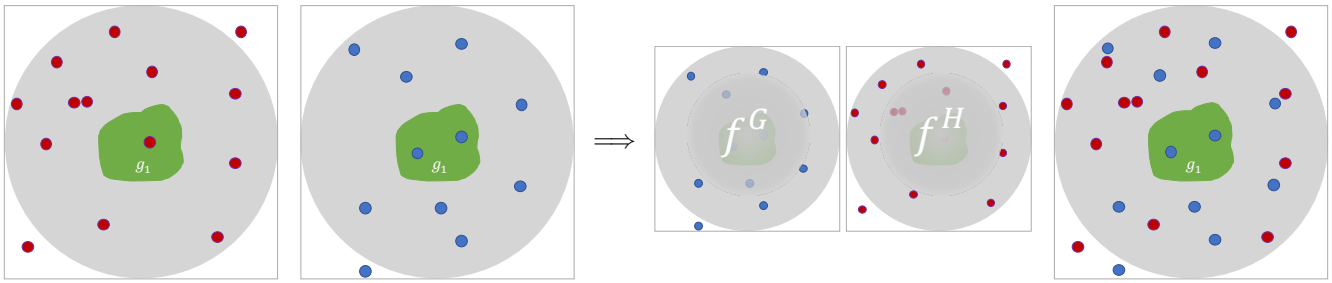
(a) CD3-CD68 (b) CD3-CD163 (c) CD3-CD206 (d) CD3-MS4A4A

(e) CD3-CD68 GT (f) CD3-CD163 GT (g) CD3-CD206 GT (h) CD3-MS4A4A GT

Fig. 6: Subsamples of the four consecutive kidney nephrectomy WSIs used in this study (Patient 1, see Table I).

2) Cell Segmentation: Our dataset is composed of 4 double-stained consecutive WSIs for each patient, each staining highlighting different cell types (see Sec. IV-A). In total, 5 different cell types are highlighted: T cells (CD3) and 4 different types of M2-like macrophages (CD68, CD163, CD206 and MS4A4A). The goal of this step is to segment each cell type resulting in 5 binary images that will be used to compute features.

The image resulting from the digitisation of a WSI is a mixture of the signals from two stains (e.g. CD3 and CD68) and the counter-stain (e.g. haematoxylin). The classic method to unmix the stains from an RGB image was proposed by Ruifrok et al. and called colour deconvolution [26]. This method transforms the RGB channels of the WSI into optical



(a) Stain G (b) Stain H $\mathbf{F} = [f^G, f^H]$ (c) Multi-WSI features f^{GH} (d) Cross-WSI features

Fig. 7: After matching, the same glomerulus (in green) and surrounding tissue in two different stain modalities (a) and (b), the cells of different types are segmented (in red and blue). Two types of features can therefore be computed: features depending on each single WSI concatenated to obtain features relevant to the same glomerulus (c); and features that combine information from both stain modalities (d).

densities of each staining that are linearly related to their concentrations in the tissue. Once each slide is unmixed, a simple thresholding of the channels of interest is enough to segment the structures targeted by the main stain.

Colour deconvolution requires a predetermined stain vector for each staining that represent the proportion of optical densities of this staining in each RGB channel. In this work, the stain vectors for each staining were measured from the dataset, however, unsupervised methods have been proposed for situations in which stain vector measurement is not an option. These methods are based on singular value decomposition [27], blind deconvolution [28], dictionary learning [29], multilayer perceptron networks [30] or non-negative matrix factorisation [31].

C. Registration

Using the matching algorithm presented in Section III, bi-directional matching (i.e. WSI G to H, and H to G) is performed on three image pairs (CD3-CD68 - CD3-CD163, CD3-CD163 - CD3-CD206, and CD3-CD206 - CD3-MS4A4A). This results in glomeruli associations between all four consecutive slides.

D. Feature Extraction

Once segmentation and matching across WSIs is complete, the following two types of features that integrate information derived from different stains in the same glomerulus neighbourhood can be extracted from the corresponding segmentations, as illustrated in Fig. 7.

Multi-WSI features derived from multiple single WSIs, for example mean M0 macrophage (CD68) or M2 macrophage (CD163) densities inside each glomerulus.

Cross-WSI features, that combine information derived from multiple slides, for example, the mean distance from M0 macrophages (CD68) to a subtype of M2 macrophages (CD163).

V. Results

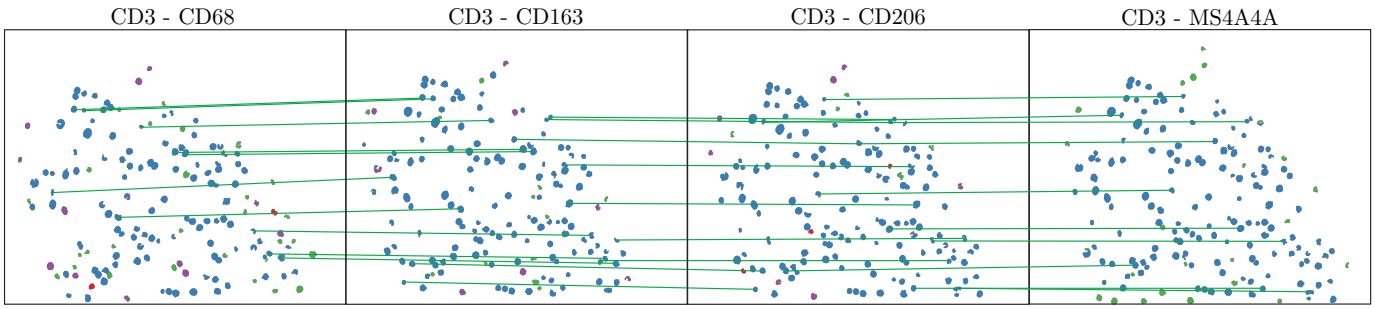
In this section, the matching and segmentation are first validated separately and then the results of the full pipeline are presented.

The following metrics are used to evaluate matching performance: Sensitivity ($S = \frac{\text{TP}}{\text{TP+FN}}$), Precision ($P = \frac{\text{TP}}{\text{TP+FP}}$), and Specificity ($SP = \frac{\text{TN}}{\text{TN+FP}}$) and Negative Predictive Value (NPV $= \frac{\text{TN}}{\text{TN+FN}}$) to account for the possibility of false positive—a centroid incorrectly associated to another—and true negative associations—unpaired centroids not associated with another correctly. The values of TP, FP, and FN were measured in terms of associations, such that a TP is a correct association, an FP is an incorrect association, and an FN is when no association is made incorrectly.

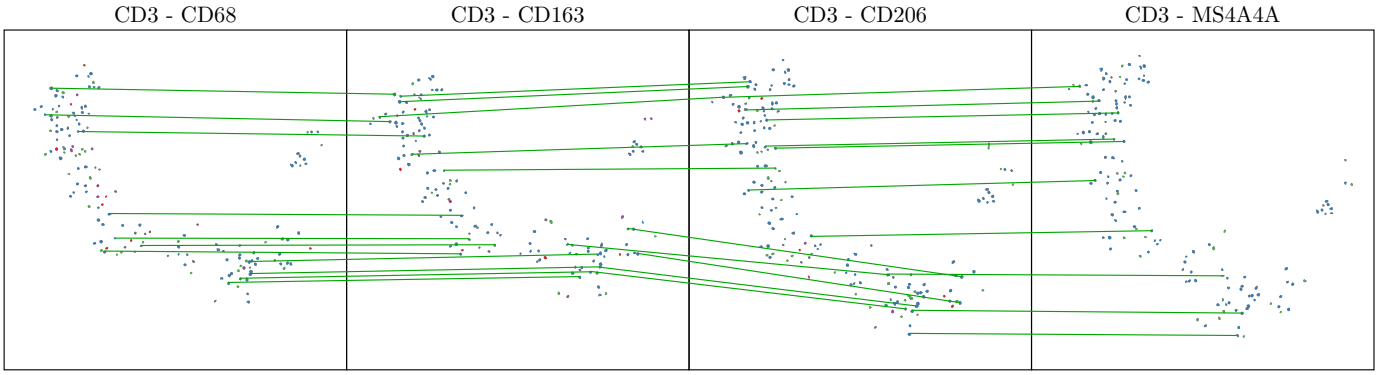
A. Validation on Glomeruli Ground-Truth Segmentation

The matching algorithm was first validated independently of the pipeline, more specifically of possible segmentation errors, by matching the glomeruli of the nephrectomy dataset obtained by manual segmentation (Fig. 5). The results of these experiments are shown in Table II and Fig. 8.

The matching algorithm has little trouble finding correct associations in all but a very few cases as shown by the very high sensitivity and precision scores. Moreover the false associations are usually understandable as they concern glomeruli that are close and for whom the ground truth matching was problematic even for experts (see Fig. 9). Most of the errors of the matching algorithm concern false detections as shown by the NPV score. The variation of the NPV is high as it is computed on a small number of samples (one more false negative association will decrease the NPV of a few tens of percent). When a clear association can not be found, the algorithm tends not to match the glomerulus, which is a desired behaviour for the discussed applications. This ensures that the associations made are reliable and will not bias further statistics that could be built upon them.



(a) Patient 1



(b) Patient 2

Fig. 8: A random subset of the glomeruli matchings between the four WSIs, in which TN matched glomeruli are in green, TP in blue (the green line represents the correct associations between WSI), FN in purple, and FP in red.

TABLE II: Matching performance on ground-truth (GT) *vs.* on segmentation (Segm) with $d_{\text{match}} = 300$ and $N = 4$.

Stain Pair	S		SP		P		NPV	
	GT	Segm	GT	Segm	GT	Segm	GT	Segm
CD3-CD68 — CD3-CD163	93%	86%	90%	89%	99%	97%	43%	58%
CD3-CD163 — CD3-CD206	98%	94%	100%	100%	100%	100%	75%	55%
CD3-CD206 — CD3-MS4A4A	96%	93%	100%	57%	100%	98%	43%	27%

Patient 1

Stain Pair	S		SP		P		NPV	
	GT	Segm	GT	Segm	GT	Segm	GT	Segm
CD3-CD68 — CD3-CD163	95%	92%	94%	74%	99%	90%	80%	76%
CD3-CD163 — CD3-CD206	98%	93%	100%	67%	100%	94%	84%	57%
CD3-CD206 — CD3-MS4A4A	94%	95%	95%	85%	99%	95%	70%	82%

Patient 2

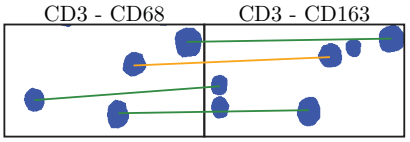


Fig. 9: False positive matching occurring in patient 1 between CD3-CD68 and CD3-CD163 (in orange) when applying the matching algorithm to the ground-truth segmentations. The correctness of this association is debatable even for experts

B. Glomeruli Segmentation and Matching

The stainings in this study present similar visual characteristics, see Fig. 6, which lends to training one 'multi-stain' U-Net by combining the training sets of each stain and applying the same network to all stains. To better utilise the limited amount of data, one network was trained for each patient in a leave-one-out fashion, such that the segmentor for patient 1 was trained using data from patients 2, 3, and 4; and patient 2 was trained using the data derived from patient 1, 3, and 4. The training set comprised patches centred on all glomeruli from the training patients and seven times the number of tissue patches (to account for the variance observed in non-glomeruli tissue), 20% of this data was reserved for validation.

The segmentation performance of this approach is described in Table III.

The centroids of each detected glomerulus were then extracted to form the sets G and H, which are the input to the matching algorithm. Pairwise matching is then performed on each consecutive image to determine the associations between

all WSIs. These results are presented in Table II. The sensitivity and precision are still very high compared with the ground-truth baseline which demonstrates the algorithm's detection robustness. The specificity sometimes drops significantly in some staining associations. The high specificity drops without a significant sensitivity or precision drop and is explained by the very small number of negative matchings is these stainings. Each single false positive match yields a large specificity drop. This behaviour is not problematic in a global scale as the number of false positive match remains very low.

C. Multi-WSI Analysis

At this stage of the pipeline, it is possible to register each matched glomerulus and its surrounding, allowing the superimposition of the segmentations from each consecutive WSI. Fig. 10 shows the result of this for a glomerulus of Patient 1. With this, both Multi-WSI and Cross-WSI features can be computed and used for diagnosis and research purpose.

VI. Conclusions

In summary, this article has presented a novel framework for the study of tissue micro-environment of renal glomeruli across multiple WSIs that allows their comprehensive evaluation without technically challenging multiplexing, by integrating multiple staining modalities in consecutive tissue sections. The framework involves segmenting glomeruli and cells in each WSI, then matching them across the WSIs to integrate the information contained within each. The result of this can then be used to perform analyses on the glomeruli and surrounding tissue.

The proposed framework is generic and independent of the presented use-cases. It can be used for the analysis of the micro-environment surrounding other large structures, under the assumption that such structures are large enough to exist across multiple WSIs and can be segmented (either manually or automatically). Furthermore, it is independent of the segmentation algorithm used and can therefore be applied to a variety of biomedical research questions beyond transplantation medicine, for example immuno-oncology and other scientific fields working with biopsy samples.

In the future, this approach could support the diagnosis of renal grafts by time-efficient quantification and evaluation of glomeruli (e.g. fibrosis) and precise number and localisation of infiltrating leukocytes (e.g. glomerulitis according to the internationally used BANFF classification for renal grafts [32]). Counting glomeruli with the described methods could also be performed for 3D reconstruction (research purposes) in consecutive tissue slides and thus enable an estimation of glomeruli numbers in the whole kidney: reduced renal allograft survival [33], hypertension and the risk of chronic kidney disease [34] are associated with low glomeruli number.

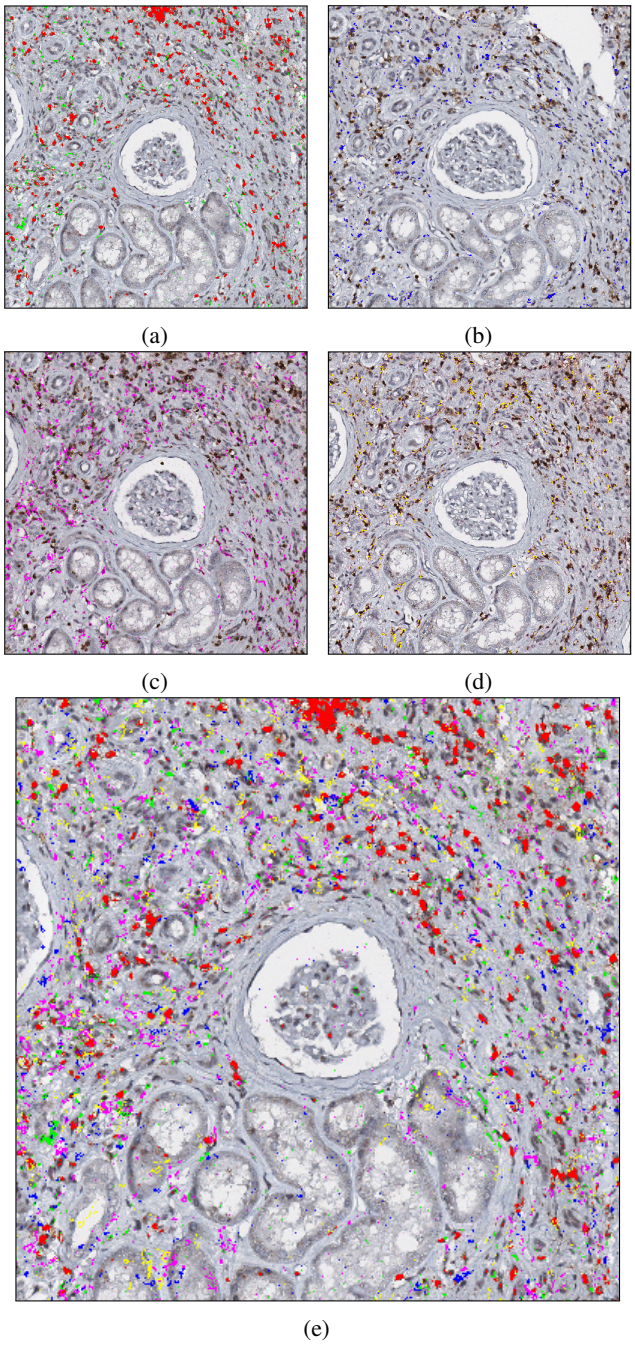


Fig. 10: Illustration of the segmentation of different cell types around the same glomerulus through 4 consecutive slides. (a) T cells (red) and M0 macrophages (green) in CD3-CD38 WSI, (b) M2 macrophages (blue) in CD3-CD163 WSI, (c) M2 macrophages (pink) in CD3-CD206 WSI, (d) M2 macrophages (yellow) in CD3-MS4A4A WSI, (e) superimposition of all cell types in CD3-CD68 WSI. Note that each subtype of macrophage is present in only one of the consecutive slides.

Acknowledgement

The authors thank Nvidia Corporation for donating a Quadro P6000 GPU, the *Centre de Calcul de l'Université de Strasbourg* for access to the GPUs used for this research, and Nicole Krönke for her excellent technical assistance in data preparation.

References

[1] Y. Liu, "Cellular and molecular mechanisms of renal fibrosis," *Nat. Rev. Nephrol.*, vol. 7, no. 12, pp. 684–696, 2011.
[2] X. Li and S. Zhuang, "Recent advances in renal interstitial fibrosis and tubular atrophy after kidney transplantation," *Fibrogenesis & Tissue Repair*, vol. 7, p. 15, 2014.

TABLE III: Segmentation performance of the detection algorithm based on pixels (average F_1 score of five repetitions, with standard deviations in parentheses).

Patient	CD3-CD68	CD3-CD163	CD3-CD206	CD3-MS4A4A	Overall
1	0.803 (0.010)	0.801 (0.020)	0.822 (0.015)	0.818 (0.014)	0.811 (0.014)
2	0.860 (0.003)	0.863 (0.005)	0.859 (0.015)	0.865 (0.011)	0.862 (0.003)

[3] P. Tang *et al.*, "Macrophages: versatile players in renal inflammation and fibrosis," *Nat. Rev. Nephrol.*, vol. 15, no. 3, pp. 144–158, 2019.

[4] J. Bräsen *et al.*, "Macrophage density in early surveillance biopsies predicts future renal transplant function," *Kidney International*, vol. 92, no. 2, pp. 479–489, 2017.

[5] J. Casper *et al.*, "Increased urinary tract infection rate and altered medullary macrophage polarization marker expression in renal transplant recipients receiving loop diuretic therapy," *Kidney International*, vol. 94, no. 5, pp. 993–1001, 2018.

[6] M. Angelova *et al.*, "Evolution of metastases in space and time under immune selection," *Cell*, vol. 175, no. 3, pp. 751–765, 2018.

[7] D. Johnson *et al.*, "A case report of clonal EBV-like memory CD4$^+$ T cell activation in fatal checkpoint inhibitor-induced encephalitis," *Nature Medicine*, vol. 25, no. 8, pp. 1243–1250, 2019.

[8] A. Can *et al.*, "Multi-modal imaging of histological tissue sections," in *IEEE ISBI*, 2008, pp. 288–291.

[9] K. Mosaliganti *et al.*, "Registration and 3D visualization of large microscopy images," in *Medical imaging 2006: Image processing*, vol. 6144, 2006, p. 61442V.

[10] D. Mueller, D. Vossen, and B. Hulsken, "Real-time deformable registration of multi-modal whole slides for digital pathology," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7–8, pp. 542–556, 2011.

[11] L. Cooper *et al.*, "Feature-based registration of histopathology images with different stains: An application for computerized follicular lymphoma prognosis," *Computer methods and programs in biomedicine*, vol. 96, no. 3, pp. 182–192, 2009.

[12] Y. Song *et al.*, "Unsupervised content classification based nonrigid registration of differently stained histology images," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 1, pp. 96–108, 2014.

[13] R. Raveaux, J.-C. Burie, and J.-M. Ogier, "A graph matching method and a graph matching distance based on subgraph assignments," *Pattern Recognition Letters*, vol. 31, no. 5, pp. 394–406, 2010.

[14] R. Marée *et al.*, "Collaborative analysis of multi-gigapixel imaging data using cytomine," *Bioinformatics*, vol. 32, no. 9, pp. 1395–1401, 2016.

[15] S. Samsi *et al.*, "Glomeruli segmentation in H&E stained tissue using perceptual organization," in *IEEE SPSMB*, 2012, pp. 1–5.

[16] M. Gadermayr *et al.*, "Do we need large annotated training data for detection applications in biomedical imaging? a case study in renal glomeruli detection," in *MLMI*, 2016, pp. 18–26.

[17] ——, "Domain adaptive classification for compensating variability in histopathological whole slide images," in *ICIAR*, 2016, pp. 616–622.

[18] R. Marée *et al.*, "An approach for detection of glomeruli in multisite digital pathology," in *IEEE ISBI*, 2016, pp. 1033–1036.

[19] J. Gallego *et al.*, "Glomerulus classification and detection based on convolutional neural networks," *J. Imaging*, vol. 4, no. 1, p. 20, 2018.

[20] T. de Bel *et al.*, "Automatic segmentation of histopathological slides of renal tissue using deep learning," in *Medical Imaging 2018: Digital Pathology*, vol. 10581, 2018, p. 1058112.

[21] S. Kannan *et al.*, "Segmentation of glomeruli within trichrome images using deep learning," *Kidney Int. Rep.*, vol. 4, no. 7, pp. 955–962, 2019.

[22] T. Lampert *et al.*, "Strategies for training stain invariant CNNs," in *IEEE ISBI*, 2019, pp. 905–909.

[23] O. Simon *et al.*, "Multi-radial LBP features as a tool for rapid glomerular detection and assessment in whole slide histopathology images," *Scientific Report*, vol. 8, p. 2032, 2018.

[24] O. Ronneberger *et al.*, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[25] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[26] A. Ruifrok and D. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, 2001.

[27] M. Macenko *et al.*, "A method for normalizing histology slides for quantitative analysis," in *IEEE ISBI*, 2009, pp. 1107–1110.

[28] M. Gavrilovic *et al.*, "Blind color decomposition of histological images," *IEEE Trans. Med. Imaging*, vol. 32, no. 6, pp. 983–994, 2013.

[29] A. Vahadane *et al.*, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. Med. Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.

[30] C. Wemmert *et al.*, "Stain unmixing in brightfield multiplexed immunohistochemistry," in *IEEE ICIP*, 2013, pp. 1125–1129.

[31] A. Rabinovich *et al.*, "Unsupervised color decomposition of histologically stained tissue samples," in *NIPS*, 2004, pp. 667–674.

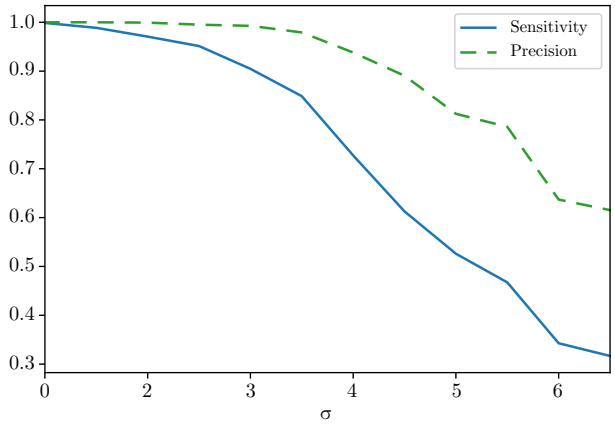
[32] M. Haas *et al.*, "The Banff 2017 kidney meeting report: Revised diagnostic criteria for chronic active t cell-mediated rejection, antibody-mediated rejection, and prospects for integrative endpoints for next-generation clinical trials," *American Journal of Transplantation*, vol. 18, no. 2, pp. 293–307, 2018.

[33] H. Mackenzie and B. Brenner, "Antigen-independent determinants of late renal allograft outcome: the role of renal mass," *Current Opinion in Nephrology and Hypertension*, vol. 5, no. 4, pp. 289–296, 1996.

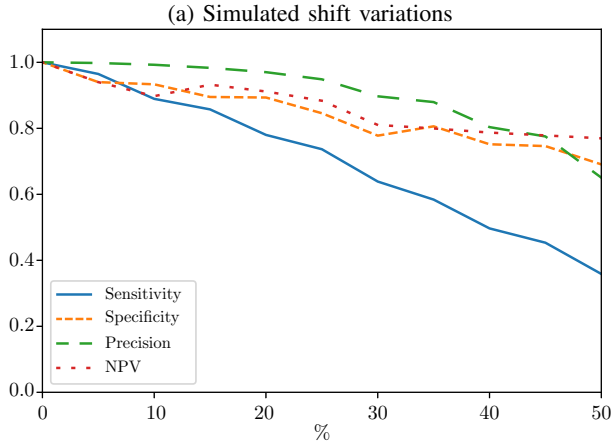
[34] N. Tsuboi *et al.*, "Clinicopathological assessment of the nephron number," *Clinical Kidney Journal*, vol. 7, no. 2, pp. 107–114, 2014.

SUPPLEMENTARY MATERIAL

A. Glomeruli matching synthetic validation



(a) Simulated shift variations



(b) Simulated unpaired centroids

Fig. 11: The results of the matching algorithm on synthetic data with $d_{\text{match}} = 80$ and $N = 3$.

To evaluate the performance of the proposed algorithm to variations in the data, a simulated dataset was created. Fifty 300×300 image pairs were generated, each pair representing two consecutive slides. For each image in a pair, 30 centroids were generated located at the same position in both images. The following two variations to the centroids were analysed.

Shift — For each second image in a pair, the x and y position of each centroid was shifted by values drawn independently from a Gaussian distribution with $\mu = 0$ and $\sigma \in \{0 \dots 11\}$ (that is 0 to 3.6% of the image size in each dimension).

Unpaired — Spurious unpaired centroids were randomly added to each image in a pair. The number of centroids added to each image ranged from 0% to 50% of the initial number of centroids in the image. An addition of 50% means that both images in a pair contains 45 centroids but only 30 should be matched.

The Shift experiment was designed to evaluate the normal spatial variations of glomeruli slices in consecutive WSI whereas the Unpaired experiment evaluates the algorithm's behavior to glomeruli appearance and disappearance between slides, and to errors during glomeruli segmentation.

Sensitivity ($S = \frac{\text{TP}}{\text{TP+FN}}$), and precision ($P = \frac{\text{TP}}{\text{TP+FP}}$) were measured during the Shift experiment. The values of TP, FP, and FN were measured in terms of associations, such that a TP is a correct association, an FP is an incorrect association, and an FN is when no association is made incorrectly. During the Unpaired experiment, Specificity ($SP = \frac{\text{TN}}{\text{TN+FP}}$) and Negative Predictive Value (NPV $= \frac{\text{TN}}{\text{TN+FN}}$) were also measured to account for the possibility of false positive—a centroid incorrectly associated to another—and true negative associations—unpaired centroids not associated with another correctly.

The average measure (over the 50 repetitions of each setup) for each experiment is presented in Fig. 11. These experiments show that the proposed algorithm is robust to shift and unpaired centroids. It is interesting to see that precision remains high with the increase of each parameter (shift and the number of added centroids) even though the specificity decreases more quickly. This means that the algorithm tends to avoid falsely associating glomeruli, which is a highly desirable behavior when the goal is to extract statistical measures based on quantitative data extracted from image processing algorithms.