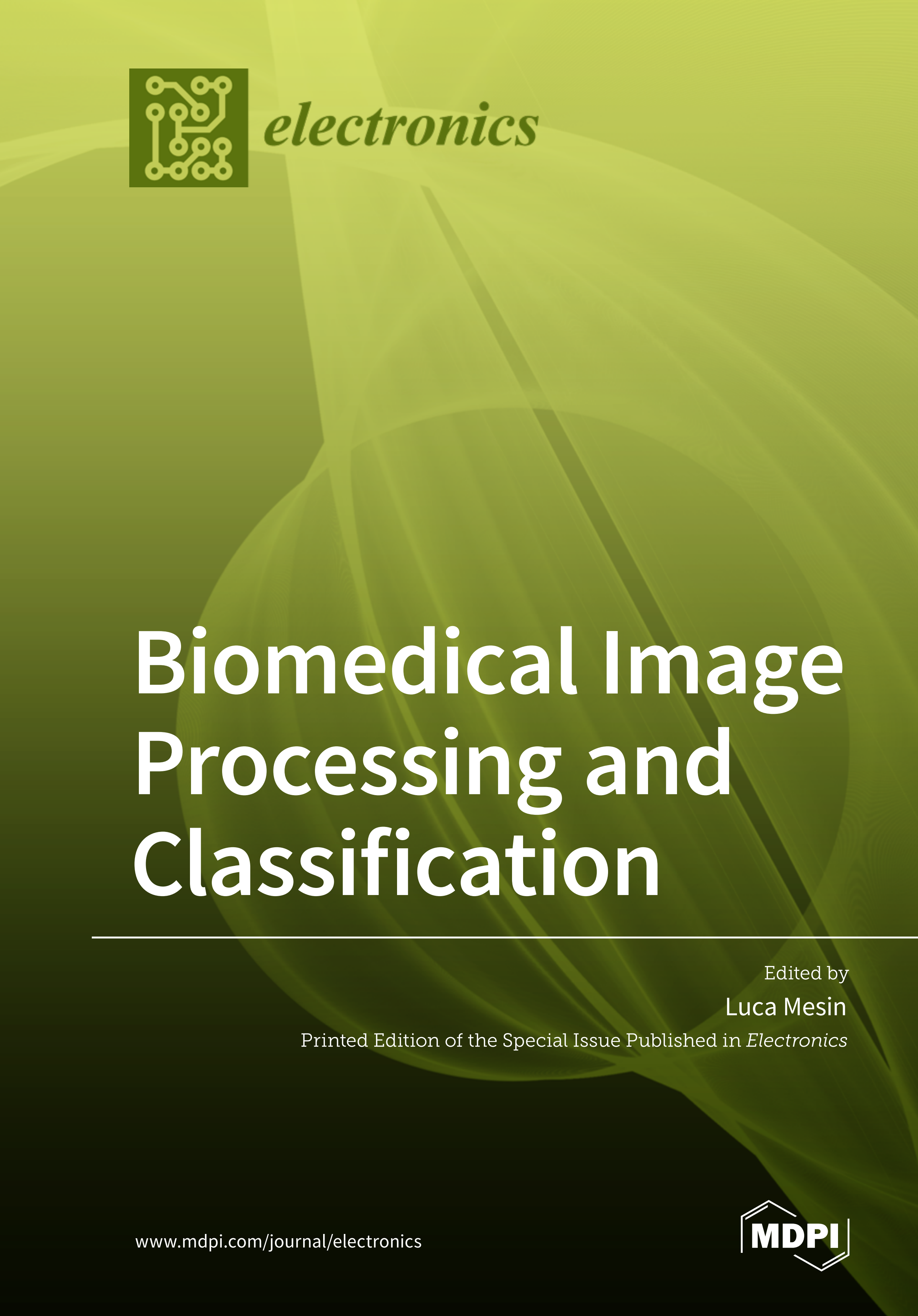
Biomedical Image Processing and Classification

Original
Biomedical Image Processing and Classification / Mesin, Luca. - STAMPA. - (2021). [10.3390/books978-3-0365-0347-9]

Publisher:
MDPI

Published
DOI:10.3390/books978-3-0365-0347-9



electronics

Biomedical Image Processing and Classification

Edited by
Luca Mesin

Printed Edition of the Special Issue Published in *Electronics*

www.mdpi.com/journal/electronics

MDPI

Biomedical Image Processing and Classification

Biomedical Image Processing and Classification

Editor

Luca Mesin

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editor
Luca Mesin
Polytechnic University of Turin
Italy

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Electronics* (ISSN 2079-9292) (available at: https://www.mdpi.com/journal/electronics/special_issues/bioimg_classification).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-0346-2 (Hbk)
ISBN 978-3-0365-0347-9 (PDF)

Contents

About the Editor

Luca Mesin (Ph.D.) is currently employed at the Department of Electronics and Telecommunications, Polytechnic University of Turin, 10129 Turin, Italy. He graduated with a degree in Electronics Engineering in 1999, and he received a Ph.D. in Applied Mathematics in 2003 from Politecnico di Torino, Italy. His present position is that of Associate Professor in Biomedical Engineering and he is also the supervisor of the Mathematical Biology and Physiology group of the Department of Electronics and Telecommunications, Politecnico di Torino. His main research activities concern biomedical signal/image processing and classification, biophysical modeling and human–machine interactions.

electronics

Editorial

Biomedical Image Processing and Classification

Luca Mesin

Mathematical Biology and Physiology, Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy; luca.mesin@polito.it; Tel.: +39-0110904085

Received: 30 November 2020; Accepted: 14 December 2020; Published: 1 January 2021

---

1. Introduction

Biomedical image processing is an interdisciplinary field [1] that spreads its foundations throughout a variety of disciplines, including electronic engineering, computer science, physics, mathematics, physiology, and medicine. Several imaging techniques have been developed [2], providing many approaches to the study of the body, including X-rays for computed tomography, ultrasounds, magnetic resonance, radioactive pharmaceuticals used in nuclear medicine (for positron emission tomography and single-photon emission computed tomography), elastography, functional near-infrared spectroscopy, endoscopy, photoacoustic imaging, and thermography. Even bioelectric sensors, when using high-density systems sampling a two dimensional surface (e.g., in electroencephalography or electromyography [3]), can provide data that can be studied by image processing methods. Biomedical image processing is finding an increasing number of important applications, for example, to make image segmentation of an organ to study its internal structure and to support the diagnosis of a disease or the selection of a treatment [4].

Classification theory is another well developed field of research [5] connected to machine learning, which is an important branch of artificial intelligence. Different problems have been addressed, from the supervised identification of a map relating input features to a desired output, to the exploration of data by unsupervised learning (cluster analysis, data mining) or online training through experience. The estimation of informative features and their further processing (by feature generation) and selection (either by filtering or with approaches wrapped to the classifier) are important steps, both to improve classification performance (avoiding overfitting) and to investigate the information provided by candidate features to the output of interest. Excellent results have also been recently documented by deep learning approaches [6], in which optimal features are automatically extracted in deep layers on the basis of training examples and then used for classification.

When classification methods are associated with image processing, computer-aided diagnosis (CAD) systems can be developed, e.g., for the identification of diseased tissues [7] or a specific lesion or malformation [4]. These results indicate interesting future prospects in supporting the diagnosis of diseases [8].

2. This Special Issue

The present issue consists of six papers on a few topics in the wide range of research fields covered by biomedical image processing and classification.

In [9], the authors have proposed a CAD system for identification and assessment of glomeruli from kidney tissue slides. Their approach is based on deep learning, exploiting convolutional neural network (CNN) architectures tailored for the semantic segmentation task. The obtained results are promising, as also stated by expert pathologists. Moreover, the proposed system can easily be integrated into the existing pathologists' workflow thanks to an XML interface with Aperio ImageScope [10].

With the recent advances of techniques in digitalized scanning, tissue histopathology slides can be stored in the form of digital images [11]. In recent years, many efforts have been devoted to developing automated classification and segmentation techniques with the aim of improving accuracy and efficiency in digital pathology [12]. In kidney transplantations, pathologists evaluate the architecture of renal structures to assess the nephron status. An accurate evaluation of vascular and stromal injury is crucial for determining kidney acceptance, which is currently based on the pathologists' histological evaluations on renal biopsies in addition to clinical data. In this context, automated algorithms may offer crucial support to histopathological image analysis. An example is given in this Special Issue [13].

Although the performance of a machine learning algorithm depends on the amount of available data, few studies have explored the minimal amount of data required to train a CNN in medical deep learning or the possibility of having scarce annotations [14]. An innovative contribution is given in this Special Issue [15]. The paper explores the minimum number of patients required to train a U-Net that accurately segments the prostate on T2-weighted MRI images. A U-Net was trained on patient numbers that ranged from 8 to 320 and its performance was measured. The Dice score significantly increased from training sizes of 8 to 120 patients and then plateaued with minimal improvement after 160 cases. This study suggests that modest dataset sizes could be sufficient to segment other organs effectively as well.

The correlation between conjunctival pallor (on physical examinations) and anemia paved the way for new non-invasive methods for monitoring and identifying the potential risks of this important pathology. A critical research challenge for this task is represented by designing a reliable automated segmentation procedure for the eyelid conjunctiva. A graph partitioning segmentation approach is proposed in [16], exploiting normalized cuts for perceptual grouping, thereby introducing a bias towards spectrophotometry features of hemoglobin. The segmentation task has been further investigated by a subsequent work, proposing a deep-learning-based approach involving a deconvolutional neural network [17]. The overall pipeline for building a reliable estimator is composed of several smaller tasks having multiple research challenges [18,19]. For instance, starting from the digital image capturing phase, the process is affected by heterogeneous ambient lighting conditions and intrinsic color balancing techniques by the device [20].

An efficient framework for enhancing and segmenting brain MRIs to identify a tumor is discussed in [21]. The hybridized fuzzy clustering and distance regularized level set (DRLS) technique effectively extracted the region of interest (ROI) in the brain slices. For identifying the ROI, fuzzy clustering was employed by selecting the number of clusters k, validated using the silhouette metric. In post-processing, the ROI mining techniques, marker controlled watershed segmentation, seed region growing and DRLS were adopted to extract the anomalous section from the segmented objects [22,23]. Tumor volume computation and 3D-modeling of the clinical dataset abnormalities were performed using the physical spacing metadata available in the header of the DICOM images considered. This can help physicists locate the tumor and determine other information (e.g., size and shape) during initial diagnosis, and thereby the process of treating the tumor may be enhanced.

Finally, one paper in this Special Issue has addressed the problem of identifying the volume status of patients [24]. The method was developed within a long-standing research activity on the automated investigation of the pulsatility of the inferior vena cava (IVC) from ultrasound measurements. The clinical approach is based on the subjective choice of a fixed direction along which to investigate IVC pulsations. However, the vein may have a complicated shape and show respirophasic movements, which introduce uncertainties into the clinical evaluation. Two automated methods have been introduced to delineate the IVC edges along sections either transverse or longitudinal to the blood vessel [25–27]. Preliminary results have shown the importance of using these automated methods to obtain more repeatable, reliable, and accurate information on IVC pulsatility than when using subjective clinical methods [28–31]. In this Special Issue, the two views are used to extract features that, integrated by a classification algorithm, can result in improved performance in diagnosing the volemic status of patients [24].

3. Future Perspectives

The research fields of biomedical image processing and classification have reached high levels of insight. Their integration into CAD systems can greatly contribute to supporting medical doctors to refine their clinical picture. In the near future, further growth in contributions to this field is expected; for example, taking advantage of increasing digitalization, deep learning has the potential to provide efficient solutions to many medical problems.

However, the real challenge is to bring an increasing number of systems into the hands of doctors, so that they can be applied to patients. This requires leaving the laboratory, engineering the systems, certifying the products, and identifying the correct target market that can accommodate the new devices and allow adequate support for these activities. In order to speed up this innovation process, the collaboration between researchers, institutions, funders, and entrepreneurs is always more important. The ”do-it-all-yourself” approach only makes sense in a world of scarce external knowledge, but today knowledge is spread as it has never been before. Thus, in order to improve the wellness of the whole community [32], a dynamic environment in which new high-impact solutions can be created will be able to grow only if there is collaboration among organizations.

Funding: This research was carried out as part of the project “Method and apparatus to characterize non-invasive images containing venous blood vessels”, funded through the PoC Instrument initiative, implemented by LINKS, with the support of LIFFT, with funds from Campagnia di San Paolo.

Acknowledgments: I would like to thank all authors who contributed to this Special Issue, the reviewers for their help in refining the papers and the MDPI editorial board and staff for the opportunity to be guest-editor.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Deserno, T.M. *Biomedical Image Processing*; Springe: New York, NY, USA, 2011.
2. Maier, A.; Steidl, S.; Christlein, V.; Hornegger, J. *Medical Imaging Systems: An Introductory Guide, New York*; Springer: Berlin, Germany, 2018; Volume 11111.
3. Merletti, R.; Muceli, S. Tutorial. Surface EMG detection in space and time: Best practices. *J. Electromyogr. Kinesiol.* **2019**, *49*, 102363. [CrossRef] [PubMed]
4. Mesin, L.; Mokabberi, F.; Carlino, C.F. Automated Morphological Measurements of Brain Structures and Identification of Optimal Surgical Intervention for Chiari I Malformation. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3144–3153. [CrossRef] [PubMed]
5. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*; Academic Press: Cambridge, MA, USA, 2008.
6. Bevilacqua, V.; Brunetti, A.; Guerriero, A.; Trotta, G.F.; Telegrafo, M.; Moschetta, M. A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cogn. Syst. Res.* **2019**, *53*, 3–19. [CrossRef]
7. Tiwari, A.; Srivastava, S.; Pant, M. Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. *Pattern Recognit. Lett.* **2020**, *131*, 244–260. [CrossRef]
8. Yanase, J.; Triantaphyllou, E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Syst. Appl.* **2019**, *138*, 112821. [CrossRef]
9. Altini, N.; Cascarano, G.; Brunetti, A.; Marino, F.; Rocchetti, M.; Matino, S.; Venere, U.; Rossini, M.; Pesce, F.; Gesualdo, L.; et al. Semantic Segmentation Framework for Glomeruli Detection and Classification in Kidney Histological Sections. *Electronics* **2020**, *9*, 503. [CrossRef]
10. Altini, N.; Cascarano, G.D.; Brunetti, A.; De Feudis, I.; Buongiorno, D.; Rossini, M.; Pesce, F.; Gesualdo, L.; Bevilacqua, V. A deep learning instance segmentation approach for global glomerulosclerosis assessment in donor kidney biopsies. *Electronics* **2020**, *9*, 1768. [CrossRef]
11. Chen, H.; Qi, X.; Yu, L.; Dou, Q.; Qin, J.; Heng, P.-A. DCAN: Deep contour-aware networks for object instance segmentation from histology images. *Med. Image Anal.* **2017**, *36*, 135–146. [CrossRef]
12. Janowczyk, A.; Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J. Pathol. Inform.* **2016**, *7*, 29. [CrossRef]

13. Salvi, M.; Mogetta, A.; Meiburger, K.; Gambella, A.; Molinaro, L.; Barreca, A.; Papotti, M.; Molinari, F. Karpinski Score under Digital Investigation: A Fully Automated Segmentation Algorithm to Identify Vascular and Stromal Injury of Donors' Kidneys. *Electronics* **2020**, *9*, 1644. [CrossRef]
14. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **2020**, *63*, 101693. [CrossRef] [PubMed]
15. Bardis, M.; Houshyar, R.; Chantaduly, C.; Ushinsky, A.; Glavis-Bloom, J.; Shaver, M.; Chow, D.; Uchio, E.; Chang, P. Deep Learning with Limited Data: Organ Segmentation Performance by U-Net. *Electronics* **2020**, *9*, 1199. [CrossRef]
16. Dimauro, G.; Simone, L. Novel Biased Normalized Cuts Approach for the Automatic Segmentation of the Conjunctiva. *Electronics* **2020**, *9*, 997. [CrossRef]
17. Kasiviswanathan, S.; Bai Vijayan, T.; Simone, L.; Dimauro, G. Semantic Segmentation of Conjunctiva Region for Non-Invasive Anemia Detection Applications. *Electronics* **2020**, *9*, 1309. [CrossRef]
18. Dimauro, G.; De Ruvo, S.; Di Terlizzi, F.; Ruggieri, A.; Volpe, V.; Colizzi, L.; Girardi, F. Estimate of Anemia with New Non-Invasive Systems—A Moment of Reflection. *Electronics* **2020**, *9*, 780. [CrossRef]
19. Dimauro, G.; Caivano, D.; Di Pilato, P.; Dipalma, A.; Camporeale, M.G. A Systematic Mapping Study on Research in Anemia Assessment with Non-Invasive Devices. *Appl. Sci.* **2020**, *10*, 4804. [CrossRef]
20. Dimauro, G.; Guarini, A.; Caivano, D.; Girardi, F.; Pasciolla, C.; Iacobazzi, A. Detecting Clinical Signs of Anaemia from Digital Images of the Palpebral Conjunctiva. *IEEE Access* **2019**, *7*, 113488–113498. [CrossRef]
21. Kanniappan, S.; Samiayya, D.; Vincent P M, D.; Srinivasan, K.; Jayakody, D.; Reina, D.; Inoue, A. An Efficient Hybrid Fuzzy-Clustering Driven 3D-Modeling of Magnetic Resonance Imagery for Enhanced Brain Tumor Diagnosis. *Electronics* **2020**, *9*, 475. [CrossRef]
22. Srinivasan, K.; Gowthaman, T.; Nema, A. Application of structural group sparsity recovery model for brain MRI. In Proceedings of the SPIE 10806, Tenth International Conference on Digital Image Processing, Shanghai, China, 11–14 May 2018; p. 108065H.
23. Srinivasan, K.; Ankur, A.; Sharma, A. Super-resolution of Magnetic Resonance Images using deep Convolutional Neural Networks. In Proceedings of the IEEE International Conference on Consumer Electronics—Taiwan (ICCE-TW), Taipei, Taiwan, 12–14 June 2017; pp. 41–42.
24. Mesin, L.; Roatta, S.; Pasquero, P.; Porta, M. Automated Volume Status Assessment Using Inferior Vena Cava Pulsatility. *Electronics* **2020**, *9*, 1671. [CrossRef]
25. Mesin, L.; Pasquero, P.; Albani, S.; Porta, M.; Roatta, S. Semi-automated tracking and continuous monitoring of inferior vena cava diameter in simulated and experimental ultrasound imaging. *Ultrasound Med. Biol.* **2015**, *41*, 845–857. [CrossRef]
26. Mesin, L.; Pasquero, P.; Roatta, S. Tracking and Monitoring Pulsatility of a Portion of Inferior Vena Cava from Ultrasound Imaging in Long Axis. *Ultrasound Med. Biol.* **2019**, *45*, 1338–1343. [CrossRef] [PubMed]
27. Mesin, L.; Pasquero, P.; Roatta, S. Multi-directional assessment of Respiratory and Cardiac Pulsatility of the Inferior Vena Cava from Ultrasound Imaging in Short Axis. *Ultrasound Med. Biol.* **2020**, *46*, 3475–3482. [CrossRef] [PubMed]
28. Mesin, L.; Giovinazzo, T.; D'Alessandro, S.; Roatta, S.; Raviolo, A.; Chiacchiarini, F.; Porta, M.; Pasquero, P. Improved repeatability of the estimation of pulsatility of inferior vena cava. *Ultrasound Med. Biol.* **2019**, *45*, 2830–2843. [CrossRef] [PubMed]
29. Mesin, L.; Albani, S.; Sinagra, G. Non-invasive Estimation of Right Atrial Pressure using the Pulsatility of Inferior Vena Cava. *Ultrasound Med. Biol.* **2019**, *45*, 1331–1337. [CrossRef] [PubMed]
30. Albani, S.; Pinamonti, B.; Giovinazzo, T.; de Scordilli, M.; Fabris, E.; Stolfo, D.; Perkan, A.; Gregorio, C.; Barbati, G.; Geri, P.; et al. Accuracy of right atrial pressure estimation using a multi-parameter approach derived from inferior vena cava semi-automated edge-tracking echocardiography: A pilot study in patients with cardiovascular disorders. *Int. J. Cardiovasc. Imaging* **2020**, *36*, 1213–1225. [CrossRef]
31. Folino, A.; Benzo, M.; Pasquero, P.; Laguzzi, A.; Mesin, L. Messere, A.; Porta, M.; Roatta, S. Vena Cava Responsiveness to Controlled Isovolumetric Respiratory Efforts. *J. Ultrasound Med.* **2017**, *36*, 2113–2123. [CrossRef]

32. Chesbrough, H.W. *Open Innovation. The New Imperative for Creating and Profiting from Technology*; Harvard Business Review Press: Brighton, Boston, MA, USA, 2003.

electronics

MDPI

Article

Semantic Segmentation Framework for Glomeruli Detection and Classification in Kidney Histological Sections

Nicola Altini [1], Giacomo Donato Cascarano [1], Antonio Brunetti [1], Francescomaria Marino [1], Maria Teresa Rocchetti [2], Silvia Matino [2], Umberto Venere [2], Michele Rossini [2], Francesco Pesce [2], Loreto Gesualdo [2] and Vitoantonio Bevilacqua [1,*]

[1] Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari, 70126 Bari, Italy; nicola.altini@poliba.it (N.A.); giacomodonato.cascarano@poliba.it (G.D.C.); antonio.brunetti@poliba.it (A.B.); francescomaria.marino@poliba.it (F.M.)

[2] Department of Emergency and Organ Transplantation (DETO), Nephrology Unit, University of Bari Aldo Moro, 70126 Bari, Italy; mariateresa.rocchetti@unifg.it (M.T.R.); silviamatino@gmail.com (S.M.); u.venere@gmail.com (U.V.); michelerossini@libero.it (M.R.); f.pesce81@gmail.com (F.P.); loreto.gesualdo@uniba.it (L.G.)

* Correspondence: vitoantonio.bevilacqua@poliba.it

Received: 28 February 2020; Accepted: 17 March 2020; Published: 19 March 2020

Abstract: the evaluation of kidney biopsies performed by expert pathologists is a crucial process for assessing if a kidney is eligible for transplantation. In this evaluation process, an important step consists of the quantification of global glomerulosclerosis, which is the ratio between sclerotic glomeruli and the overall number of glomeruli. Since there is a shortage of organs available for transplantation, a quick and accurate assessment of global glomerulosclerosis is essential for retaining the largest number of eligible kidneys. In the present paper, the authors introduce a Computer-Aided Diagnosis (CAD) system to assess global glomerulosclerosis. The proposed tool is based on Convolutional Neural Networks (CNNs). In particular, the authors considered approaches based on Semantic Segmentation networks, such as SegNet and DeepLab v3+. The dataset has been provided by the Department of Emergency and Organ Transplantations (DETO) of Bari University Hospital, and it is composed of 26 kidney biopsies coming from 19 donors. The dataset contains 2344 non-sclerotic glomeruli and 428 sclerotic glomeruli. The proposed model consents to achieve promising results in the task of automatically detecting and classifying glomeruli, thus easing the burden of pathologists. We get high performance both at pixel-level, achieving mean F-score higher than 0.81, and Weighted Intersection over Union (IoU) higher than 0.97 for both SegNet and Deeplab v3+ approaches, and at object detection level, achieving 0.924 as best F-score for non-sclerotic glomeruli and 0.730 as best F-score for sclerotic glomeruli.

Keywords: semantic segmentation; convolutional neural networks; kidney biopsy; kidney transplantation; glomerulus detection; glomerulosclerosis

---

1. Introduction

The spread of Deep Learning (DL) techniques and frameworks has led to a revolution in the medical imaging field. The assessment of organ viability, by donor kidney biopsy examination, is essential prior to transplantation. The traditional evaluation of biopsies was based on the visual analysis by trained pathologists of biopsy slides using a light microscope which is a time consuming and highly variable procedure. The high variability between the observers resulted in poor reproducibility among pathologists, which may cause an inappropriate organ discard.

Therefore, the development of new techniques able to objectively and rapidly interpret donor kidney biopsy to support pathologist's decision making is strongly fostered. The increasing availability of whole-slide scanners, which facilitate the digitization of histopathological tissue, led to a new research field denoted as digital pathology and generated a strong demand for the development of Computer-Aided Diagnosis (CAD) systems. As stated in the literature, the application of deep learning techniques for the analysis of Whole-Slide Images (WSIs) has shown significant results and suggest that the integration of DL framework with CAD systems is a valuable solution.

In the realm of digital pathology, several recent studies have proposed CAD systems for glomerulus identification and classification in renal biopsies [1–8]. The eligibility for transplantation of a kidney retrieved from Expanded Criteria Donors (ECD) relies on rush histological examination of the organ to evaluate suitability for transplant [9]. The Karpinski score is based on the microscopic examination of four compartments: glomerular, tubular, interstitial and vascular, in order to assess the degree of chronic injury. For each compartment is assigned a score from 0 to 3 where 0 corresponds to normal histology and 3 to the highest degree of, respectively, global glomerulosclerosis, tubular atrophy, interstitial fibrosis and arterial and arteriolar narrowing [9,10]. The evaluation of global glomerulosclerosis requires detection and classification of all the glomeruli present in a kidney biopsy, distinguishing between healthy (non-sclerotic) and non-healthy (sclerotic) ones.

The two fundamental components that characterize a non-sclerotic glomerulus are the capillary tuft with the mesangium and the Bowman's capsule. The first one is placed inside the glomerulus while the second one is peripheral and has the function to contain the tuft. The space between these two components is called Bowman's space. From a morphological point of view, the non-sclerotic glomerulus generally has an elliptic form. The capillary tuft has a pomegranate form, caused by the contemporary presence of blue points (nuclei of cells), white areas (capillary lumens) and variable amount of regions with similar tonality and different levels of saturation (mesangial matrix). A non-healthy glomerulus, from the point of view of Karpinski's score, is a globally sclerotic glomerulus, namely a glomerulus where capillary lumens are completely obliterated for increase in extracellular matrix and Bowman's space is completely filled by collagenous material. Examples of non-sclerotic and sclerotic glomeruli are depicted in Figure 1.

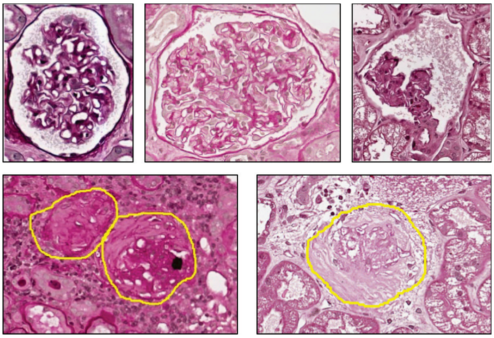


Figure 1. Glomeruli. Top row: non-sclerotic glomeruli. Bottom row: sclerotic glomeruli.

Ledbetter et al. proposed a Convolutional Neural Network to predict kidney function (evaluated as the quantity of primary filtrate that passes from the blood through the glomeruli per minute) in chronic kidney disease patients from whole-slide images of their kidney biopsies [3]. Gallego et al. proposed a method based on the pretrained AlexNet model [11] to perform glomerulus classification and detection in kidney tissue segments [2]. Gadermayr et al. focused on the segmentation of the glomeruli. The authors proposed two different CNN cascades for segmentation applications with sparse objects. They applied these approaches to the glomerulus segmentation task and compared them with conventional fully convolutional networks, coming to the conclusion that cascade networks can be a powerful tool for segmenting renal glomeruli [4]. Temerinac-Ott et al. compared the performance between a CNN classifier and a support-vector machines (SVM) classifier which exploits features extracted by histogram of oriented gradients (HOG) [12] for the task of glomeruli detection in WSIs with multiple stains, using a sliding window approach. The obtained results showed that the CNN method outperformed the HOG and SVM classifier [1]. Kawazoe et al. faced the task of glomeruli detection in multistained human kidney biopsy slides by using a Deep Learning approach based on Faster R-CNN [6]. Marsh et al. developed a deep learning model that recognizes and classifies sclerotic and non-sclerotic glomeruli in whole-slide images of frozen donor kidney biopsies. They used a Fully Convolutional Network (FCN) followed by a blob-detection algorithm [13], based on Laplacian-of-Gaussian, to post-process the FCN probability maps into object detection predictions [8]. Ginley et al. proposed a CAD to classify renal biopsies of patients with diabetic nephropathy [7], using a combination of classical image processing and novel machine learning techniques. Hermsen et al. adopted CNNs, namely an ensemble of five U-Nets, for segmentation of ten tissue classes from WSIs of periodic acid-Schiff (PAS) stained kidney transplant biopsies [14].

The analysis of the literature suggests that main works focused on the glomerular detection task only, without considering the further classification into sclerotic and non-sclerotic [1,2,4,6]. Few papers considered the assessment of global glomerulosclerosis from kidney biopsies [7,8,14].

In our previous works we focused on other kidney biopsies analysis tasks, such as classification of tubules and vessels [15] and classification of non-sclerotic and sclerotic glomeruli [5]. In this work, we propose a CAD system to address the segmentation and the classification tasks of glomeruli, in order to obtain a reliable estimate of Karpinski histological score. The proposed work allowed us to obtain better results than the literature in the classification task.

2. Materials

The kidney biopsies dataset analyzed in this paper has been provided by the Department of Emergency and Organ Transplantations (DETO) of the Bari University Hospital. Slides were digitized using a high-resolution whole-slide scanner with a scanning objective which has a 20× magnification corresponding to a resolution of 0.50 µm/pixel. All the biopsies provided by DETO clinicians are PAS stained sections from formalin fixed paraffin embedded tissue. The complete dataset is composed of 26 kidney biopsies coming from 19 donors. It contains 2344 non-sclerotic glomeruli and 428 sclerotic glomeruli. The dataset has been split into a train-validation (trainval) set and a test set. The trainval set has been further split into a train set and a validation set; the last one is used for tuning hyperparameters and for assessing the trend of the loss function and of accuracy during the training process. A detailed overview of the dataset is reported in Table 1.

Table 1. Dataset info.

Set	WSIs	Non-Sclerotic	Sclerotic	Ratio
Trainval set	19	1852	341	5.43 : 1
Test set	7	492	87	5.66 : 1
Dataset	26	2344	428	5.48 : 1

3. Methods

3.1. Semantic Segmentation Framework

Convolutional Neural Networks have had a widespread adoption in all kinds of image analysis tasks, starting from AlexNet which won ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC 2012) [16] by a huge margin [11], though pioneering work was already done by LeCun much earlier for handwritten digit recognition [17].

Semantic segmentation is a task which consists of classifying all the pixels belonging to an input image. In order to accomplish this task, most CNN semantic segmentation architectures are based on encoder-decoder networks. The encoder is devoted to the feature extraction process, shrinking the spatial dimensions while increasing the depth. The decoder has the task to recover the spatial information from the output of the encoder. Due to the several application in the medical imaging field, in this work we considered two main approaches based on SegNet and DeepLab v3+ architectures. The main SegNet applications regard segmentation tasks such as semantic segmentation of prostate cancer [18], gland segmentation from colon cancer histology images [19] and brain tumor segmentation from multi-modal magnetic resonance images [20]. DeepLab v3+ has been used for the semantic segmentation of colorectal polyps [21] and the automatic liver segmentation [22,23].

SegNet is a CNN architecture for semantic segmentation proposed by researchers at University of Cambridge [24]. As other semantic segmentation architectures, SegNet is composed of an encoder network and a corresponding decoder network, followed by a final pixel-wise classification layer. One clever point of SegNet is that it removes the necessity of learning the upsampling process, by storing indices used in max-pooling step in encoder and applying them when upsampling in the corresponding layers of the decoder.

DeepLab is an architecture proposed by Chen et al. [25]. One of the interesting novelties proposed by the authors of DeepLab is the atrous convolution, also known as dilated convolution. The idea has been commonly used in wavelet transform before being adapted to convolutions for deep learning. Atrous convolution consents to broaden the field of view of filters to incorporate larger context. It is, therefore, a valuable tool to tune the field of view, permitting identification of the right balance between context assimilation (large field of view) and fine localization (small field of view). We adopted DeepLab v3+ [26] with ResNet-18 [27] as backbone in our tests.

We replaced the last layer of both SegNet and DeepLab v3+ networks with a pixel-wise classification layer with 3 output classes (background, sclerotic glomeruli and non-sclerotic glomeruli); we used inverse class frequencies as class weights and pixel-wise cross-entropy as loss function.

3.2. Proposed Workflow

3.2.1. CAD Architecture

A high-level overview of the proposed CAD is depicted in Figure 2. The physicians can visualize the WSIs using Aperio ImageScope software. In order to perform supervised learning, we need labelled data. Pathologists can annotate the slides using ImageScope, and export the results in XML files, which we can use to feed our neural networks. After having trained our models, we can export the output in XML files, and physicians can see the CAD annotations always in ImageScope, with seamless integration. To accomplish the task of calculating the Karpinski histological score, we must make a careful choice for the architecture of the network. All the models have been trained and validated on a machine with the characteristics reported in Table 2.

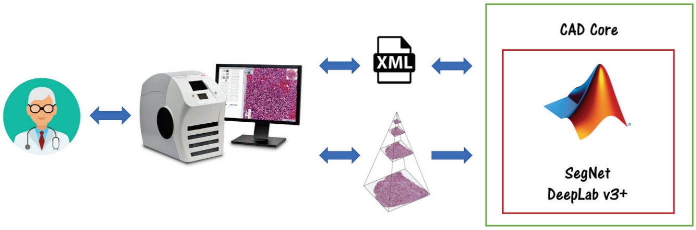


Figure 2. CAD architecture. Physicians can visualize and annotate the WSIs using Aperio ImageScope software. The developed Deep Learning models can interact with ImageScope through an XML interface.

Table 2. System Details.

System Details	
GPU	NVIDIA GTX 1060 with 6 GB of RAM
CPU	Intel Core i7-4790 CPU @ 3.60 Ghz
RAM	32 GB
OS	Microsoft Windows 10 Home
Tool	MATLAB R2019a

3.2.2. Semantic Segmentation Workflow

To obtain an estimate of the Karpinski score, we must detect and classify all the glomeruli which appear in the WSI. We first use a semantic segmentation CNN to obtain a pixel-level classification, distinguishing between pixels which belongs to background, sclerotic and non-sclerotic glomeruli. Then, we must turn these pixel-level classifications into object detections, so that we can count the number of sclerotic and non-sclerotic glomeruli. The general schema for our semantic segmentation-based glomerular detector is depicted in Figure 3.

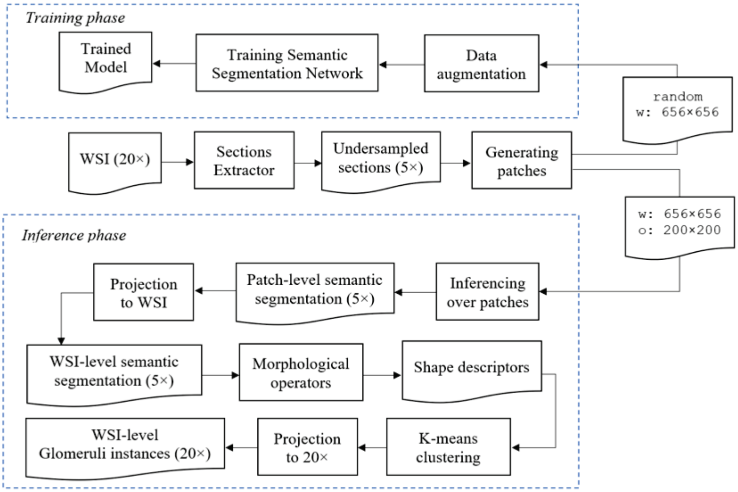


Figure 3. Semantic Segmentation approach architecture. The top part describes how to train the CNN. The bottom part explains how to use the trained model for performing inference, and the related morphological and clustering post-processing steps.

The first step in our workflow consists of segmenting the sections present in the WSI. At this purpose, we used classical Image Processing techniques as thresholding, morphological operators, connected components labelling, and eventually, clustering. A similar preprocessing step has also been done by Ledbetter et al. [3]. We refer to the module performing this step as Sections Extractor. To reduce the very large dimension of WSIs, which can be overwhelming for Deep Learning algorithms, we undersampled the sections by a factor of 4. The original WSIs have a magnification of 20×, after undersampling it becomes equivalent to a magnification of 5×. This operation leads to an effective downsampling of the images from a resolution of about 8000 × 8000 pixels to a resolution of about 2000 × 2000 pixels. Since the section obtained this way was still too large to fit in our GPU, we divided it in patches. During training, we randomly sampled patches of size 656 × 656, with a mechanism to avoid to take too many patches only with negatives samples. The random patches sampled during the training process are then fed to a data augmentation block that performs different augmentations, as reported in Table 3. Augmentations are generated on-the-fly for each epoch within random ranges, so the network always processes slightly different input data, thus reducing the risk of overfitting. In the inference phase, we take patches of size 656 × 656 pixels, with an overlap between successive windows of 200 × 200 pixels. Please note that in semantic segmentation is important to have a larger context for performing inference, when the approach involves a sliding window processing [28]. After we get the predicted masks for glomeruli at patch-level, we project them to the original WSI, to get the WSI-level predicted mask. At this point, we apply morphological operators to remove noisy points and smooth the glomeruli shapes. We then analyze shape descriptors to understand if it is necessary to perform a clustering operation. In the end, the obtained mask is projected to 20× resolution, corresponding to oversampling by 4, using nearest-neighbour interpolation. Please note that in this work, all the resizing operations involving the digital pathology images are obtained using bicubic interpolation, while all the resizing operations involving the categorical masks are obtained using nearest-neighbour interpolation.

3.2.3. Morphological Operators and Clustering

Adapting a semantic segmentation network to perform object detection poses some challenges. The task of semantic segmentation consists of labelling only individual pixels, which mainly captures textural information. In contrast to architectures explicitly tailored to Object Detection, such as Faster R-CNN [29] or Mask R-CNN [30], where there are anchor boxes, the network does not look for objects, it just tries to classify individual pixels. To extend the semantic segmentation model into an instance segmentation one, we must use different morphological operators and clustering algorithms as post-processing steps.

Morphological operators are applied only to binary masks obtained as the output of the semantic segmentation networks. First, we smooth the shapes of objects performing a morphological closing operation, with a disk of radius 5 pixels as structuring element, and with the morphological flood-fill operation. Then, we delete small objects and noisy points using opening operator, with a disk of radius 10 pixels as structuring element, and area opening operator, removing connected regions with an area below 1000 pixels. Examples are depicted in Figure 4, where binary masks are overlapped to the biopsy images for visualization purposes. Masks relative to non-sclerotic and sclerotic glomeruli are green and red colored, respectively. Lastly, we analyze the shape descriptors for each of these objects to understand if there are touching objects we need to cluster. The sequence of morphological operators used is depicted in Figure 5.

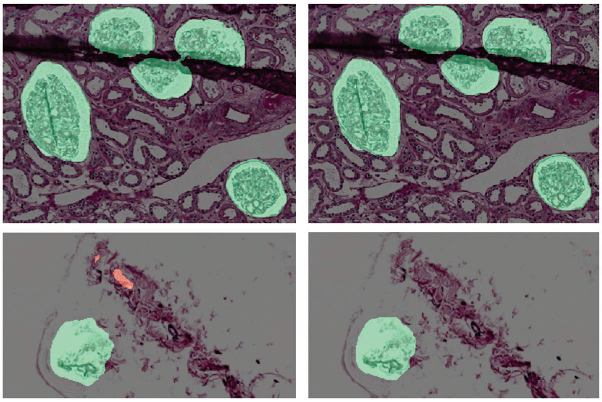


Figure 4. (**Left**) Semantic Segmentation output. (**Right**) After Morphological Operators.

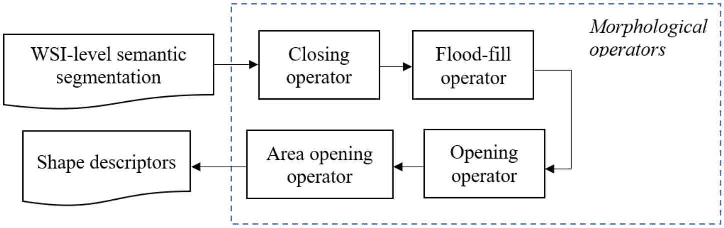


Figure 5. Morphological operators sequence applied to the output masks from the semantic segmentation network. The output of the morphological post-processing is used for calculating shape descriptors to eventually perform clustering.

An important observation is that individual glomeruli have convex shapes, so their area is pretty similar to their convex area. We perform a *K*-means clustering based on the difference between the convex area and the area, as specified in Equation (1).

$$deltaArea = convexHullArea - area \tag{1}$$

We decide the number *K* of clusters according to `deltaArea`: if `deltaArea` $\leq$ `900`, $K = 1$; if `deltaArea` $>$ `900` and `deltaArea` $\leq$ `5000`, $K = 2$; if `deltaArea` $>$ `5000`, $K = 3$. The values of `deltaArea` and the corresponding *K* have been empirically determined on the trainval set. Confusion matrices reported later have been obtained after the clustering with the configuration based on `deltaArea`. Examples of glomeruli before clustering are depicted in Figure 6a,c. The corresponding images after clustering are shown in Figure 6b,d.

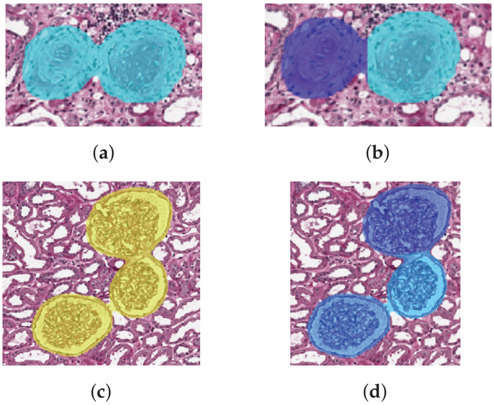

(**a**) (**b**) (**c**) (**d**)

Figure 6. Examples of K-means clustering for both sclerotic and non-sclerotic glomeruli. The number *K* of clusters is determined according to `deltaArea` defined in (1). (**a**) Sclerotic glomeruli before clustering. (**b**) Sclerotic glomeruli after clustering, with $K = 2$. (**c**) Non-sclerotic glomeruli before clustering. (**d**) Non-sclerotic glomeruli after clustering, with $K = 3$.

3.2.4. Data Augmentation

Tellez et al. analyzed the problem of stain color variation in digital pathology very deeply [31]. They proposed different solutions for both stain color augmentation and stain color normalization. In this work, we exploited techniques proposed by them such as morphological transformations and Hue-Saturation-Value (HSV) shifts. An interesting morphological transformation is the elastic deformation; it was originally proposed by Simard et al. [32] for the analysis of visual documents, and then has had a widespread application in medical imaging, as also shown by U-Net authors [28]. We used elastic deformation to generate plausible alterations of glomeruli shapes, increasing the variability of training images and thus reducing the risk of overfitting. An example of elastic deformation applied to our images is depicted in Figure 7. Examples of HSV shift are depicted in Figure 8.

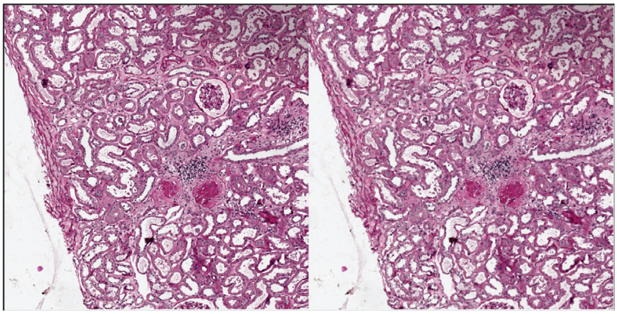


Figure 7. Elastic deformation example. Left: original image. Right: after elastic deformation with $\sigma = 6.29$, $\alpha = 340$.

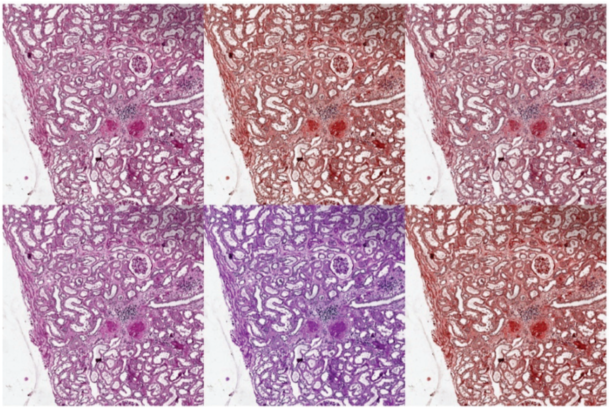


Figure 8. HSV shift examples. Top Left: original image. Top Center: $\Delta H = +0.18$, $\Delta S = +0.03$. Top Right: $\Delta H = +0.06$, $\Delta S = -0.06$. Bottom Left: $\Delta H = -0.04$, $\Delta S = -0.02$. Bottom Center: $\Delta H = -0.11$, $\Delta S = +0.10$. Bottom Right: $\Delta H = +0.18$, $\Delta S = +0.09$.

A summary of the data augmentation techniques used for the training process is reported in Table 3. The augmentations in group 1 are independently performed, each with a given probability p. Resize augmentation used here is slightly different from standard resize; in fact, we apply mirroring padding (instead of zero padding) when we perform a resize which shrinks the image size. Augmentations, such as mirroring padding, which alter the morphology of the image are also executed for the mask. From the augmentations reported in group 2, only one is made. Group 3 contains only one augmentation, which is performed with a given probability. The augmentations are performed in the order they compare in the table, i.e., before the four in group 1, then one of group 2 and in the end the one of group 3.

Table 3. Augmentations.

Data Augmentation	
Type	Details
Group 1	
Rotate	$\theta = 90$, $p = 0.25$
Flip left-right	$p = 0.25$
Flip upside-down	$p = 0.25$
Resize	$resize \in [0.8, 1.2]$, $p = 0.25$
Group 2	
Gaussian Noise	$\sigma \in [0, 0.01]$, $p = 0.1$
Gaussian Blur	$\sigma \in [0, 0.1]$, $p = 0.1$
Elastic Deformation	$\sigma \in [2, 5]$, $\alpha \in [100, 300]$, $p = 0.2$
Group 3	
HSV shift	$\Delta S \in [-0.1, 0.1]$, $\Delta H \in [-0.1, 0.1]$, $p = 0.5$

3.2.5. Hyperparameters Tuning

We tried different semantic segmentation network architectures. For SegNet and DeepLab v3+ we tuned hyperparameters according to Table 4. Please note that DeepLab v3+ with ResNet-18 backbone is more lightweight than SegNet, and this allowed us to use a larger mini-batch size, with eight patches per mini-batch. With our GPU, SegNet was trained with only one patch per mini-batch. More details about hyperparameters can be found in MATLAB documentation [33].

Table 4. Hyperparameters.

Hyperparameter	**SegNet**	**Deeplab v3+**
Optimizer	SGDM	SGDM
LearnRateSchedule	'piecewise'	'piecewise'
LearnRateDropPeriod	10	10
LearnRateDropFactor	0.3	0.3
Momentum	0.9	0.9
InitialLearnRate	0.001	0.001
L2Regularization	0.005	0.005
MaxEpochs	30	30
MiniBatchSize	1	8
Shuffle	'every-epoch'	'every-epoch'
ValidationPatience	10	10
ValidationFrequency	1 per epoch	1 per epoch

4. Experimental Results

We distinguish between the results obtained at pixel-level (semantic segmentation task) and at object detection level.

In particular, for the semantic segmentation task we group the metrics in Dataset Metrics and Class Metrics [33].

The group of **Dataset Metrics** includes semantic segmentation metrics aggregated over the data set: *Global Accuracy*, *Mean Accuracy* (the mean of the accuracies calculated per class), *Mean IoU* (the mean of the IoUs calculated per class), *Weighted IoU* (mean of the IoUs, weighted by the number of pixels in the class) and *Mean F-score* (mean of the F-measures calculated per class).

The group of **Class Metrics** includes semantic segmentation metrics calculated for each class, namely: *Accuracy* (2), *IoU* (3) and *Mean F-score* (F-measure for each class, averaged over all images).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{3}$$

For the object detection task, confusion matrices are calculated assuming that a true positive match between predicted mask and ground truth mask has pixel-wise IoU (3) of at least 0.2. Besides confusion matrices, the metrics used for assessing the results of the object detection task are:

$$Precision = \frac{TP}{TP + FP}, \tag{4}$$

$$Recall = \frac{TP}{TP + FN}, \tag{5}$$

$$F_1\ Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \tag{6}$$

The best results on non-sclerotic glomeruli have been obtained using DeepLab v3+, while for sclerotic glomeruli the best model was SegNet. An example of the output of our semantic segmentation framework is depicted in Figure 9.

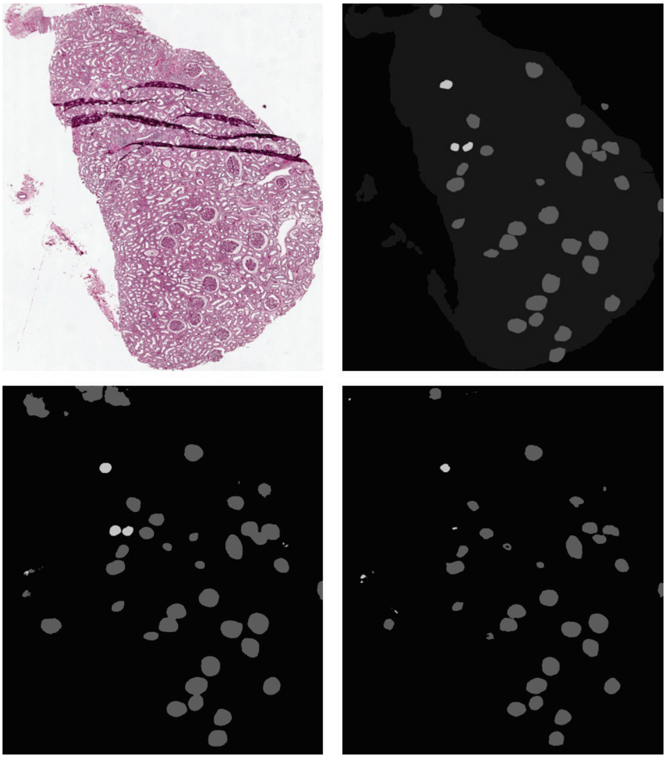


Figure 9. Top Left: original image. Top Right: ground truth. Bottom Left: SegNet prediction. Bottom Right: DeepLab v3+ prediction. Sclerotic glomeruli and non-sclerotic ones are white and gray colored, respectively.

4.1. Pixel-Level Metrics

Pixel-level dataset metrics for both SegNet and DeepLab v3+ are reported in Table 5. The pixel-level class metrics of SegNet and DeepLab v3+ are reported in Tables 6 and 7, respectively.

The normalized pixel-level confusion matrix are in Tables 8 and 9. Pixel-level confusion matrices are normalized per row; B, NS, S stand for Background, Non-sclerotic and sclerotic, respectively.

Table 5. Dataset Metrics.

CNN	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean F-Score
SegNet	0.98346	0.86385	0.71352	0.97156	0.81784
Deeplab v3+	0.99179	0.76884	0.72873	0.98434	0.84614

Table 6. Class Metrics SegNet.

Class	Accuracy	IoU	Mean F-Score
Background	0.98636	0.98294	0.99243
Non-sclerotic	0.91925	0.66546	0.83239
sclerotic	0.68594	0.49215	0.69686

Table 7. Class Metrics Deeplab v3+.

Class	Accuracy	IoU	Mean F-Score
Background	0.99690	0.99172	0.96684
Non-sclerotic	0.88199	0.80872	0.93306
Sclerotic	0.42764	0.38574	0.63852

Table 8. Normalized pixel-level Confusion Matrix SegNet.

		Prediction		
		B	NS	S
Ground Truth	B	98.64%	1.26%	0.10%
	NS	8.07%	91.93%	0.00%
	S	30.97%	0.44%	68.59%

Table 9. Normalized pixel-level Confusion Matrix Deeplab v3+.

		Prediction		
		B	NS	S
Ground Truth	B	99.69%	0.28%	0.03%
	NS	11.78%	88.20%	0.02%
	S	50.57%	6.67%	42.76%

4.2. Object Detection Metrics

In object detection confusion matrices B, NS, S stand for Background, Non-sclerotic and Sclerotic, respectively.

The object detection confusion matrices for SegNet and DeepLab v3+ are reported in Tables 10 and 11, respectively. The detection metrics for both the proposed models and a comparison with the method proposed by Marsh et al. [8] are reported in Table 12. The SegNet-based model obtained a better F-score for both the glomeruli classes. The DeepLab v3+-based model obtained a better F-score for non-sclerotic glomeruli and a slightly worse F-score for sclerotic glomeruli.

Table 10. Object Detection Confusion Matrix SegNet.

		Prediction		
		NS	S	B
Ground Truth	NS	436	0	56
	S	1	58	28
	B	86	14	–

Table 11. Object Detection Confusion Matrix Deeplab v3+.

		Prediction		
		NS	S	B
Ground Truth	NS	449	0	43
	S	7	41	39
	B	24	1	–

Table 12. Performance Comparison for Detection Metrics.

Author	**Model**	**Class**	**Recall**	**Precision**	**F-Score**
Marsh et al. [8]	FCN + blob-detection	NS	0.885	0.813	0.848
		S	0.698	0.607	0.649
Proposed approach	SegNet	NS	0.886	0.834	0.859
		S	0.667	0.806	0.730
	DeepLab v3+	NS	0.913	0.935	0.924
		S	0.471	0.976	0.636

5. Conclusions and Future Work

The proposed approach allowed us to obtain high performance both at pixel and object detection level. The semantic segmentation achieved mean F-score higher than 0.81 and Weighted IoU higher than 0.97 for both SegNet and Deeplab v3+ approaches; the glomeruli detection achieved 0.924 as best F-score for non-sclerotic glomeruli and 0.730 as best F-score for sclerotic glomeruli. We compared our obtained performance with the state of the art. As stated in the Section 1, there are three main works that face the problem of glomerular classification. Ginley et al. considered the glomerular assessment for patients affected by diabetic nephropathy but not for transplantation purposes [7]. Hermsen et al. considered many tissue classes, but the number of sclerotic glomeruli in their datasets is too small for a comparison with our method [14]. Marsh et al. considered the problem of global glomerulosclerosis from kidney transplant biopsies with haematoxylin and eosin (HE) stain [8]. The performance comparison between our proposed methods and Marsh et al. work is reported in Table 12. The obtained results show an improvement over the work of Marsh et al. Thus, CNNs for Semantic Segmentation are a viable approach for the purpose of glomerular segmentation and classification, allowing the obtaining of a reliable estimate of the global glomerulosclerosis. Assessing the suitability of kidney from ECD donors relies in many centers on the histological examination of kidney biopsies performed at the time of organ retrieval and processed and evaluated by on-call pathologist that, not necessarily, is an expert trained in renal pathology. The importance of training in renal pathology when assessing biopsy of such cases has been evaluated in some studies reporting better correlation with subsequent allograft outcome of histological scores provided by renal pathologists compared to those provided by general pathologist with potential risk of "overscoring" and the potential of discarding kidneys that could have been potentially transplanted [34–36]. The results were validated by the renal pathologists which assessed the reliability of the proposed workflow; the applied methodology constitutes a milestone in the creation of a CAD system for the renal transplant assessment. The proposed system could help pathologists in accomplishing the laborious task of evaluating the eligibility of a kidney for transplantation, providing a rapid and accurate result. Future work will include the use of Deep Learning models explicitly designed for the detection task, such as Faster R-CNN and Mask R-CNN.

Author Contributions: Conceptualization, N.A., G.D.C. and V.B.; Data curation, M.R., F.P. and L.G.; Methodology, N.A. and G.D.C.; Supervision, F.M., L.G. and V.B.; Validation, M.T.R., M.R., F.P. and L.G.; Writing—original draft, N.A.; Writing—review & editing, G.D.C., A.B., F.M., M.T.R., S.M., U.V., M.R., F.P., L.G. and V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially funded by the Italian Apulian Region project "SOS – Smart Operating Shelter" (INNONETWORK n. 9757YR7).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Temerinac-Ott, M.; Forestier, G.; Schmitz, J.; Hermsen, M.; Bräsen, J.; Feuerhake, F.; Wemmert, C. Detection of glomeruli in renal pathology by mutual comparison of multiple staining modalities. In Proceedings of the IEEE 10th International Symposium on Image and Signal Processing and Analysis, Ljubljana, Slovenia, 18–20 Sepember 2017; pp. 19–24.
2. Gallego, J.; Pedraza, A.; Lopez, S.; Steiner, G.; Gonzalez, L.; Laurinavicius, A.; Bueno, G. Glomerulus classification and detection based on convolutional neural networks. *J. Imaging* **2018**, *4*, 20. [CrossRef]
3. Ledbetter, D.; Ho, L.; Lemley, K.V. *Prediction of Kidney Function from Biopsy Images Using Convolutional Neural Networks*; Los Alamos National Lab: Santa Fe, NM, USA, 2017; pp. 1–11.
4. Gadermayr, M.; Dombrowski, A.K.; Klinkhammer, B.M.; Boor, P.; Merhof, D. CNN Cascades for Segmenting Whole Slide Images of the Kidney. *arXiv* **2017**, arXiv:1708.00251.
5. Cascarano, G.D.; Debitonto, F.S.; Lemma, R.; Brunetti, A.; Buongiorno, D.; De Feudis, I.; Guerriero, A.; Rossini, M.; Pesce, F.; Gesualdo, L.; et al. *An Innovative Neural Network Framework for Glomerulus Classification Based on Morphological and Texture Features Evaluated in Histological Images of Kidney Biopsy*; Springer: Berlin, Germany, 2019; pp. 727–738._66. [CrossRef]
6. Kawazoe, Y.; Shimamoto, K.; Yamaguchi, R.; Shintani-Domoto, Y.; Uozaki, H.; Fukayama, M.; Ohe, K. Faster R-CNN-based glomerular detection in multistained human whole slide images. *J. Imaging* **2018**, *4*, 91. [CrossRef]
7. Ginley, B.; Lutnick, B.; Jen, K.Y.; Fogo, A.B.; Jain, S.; Rosenberg, A.; Walavalkar, V.; Wilding, G.; Tomaszewski, J.E.; Yacoub, R.; et al. Computational Segmentation and Classification of Diabetic Glomerulosclerosis. *J. Am. Soc. Nephrol.* **2019**, *30*, 1953–1967. [CrossRef] [PubMed]
8. Marsh, J.N.; Matlock, M.K.; Kudose, S.; Liu, T.C.; Stappenbeck, T.S.; Gaut, J.P.; Swamidass, S.J. Deep learning global glomerulosclerosis in transplant kidney frozen sections. *IEEE Trans. Med. Imaging* **2018**, *37*, 2718–2728. [CrossRef] [PubMed]
9. Karpinski, J.; Lajoie, G.; Cattran, D.; Fenton, S.; Zaltzman, J.; Cardella, C.; Cole, E. Outcome of kidney transplantation from high-risk donors is determined by both structure and function. *Transplantation* **1999**, *67*, 1162–1167. [CrossRef]
10. Remuzzi, G.; Grinyò, J.; Ruggenenti, P.; Beatini, M.; Cole, E.H.; Milford, E.L.; Brenner, B.M. Early experience with dual kidney transplantation in adults using expanded donor criteria. *J. Am. Soc. Nephrol.* **1999**, *10*, 2591–2598.
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. 2012 AlexNet. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2012. [CrossRef]
12. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IIEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
13. Lindeberg, T. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int. J. Comput. Vis.* **1993**, *11*, 283–318. [CrossRef]

14. Hermsen, M.; de Bel, T.; Den Boer, M.; Steenbergen, E.J.; Kers, J.; Florquin, S.; Roelofs, J.J.; Stegall, M.D.; Alexander, M.P.; Smith, B.H.; et al. Deep Learning–Based Histopathologic Assessment of Kidney Tissue. *J. Am. Soc. Nephrol.* **2019**, *30*, 1968–1979. [CrossRef]
15. Bevilacqua, V.; Pietroleonardo, N.; Triggiani, V.; Brunetti, A.; Di Palma, A.M.; Rossini, M.; Gesualdo, L. An innovative neural network framework to classify blood vessels and tubules based on Haralick features evaluated in histological images of kidney biopsy. *Neurocomputing* **2017**, *228*, 143–153. [CrossRef]
16. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
17. Lecun, Y.; Bottou, L.; Bengio, Y.; Ha, P. LeNet. *Proc. IEEE* **1998**. [CrossRef]
18. Ma, Z.; Li, J.; Salemi, H.; Arnold, C.; Knudsen, B.S.; Gertych, A.; Ing, N. Semantic segmentation for prostate cancer grading by convolutional neural networks. In Proceedings of the SPIE Medical Imaging 2018: Digital Pathology, Houston, TX, USA, 10–15 February 2018; p. 46. [CrossRef]
19. Tang, J.; Li, J.; Xu, X. Segnet-based gland segmentation from colon cancer histology images. In Proceedings of the IEEE 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Jiangsu, China, 18–20 May 2018; pp. 1078–1082.
20. Alqazzaz, S.; Sun, X.; Yang, X.; Nokes, L. Automated brain tumor segmentation on multi-modal MR image using SegNet. *Comput. Visual Media* **2019**, *5*, 209–219. [CrossRef]
21. Xiao, W.T.; Chang, L.J.; Liu, W.M. Semantic segmentation of colorectal polyps with DeepLab and LSTM networks. In Proceedings of the IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Taiwan, China, 19–21 May 2018; pp. 1–2.
22. Tang, W.; Zou, D.; Yang, S.; Shi, J. DSL: Automatic liver segmentation with faster R-CNN and DeepLab. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; Springer: Berlin, Germany, 2018; pp. 137–147.
23. Tang, W.; Zou, D.; Yang, S.; Shi, J.; Dan, J.; Song, G. A two-stage approach for automatic liver segmentation with Faster R-CNN and DeepLab. *Neural Comput. Appl.* **2020**, 1–10. [CrossRef]
24. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
25. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
26. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin, Germany, 2018; Volume 11211 LNCS, pp. 833–851._49. [CrossRef]
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
28. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin, Germany, 2015; Volume 9351, pp. 234–241._28. [CrossRef]
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
30. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Rio, Brazil, 14–21 October 2017. [CrossRef]
31. Tellez, D.; Litjens, G.; Bándi, P.; Bulten, W.; Bokhorst, J.M.; Ciompi, F.; van der Laak, J. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med. Image Anal.* **2019**, *58*, 101544. [CrossRef]
32. Simard, P.Y.; Steinkraus, D.; Platt, J.C. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In Proceedings of the ICDAR, Edinburgh, UK, 3–6 August 2003; Volume 3.
33. MATLAB Documentation. Available online: https://www.mathworks.com/help/matlab/ (accessed on 19 November 2019).

34. Azancot, M.A.; Moreso, F.; Salcedo, M.; Cantarell, C.; Perello, M.; Torres, I.B.; Montero, A.; Trilla, E.; Sellarés, J.; Morote, J.; et al. The reproducibility and predictive value on outcome of renal biopsies from expanded criteria donors. *Kidney Int.* **2014**, *85*, 1161–1168. [CrossRef]
35. Haas, M. Donor kidney biopsies: pathology matters, and so does the pathologist. *Kidney Int.* **2014**, *85*, 1016–1019. [CrossRef]
36. Girolami, I.; Gambaro, G.; Ghimenton, C.; Beccari, S.; Caliò, A.; Brunelli, M.; Novelli, L.; Boggi, U.; Campani, D.; Zaza, G.; et al. Pre-implantation kidney biopsy: value of the expertise in determining histological score and comparison with the whole organ on a series of discarded kidneys. *J. Nephrol.* **2020**, *33*, 167–176. [CrossRef] [PubMed]

electronics

MDPI

Article

Karpinski Score under Digital Investigation: A Fully Automated Segmentation Algorithm to Identify Vascular and Stromal Injury of Donors' Kidneys

Massimo Salvi [1,*], Alessandro Mogetta [1], Kristen M. Meiburger [1], Alessandro Gambella [2], Luca Molinaro [3], Antonella Barreca [3], Mauro Papotti [4] and Filippo Molinari [1]

[1] Department of Electronics and Telecommunications, Polytechnic of Turin, 10129 Turin, Italy; mogettaalessandro@gmail.com (A.M.); kristen.meiburger@polito.it (K.M.M.); filippo.molinari@polito.it (F.M.)
[2] Pathology Unit, Department of Medical Sciences, University of Turin, 10126 Turin, Italy; alessandro.gambella@unito.it
[3] Division of Pathology, A.O.U. Città della Salute e della Scienza Hospital, 10126 Turin, Italy; luca.molinaro@unito.it (L.M.); antonella.barreca@libero.it (A.B.)
[4] Department of Oncology, University of Turin, 10126 Turin, Italy; mauro.papotti@unito.it
* Correspondence: massimo.salvi@polito.it

Received: 14 September 2020; Accepted: 3 October 2020; Published: 8 October 2020

Abstract: In kidney transplantations, the evaluation of the vascular structures and stromal areas is crucial for determining kidney acceptance, which is currently based on the pathologist's visual evaluation. In this context, an accurate assessment of the vascular and stromal injury is fundamental to assessing the nephron status. In the present paper, the authors present a fully automated algorithm, called RENFAST (Rapid EvaluatioN of Fibrosis And vesselS Thickness), for the segmentation of kidney blood vessels and fibrosis in histopathological images. The proposed method employs a novel strategy based on deep learning to accurately segment blood vessels, while interstitial fibrosis is assessed using an adaptive stain separation method. The RENFAST algorithm is developed and tested on 350 periodic acid–Schiff (PAS) images for blood vessel segmentation and on 300 Massone's trichrome (TRIC) stained images for the detection of renal fibrosis. In the TEST set, the algorithm exhibits excellent segmentation performance in both blood vessels (accuracy: 0.8936) and fibrosis (accuracy: 0.9227) and outperforms all the compared methods. To the best of our knowledge, the RENFAST algorithm is the first fully automated method capable of detecting both blood vessels and fibrosis in digital histological images. Being very fast (average computational time 2.91 s), this algorithm paves the way for automated, quantitative, and real-time kidney graft assessments.

Keywords: kidney transplantation; digital pathology; deep learning; kidney fibrosis; blood vessel segmentation; convolutional neural networks

---

1. Introduction

Kidney allograft transplant is experiencing a broad revolution, thanks to an increasing understanding of the pathologic mechanisms behind rejection and the introduction of new techniques and procedures for transplants [1]. The primary focus during kidney transplants has always been the identification, assessment, and treatment of allograft rejection. However, recently, a new issue has come to light: a shortage of donor organs. To solve this impasse, selection criteria were revised, leading to the so-called "expanded criteria donor" approach: kidneys that once would have been excluded because of the donors' clinical history or those deriving from deceased patients are nowadays carefully used [2,3].

In this context, the preimplantation evaluation of donors' kidneys has become more and more crucial. The pathologist's challenge is to recognize early signs of degeneration to "predict" the organs' functionality and performance. This analysis, usually based on periodic acid–Schiff (PAS) and trichrome (TRIC) staining, is focused on the glomeruli, tubules, vessels, and cortical parenchyma of the donor kidney, searching for glomerulosclerosis, tubule atrophy, vascular damage, or interstitial fibrotic replacement, respectively (Figure 1). The Karpinski score is then applied to grade the injury of the donor kidney. This score is based on a semiquantitative evaluation of the structures mentioned above. For each of the four compartments (glomeruli, tubules, blood vessels, and cortical parenchyma), the pathologist summarizes the evaluation in a four-grade score, ranging from 0 (absence of injury) to 3 (marked injuries); the total score is expressed out of 12 [4]. Notably, both arteries and arterioles are considered in vascular damage assessment, characterized by progressive thickening of their wall and shrinkage of their lumen. At the same time, the cortical parenchyma could be replaced by fibrous connective tissue [5,6].

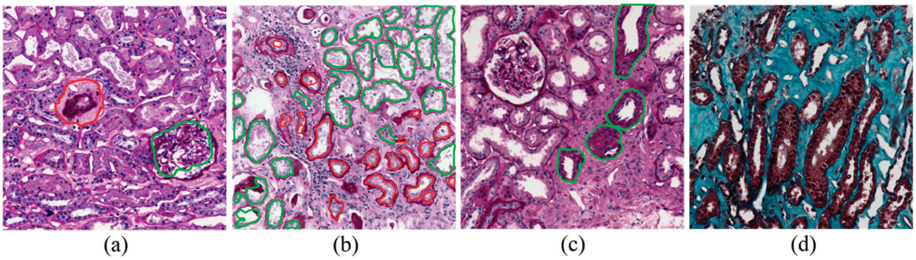


Figure 1. Histological features assessed to determine the Karpinski score. (**a**) Glomerulosclerosis: examples of a healthy and sclerotic glomerulus are shown in green and red, respectively; (**b**) Tubular atrophy: healthy and atrophic tubules are highlighted in green and red, respectively; (**c**) Vascular damage: blood vessels are outlined in green; (**d**) Cortical parenchyma: renal fibrosis is represented by the turquoise zone.

The preimplantation kidney evaluation is a delicate, crucial activity for pathology laboratories. It is time-consuming, usually performed with urgency, and has a marked impact on the daily diagnostic routine. Moreover, the evaluation is operator-dependent, with a significant rate of inter-observer difference [7]. In this challenging and evolving panorama, the introduction and application of an automated analysis algorithm would be of compelling importance.

In the last few years, several strategies have been proposed for the segmentation of kidney blood vessels and for the quantification of fibrotic tissue in biopsy images. Bevilacqua et al. [8] employed an artificial neural network (ANN) to detect blood vessels in histological kidney images. Lumen regions were firstly detected by applying fixed thresholding and morphological operators. Seeded region growing was then implemented to extract the membrane all around the segmented objects. Finally, a neural network based on Haralick texture features [9] was used to distinguish between blood vessels and tubular structures. Although well structured, this strategy suffers from several limitations. First, blood vessels with small or absent lumen cannot be segmented using the described approach. In addition, stain variability greatly influences the performance of the region growing, causing imprecise recognition of the blood vessel borders. Finally, the high variability in the shapes, dimensions, and textural characteristics of tubules seriously affects the classification provided by the network. Tey et al. [10] proposed an algorithm for the quantification of interstitial fibrosis (IF) based on color image segmentation and tissue structure identification in biopsy samples stained with Massone's trichrome (TRIC). All the renal structures were identified by employing color space transformations and structural feature extraction from the images. Then, the regions of fibrotic tissue were identified by removing all the non-fibrotic structures from the biopsy tissue area. This approach leads to fast identification of renal fibrotic tissue, but it is not free from limitations. First of all, there is a loss of information during the color space transformation and, in the presence of high stain variability, the method is not able to correctly classify all the renal structures. Moreover, being based on the identification and subsequent removal of non-fibrotic regions from the tissue, an error in the segmentation of these structures causes inaccurate quantification of interstitial fibrosis. Fu et al. [11] proposed a convolutional neural network (CNN) for fibrotic tissue segmentation in atrial tissue stained with Massone's trichrome. The network, consisting of 11 convolutional layers, was trained on a three-class problem (background vs. fibrosis vs. myocytes), giving the RGB image as input and the corresponding manual mask as the target. This approach provides fast detection of fibrotic areas of the tissue but presents one major limitation: color variability. Stain variations may affect both the training of the network and the correct segmentation of fibrotic tissue, and every mis-segmentation error leads to incorrect detection and quantification of interstitial fibrosis.

In this paper, we present a novel method for the detection of blood vessels and for the quantification of interstitial fibrosis in kidney histological images. To the best of our knowledge, no automated solution has been proposed so far to cope with the issue of stain variability in PAS and TRIC images. Our approach employs a preprocessing stage specifically designed to address the problem of color variability. The proposed algorithm for the segmentation of vascular structures exploits a deep learning approach combined with the detection of cellular structures to accurately segment blood vessels in PAS stained images. Interstitial fibrosis is assessed using an adaptive stain separation method to detect all the fibrotic areas within the histological tissue.

2. Materials and Methods

Here we present an automated method called RENFAST (Rapid EvaluatioN of Fibrosis And vesselS Thickness). The RENFAST algorithm is a deep-learning-based method for the segmentation of renal blood vessels and fibrosis. A flowchart of the proposed method is sketched in Figure 2. In the following sections, a detailed description of the algorithm is provided.

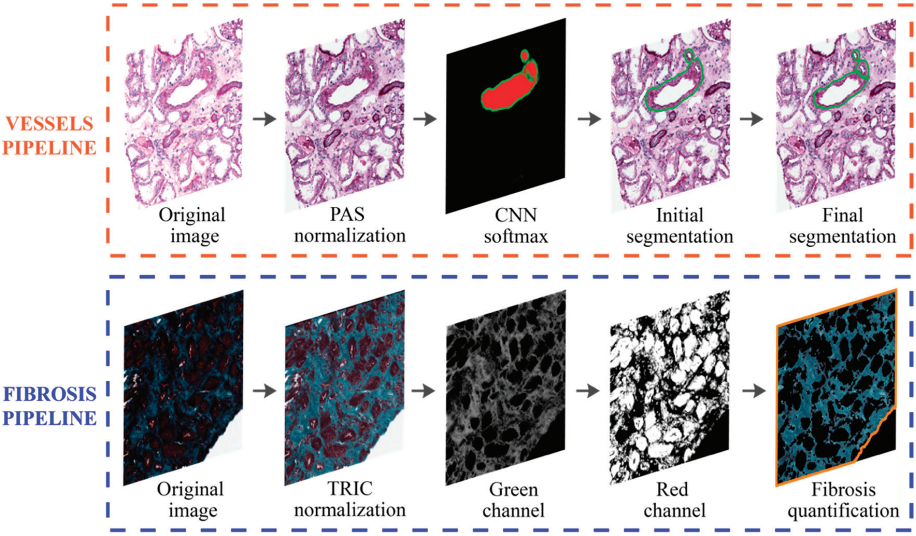


Figure 2. Flowchart of the RENFAST (Rapid EvaluatioN of Fibrosis And vesselS Thickness) algorithm for vessel and fibrosis segmentation. The first row illustrates the pipeline for blood vessel detection, while the second row shows the workflow of fibrosis segmentation. After PAS (periodic acid–Schiff) color normalization, blood vessels are detected using a deep learning method (CNN) and ad hoc post-processing. Kidney fibrosis is segmented through TRIC (Massone's trichrome) normalization followed by adaptive stain separation.

2.1. Database Description

The whole slide images (WSIs) of kidney biopsy specimens of 65 patients (median age 51 years, range 29–74 years) were used for this work; these were collected at the Division of Pathology, AOU Città della Salute e della Scienza Hospital, Turin, Italy and then anonymized. The pathology laboratory managed the biopsied samples of each kidney according to the kidney transplant biopsy's internal protocol. The tissue was fixed with Serra fixative and then processed in an urgency regimen using a microwave processor or LOGOS J processor (Milestone, Bergamo, Italy). Samples were then paraffin-embedded and serially sectioned (5 µm), mounted onto adhesive slides, and stained with PAS and TRIC. Finally, all the slides produced were scanned with a Hamamatsu NanoZoomer S210 Digital slide scanner (Turin, Italy), providing a magnification of ×100 (conversion factor: 0.934 µm/pixel). For each patient (n = 65), an expert pathologist (A.B.) manually extracted 10 images with dimensions of 512 × 512 pixels, for a total of 650 images. After consensus, manual annotations of blood vessels and fibrosis were generated by two operators (A.G. and L.M.). Table 1 shows the overall dataset composition. The image dataset, along with the annotations, is available at https://data.mendeley.com/datasets/m2t49zf6xr/1.

Table 1. Dataset used in this work.

Dataset	Subset	Stain	# Patients	# Images
Vessels	TRAIN	PAS	30	300
	TEST	PAS	5	50
Fibrosis	TRAIN	TRIC	25	250
	TEST	TRIC	5	50

2.2. Stain Normalization

The proposed algorithm employs a specific preprocessing stage, called stain normalization, to reduce the color variability of the histological samples. Previous studies have shown that stain variability significantly affects the performance of automatic algorithms in digital pathology [12,13]. The procedure of stain normalization allows for transforming a source image I into another image I_{NORM}, through the operation $I_{NORM} = f(I, I_{REF})$, where I_{REF} is a reference image and $f(\cdot)$ is the function that applies the color intensities of I_{REF} to the source image [14]. The reference image is chosen by the pathologist as the image with the most optimal tissue staining and visual appearance. For each image of the dataset, the RENFAST algorithm applies the same stain normalization method that we developed in our previous work [15]. First, the image is converted to the optical density space (OD) where the relationship between stain concentration and light intensity is linear. The algorithm then estimates the stain color appearance matrix (W) and the stain density map (H) for both the source and reference images. In order to apply the normalization, the stain density map of the source image is adjusted using the following equation:

$$I_{NORM} = W_{REF} \cdot \frac{H_{SOURCE}}{H_{REF}} \tag{1}$$

where $(\cdot)_{SOURCE}$ and $(\cdot)_{REF}$ denote the source and reference images, respectively. Finally, the normalized image is converted back from the OD space to RGB. Figure 3 illustrates the color normalization process for sample PAS and TRIC images.

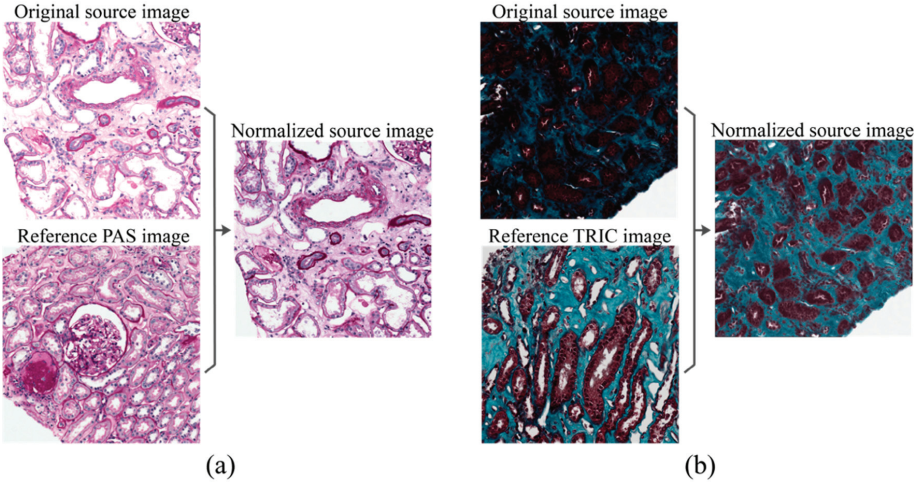


Figure 3. Stain normalization performed by the RENFAST algorithm. (**a**) PAS normalization; (**b**) TRIC normalization.

2.3. Deep Network Architecture

After stain normalization, the first step performed by the RENFAST algorithm is semantic segmentation using a convolutional neural network (CNN). To perform blood vessel segmentation, a UNET architecture with ResNet34 backbone [16] is employed using the Keras framework. The overall network architecture is shown in Figure 4. This network consists of an encoder structure that downsamples the spatial resolution of the input image through convolutional operations, to obtain a low-resolution feature mapping. These features are then resampled by a decoding structure to obtain a pixel-wise prediction of the same size of the input image. The output of the network is a probability map that assigns to each pixel a probability of belonging to a specific class. The entire network is trained on a three-class problem, giving the 512 × 512 RGB images as input and the corresponding labeled masks as the target. In each image of the dataset, pixels are labeled in three classes: (i) background, (ii) blood vessel, and (iii) blood vessel boundaries. To solve the problem of class imbalance, our network's loss function is class-weighted by taking into account how frequently a class occurs in the training set. This means that the least-represented class will have a greater contribution than a more represented one during the weight update. The class weight is computed as follows:

$$f_{classX} = \sum_{i=1}^{N} \frac{\%\, pixel_{classX}}{N} \qquad x = 1,2,3 \tag{2}$$

$$class_{WEIGHT} = \frac{median([f_{class1},\ f_{class2},\ f_{class3}])}{[f_{class1},\ f_{class2},\ f_{class3}]} \tag{3}$$

where N is the total number of images and f_{classX} is the class frequency of generic class X.

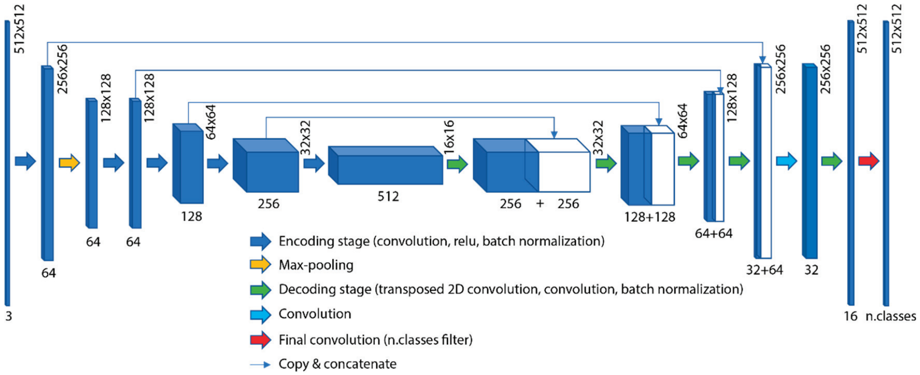


Figure 4. Architecture of the deep network employed to perform blood vessel detection. A UNET with ResNet34 backbone was implemented using Keras framework.

The encoding network was pre-trained on ILSVRC 2012 ImageNet [17]. During the training process, only the decoder weights were updated, while the encoder weights were set to non-trainable. This strategy allows for exploiting the knowledge acquired from a previous problem (ImageNet) and using the features learned to solve a new problem (vessel segmentation). This approach is useful both to speed up the training process and to create a robust model even using fewer data. The training data are real-time augmented while passing through the network, applying the same random transformations (rotation, shifting, flipping) both to the input image and to the corresponding encoded mask. Real-time data augmentation allows us to increase the amount of data available without storing the transformed data in memory. This strategy makes the model more robust to slight variations and prevents the network from overfitting.

Our network (Figure 4) was trained on 300 images with a mini-batch size of 32 and categorical cross-entropy as a loss function. The Adam optimization algorithm was employed with an initial learning rate of 0.01. The maximum number of epochs was set to 50, with a validation patience of 10 epochs for early stopping of the training process.

To preserve the information near the boundaries of the image, the RENFAST algorithm applies a specific procedure to build the CNN softmax. Briefly, a mirror border is synthesized in each direction and a sliding window approach is employed to build the probability map. To give the reader the opportunity to observe the entire procedure, we added a detailed description along with a summary figure in Appendix A.

2.4. Blood Vessel Detection

Starting from the normalized RGB image (Figure 5a), the RENFAST algorithm applies the deep network described in the previous section. Figure 5b shows the probability map obtained from the CNN, in which the red and green areas represent the pixels inside and on the edge of the blood vessels, respectively. Then, our method detects all the white and nuclear regions within the image. All the unstained structures are segmented by thresholding the grayscale image of the PAS sample, while cell nuclei are detected using the object-based thresholding developed in our previous work [15]. Figure 5c illustrates the segmentation of cellular structures performed by the RENFAST algorithm.

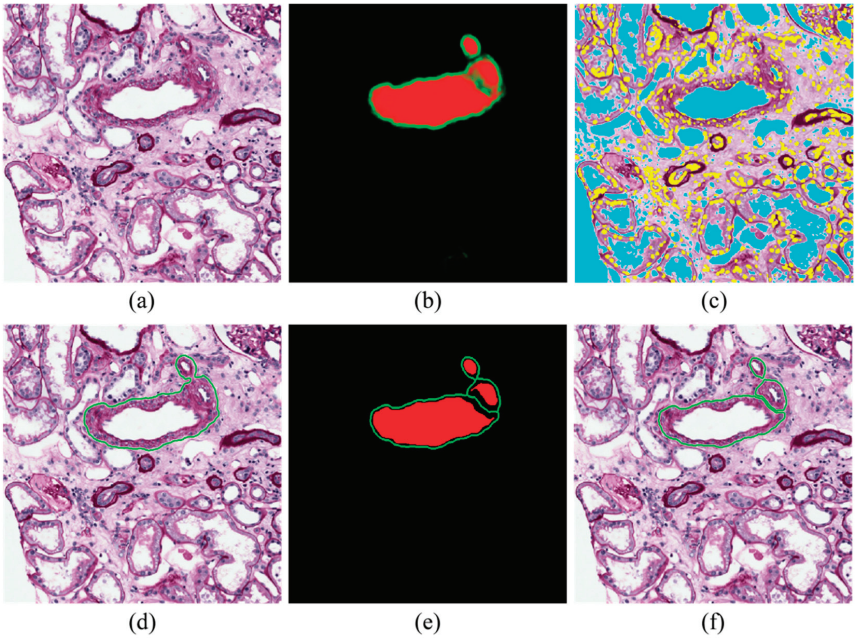


Figure 5. Steps performed by RENFAST for blood vessel detection. (**a**) Normalized image; (**b**) CNN probability map; (**c**) Cellular structure detection (yellow: nuclei, cyan: lumen); (**d**) Initial blood vessel segmentation; (**e**) Softmax with high SNR (signal-to-noise ratio); (**f**) Final blood vessel segmentation.

To obtain initial detection of the vascular structures, the probability maps of the regions inside and on the border of the blood vessels are added together and thresholded with a fixed value of 0.35. Then, morphological closing with a disk of 3-pixel radius (equal to 2.80 µm) is carried out to obtain smoother contours. As can be seen from Figure 5d, this strategy leads to accurate detection of the blood vessel boundaries but does not allow the separation of touching structures. To overcome this problem, an additional processing stage is performed to divide clustered blood vessels. The RENFAST algorithm employs a four-step procedure to increase the contrast between each blood vessel's boundary and the background:

1. Inner region mask: thresholding (0.35) and level-set on the probability map of inner regions (red layer);
2. Boundary mask: thresholding (0.35) and level-set on the probability map of boundary regions (green layer);
3. New red layer of the softmax: subtraction of the boundary mask from the inner region mask;
4. New green layer of the softmax: skeleton of the boundary mask.

This procedure generates a softmax with a high SNR (signal-to-noise ratio) where the border of each blood vessel is clearly defined (Figure 5e). Finally, for each connected component of the initial mask (Figure 5d), a simple check is performed: if by subtracting the green layer of the high-SNR softmax (Figure 5e), more than one region is generated, these regions are dilated by 1 pixel and added to the final mask. In this way, the thickness lost during the subtraction is recovered while maintaining the blood vessels' separation. Otherwise, if no additional structure is created with the subtraction, the connected component is inserted directly into the final mask.

The last step of the RENFAST algorithm for vessel segmentation is a structural check on the segmented objects: All the regions with an area less than 180 µm^2 are erased as they are too small to be considered blood vessels. In addition, objects must have at least 2.5% and 5% of the area occupied by lumen and nuclei, respectively. With these structural checks, most of the false positives generated by the CNN are deleted. The final result provided by the proposed algorithm is shown in Figure 5f.

2.5. Fibrosis Segmentation

The RENFAST algorithm is also able to quantify interstitial fibrosis in TRIC images. After stain normalization (Section 2.2), our method detects all the uncolored regions to process only TRIC stained structures. The normalized TRIC image is first converted to grayscale and Weiner filtered. The resulting image is then thresholded using a value equal to 90% of the image maximum (Figure 6a). Since fibrosis is characterized by a greenish color, the proposed algorithm applies an adaptive stain separation as described in [15]. Thanks to the stain separation (Figure 6b), it is possible to divide the regions that may manifest fibrosis (green channel) from the structural component (red channel). Segmentation of these two channels is performed using an improved version of the MANA (Multiscale Adaptive Nuclei Analysis) algorithm [18]. After min-max scaling, custom object-based thresholding is applied to the green channel (fibrosis) and red channel obtained in the previous step. For each possible threshold point $T \in [0, 1]$, the RENFAST algorithm computes the following energy function:

$$E(T) = p_0^2 \cdot var_0 \cdot log(var_0) + p_1^2 \cdot var_1 \cdot log(var_1) \tag{4}$$

where p_0 is the probability of having intensity values lower than T, p_1 is evaluated as $1 - p_0$, while var_0 and var_1 represent the variances of the probability functions of the two classes p_0 and p_1. The threshold T associated with the maximum of the energy function E represents the optimal thresholding point. The result of green and red channel segmentation is illustrated in Figure 6c. All remaining pixels not associated with one of the binary masks (white, green, red) are included in the green or red mask based on where they have the highest intensity in the stain separation channel.

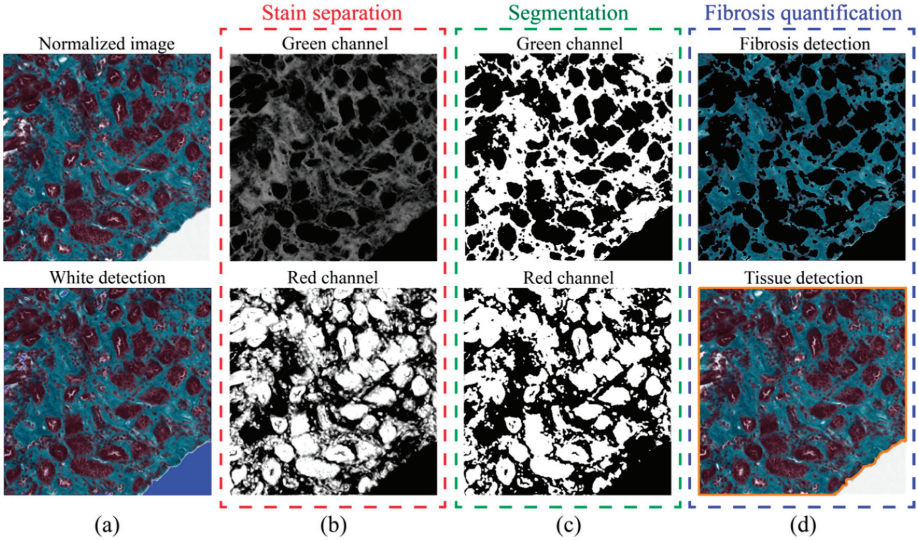


Figure 6. Steps performed by RENFAST for fibrosis segmentation. (**a**) Normalized image and white detection (in blue); (**b**) Stain separation between green and red channels; (**c**) Segmentation of green and red channels; (**d**) Fibrosis and tissue detection for interstitial fibrosis quantification.

Finally, the RENFAST algorithm quantifies the interstitial fibrosis as the ratio between the fibrotic area (segmented green channel) and the overall tissue area. Tissue detection is performed using an

RGB high-pass filter [19] where the RGB color of each pixel is treated as a 3D vector. The strength of the edge is defined as the magnitude of the maximum gradient. The raw tissue mask is generated by choosing a threshold equal to 5% of the maximum gradient. Morphological opening with a disk of 4-µm radius is then carried out to obtain the tissue contour (Figure 6d).

2.6. Performance Metrics

A comparison between manual and automatic masks was carried out to assess RENFAST's performance in the segmentation of kidney blood vessels and fibrosis. Manual annotations of blood vessels were generated using a custom graphical user interface based on MATLAB. Since fibrosis segmentation can be a long and demanding task, we designed a semi-automatic pipeline to help the pathologist during the generation of the manual mask (Appendix B). Several pixel-based metrics, such as balanced accuracy, precision, recall, and $\mathrm{F1}_{\mathrm{SCORE}}$, were evaluated for both blood vessel and fibrosis segmentation. Balanced accuracy ($\mathrm{Bal}_{\mathrm{ACCURACY}}$) is a common metric used in segmentation problems to deal with imbalanced datasets (TP vs. TN). $\mathrm{Bal}_{\mathrm{ACCURACY}}$ is calculated as the average of the correct predictions of each class individually. Precision is employed to evaluate the false detection of ghost shapes; recall quantifies the missed detection of ground truth objects; and finally, the $\mathrm{F1}_{\mathrm{SCORE}}$ is defined as the harmonic mean between precision and recall.

Accurate segmentation of blood vessel borders is fundamental for a correct evaluation of vascular damage. For this reason, we also evaluated the Dice coefficient (DSC) and the Hausdorff distance for all the true-positive vascular structures. Specifically, we computed the 95th percentile Hausdorff distance (HD95), which is defined as the maximum distance of a set (manual boundary) to the nearest point in the other set (automatic boundary). This metric is more robust towards a very small subset of outliers because it is based on the calculation of the 95th percentile of distances. During fibrosis assessment, the pathologist computes the ratio between fibrotic tissue and the whole tissue area. For each image, the absolute error (AE) between manual and automatic estimation was calculated as

$$AE = \left| \left(\frac{fibrosis_{AREA}}{tissue_{AREA}}\right)_{MANUAL} - \left(\frac{fibrosis_{AREA}}{tissue_{AREA}}\right)_{RENFAST} \right| \tag{5}$$

where $(\cdot)_{MANUAL}$ and $(\cdot)_{RENFAST}$ denote the manual and the automatic annotations, respectively.

3. Results

The automatic results provided by the RENFAST method are compared herein both with manual annotations and with previously published works. For blood vessel segmentation, we compared our algorithm with the one proposed by Bevilacqua et al. [8], while we used the methods published by Tey et al. [10] and Fu et al. [11] as benchmarks for interstitial fibrosis segmentation. As datasets and manual annotations of these works are not publicly available, all the described methods were applied to the same dataset used in this paper. The processing was performed on a custom workstation with a 3.5 GHz 10-core CPU with 64 Gb of RAM (Turin, Italy).

3.1. Blood Vessel Detection

Both pixel-based metrics ($\mathrm{Bal}_{\mathrm{ACCURACY}}$, precision, recall, $\mathrm{F1}_{\mathrm{SCORE}}$) and object-based metrics (DSC, HD95) were calculated to assess the performance of the RENFAST algorithm. To demonstrate the superiority of our strategy, we also evaluated the results obtained using a simple two-class CNN (background vs. vessel) and a three-class CNN without our post-processing. Tables 2 and 3 summarize the metrics calculated for blood vessel detection.