







Article

# A Deep Learning Instance Segmentation Approach for Global Glomerulosclerosis Assessment in Donor Kidney Biopsies

Nicola Altini <sup>1</sup>, Giacomo Donato Cascarano <sup>1</sup>, Antonio Brunetti <sup>1</sup>, Irio De Feudis <sup>1</sup>, Domenico Buongiorno <sup>1</sup>, Michele Rossini <sup>2</sup>, Francesco Pesce <sup>2</sup>, Loreto Gesualdo <sup>2</sup> and Vitoantonio Bevilacqua <sup>1,\*</sup>

- <sup>1</sup> 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.); irio.defeudis@poliba.it (I.D.F.); domenico.buongiorno@poliba.it (D.B.)
- <sup>2</sup> Nephrology Unit, Department of Emergency and Organ Transplantation (DETO), University of Bari Aldo Moro, 70126 Bari, Italy; michelerosini@libero.it (M.R.); francesco.pesce12@alumni.imperial.ac.uk (F.P.); loreto.gesualdo@uniba.it (L.G.)
- \* Correspondence: vitoantonio.bevilacqua@poliba.it

Received: 23 September 2020; Accepted: 21 October 2020; Published: 25 October 2020



**Abstract:** The histological assessment of glomeruli is fundamental for determining if a kidney is suitable for transplantation. The Karpinski score is essential to evaluate the need for a single or dual kidney transplant and includes the ratio between the number of sclerotic glomeruli and the overall number of glomeruli in a kidney section. The manual evaluation of kidney biopsies performed by pathologists is time-consuming and error-prone, so an automatic framework to delineate all the glomeruli present in a kidney section can be very useful. Our experiments have been conducted on a dataset provided by the Department of Emergency and Organ Transplantations (DETO) of Bari University Hospital. This dataset is composed of 26 kidney biopsies coming from 19 donors. The rise of Convolutional Neural Networks (CNNs) has led to a realm of methods which are widely applied in Medical Imaging. Deep learning techniques are also very promising for the segmentation of glomeruli, with a variety of existing approaches. Many methods only focus on semantic segmentation—which consists in segmentation of individual pixels—or ignore the problem of discriminating between non-sclerotic and sclerotic glomeruli, so these approaches are not optimal or inadequate for transplantation assessment. In this work, we employed an end-to-end fully automatic approach based on Mask R-CNN for instance segmentation and classification of glomeruli. We also compared the results obtained with a baseline based on Faster R-CNN, which only allows detection at bounding boxes level. With respect to the existing literature, we improved the Mask R-CNN approach in sliding window contexts, by employing a variant of the Non-Maximum Suppression (NMS) algorithm, which we called Non-Maximum-Area Suppression (NMAS). The obtained results are very promising, leading to improvements over existing literature. The baseline Faster R-CNN-based approach obtained an F-Measure of 0.904 and 0.667 for non-sclerotic and sclerotic glomeruli, respectively. The Mask R-CNN approach has a significant improvement over the baseline, obtaining an F-Measure of 0.925 and 0.777 for non-sclerotic and sclerotic glomeruli, respectively. The proposed method is very promising for the instance segmentation and classification of glomeruli, and allows to make a robust evaluation of global glomerulosclerosis. We also compared Karpinski score obtained with our algorithm to that obtained with pathologists' annotations to show the soundness of the proposed workflow from a clinical point of view.

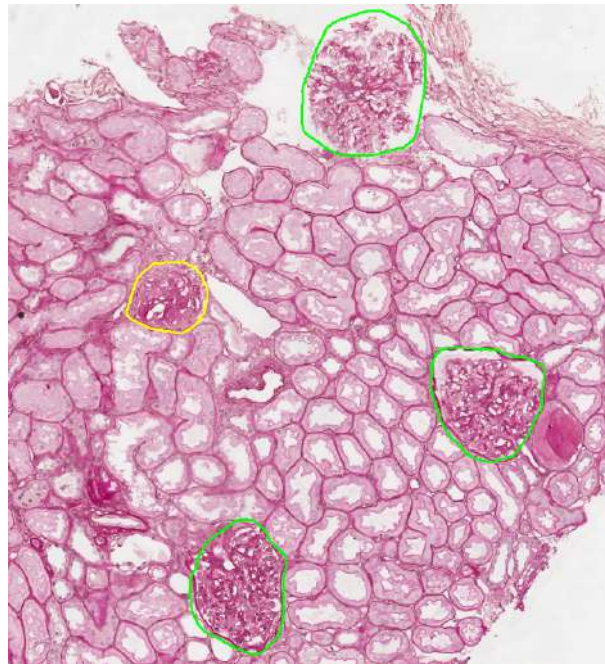
**Keywords:** glomerulosclerosis; kidney transplantation; Faster R-CNN; Mask R-CNN; object detection; instance segmentation

---

## 1. Introduction

In order to evaluate if a kidney is eligible for transplantation, a key step is the histological assessment of renal biopsies by expert pathologists. The determination, by a pathologist, of the number of globally sclerosed glomeruli with respect to the total number of glomeruli is a fundamental criteria for accepting or discarding donor kidneys. Considering the shortage of organs suitable for transplantation, the possibility to have an automatic system for a rapid and effective evaluation of global glomerulosclerosis would be very important, permitting to retain the largest quantity of eligible kidneys. In this paper, we propose a Computer-Aided Diagnosis (CAD) system which has the purpose to support the expert pathologists in the glomerular detection and classification task, allowing them to easily obtain global glomerulosclerosis information. Automated systems proved to be useful in a variety of medical applications, including biometrical analysis for personal identification [1], cancer system biology [2], blood parameters evaluation [3], breast cancer classification [4], diagnosis of neurological disorders [5], analysis of nasal cytology [6], segmentation and investigation of the conjunctiva [7,8], and prediction of the on-target cleavage efficiency from sgRNA sequences [9]. The rise of Convolutional Neural Networks (CNNs) opened many opportunities for Computer Vision tasks like object detection, semantic segmentation, and instance segmentation. This has led to a large development of deep learning methods and techniques in these tasks, which cannot be extensively detailed here. A comprehensive review on object detection and instance segmentation approaches can be found in [10], whereas one for semantic segmentation is [11]. In the realm of Digital Pathology, several recent studies have employed CNNs for glomerulus identification in renal biopsies [12–23]. Glomerulus detection has been approached as object detection task (e.g., [13]) or as semantic segmentation task (e.g., [17,22]). In this paper, we treat it like an instance segmentation task (e.g., [23]). CNN and medical imaging techniques have proven to be useful for evaluation of eligibility of donor kidneys [14,15,17,22,23].

A fundamental quantitative measure for assessing the eligibility for transplantation of kidneys from expanded criteria donors (ECD) is the Karpinski score [24]. Glomerular, tubular, interstitial, and vascular compartments are evaluated from an histological point of view. Then, for each of these compartments, it is assigned a score in the range 0 to 3, where 0 corresponds to normal histology and 3 to the worst degree of, respectively, global glomerulosclerosis, tubular atrophy, interstitial fibrosis, and arterial and arteriolar narrowing [24,25]. The identification of all non-sclerotic and sclerotic glomeruli in the kidney biopsy is the preliminary task required to define a score for global glomerulosclerosis. Non-sclerotic glomeruli tend to have an elliptic shape. They are characterized by the Bowman's capsule and by the capillary tuft with the mesangium. The latter is sited inside the glomerulus, whereas the first is peripheral and contains the tuft. There is a space between these two elements, which is known as Bowman's space. The capillary tufts features nuclei of cells (blue points), capillary lumens (white areas), and the mesangial matrix (regions with similar tonality and different levels of saturation), so it resembles a pomegranate. A (globally) sclerotic glomerulus is characterized by capillary lumens which are obliterated for an increase in extracellular matrix, and collagenous material which completely fills the Bowman's space. Examples of non-sclerotic and sclerotic glomeruli are depicted in Figure 1.



**Figure 1.** Glomeruli. Green: non-sclerotic glomeruli. Yellow: sclerotic glomeruli. Ground truth annotations provided by pathologists.

In this paper, we propose a deep learning framework, based on Mask R-CNN [26], for glomerular detection and classification with an end-to-end instance segmentation approach. Semantic segmentation networks can guarantee very high pixel-level results, but they may perform worse in the object detection task, if compared to specialized architectures [15]. The key points of the proposed method are:

- the possibility to train an end-to-end instance segmentation neural network, by exploiting Mask R-CNN, strongly reducing the need of post processing operations and allowing to learn all the required features in a unified process;
- the use of a variant of the standard Non-Maximum Suppression (NMS) algorithm, which we called Non-Maximum-Area Suppression (NMAS) that led to an improvement of the performances in our sliding window approach. Note that NMAS, like NMS, is a general purpose algorithm and can be useful also for other detection tasks;
- it shows superior performances to other alternatives proposed in literature, without computational drawbacks. Alternatives include object detection approaches, as Faster R-CNN (adopted in [13]), which is herein used as baseline, and semantic segmentation approaches (adopted in [15,17]).

## 2. Methods and Materials

### 2.1. Dataset

All the experiments conducted in this paper exploited a dataset provided by the Department of Emergency and Organ Transplantations (DETO) of Bari University Hospital. This dataset is composed of 26 kidney biopsies coming from 19 donors. Kidney donors sections contain 2344 non-sclerotic glomeruli and 428 sclerotic glomeruli [15]. The dataset has been split into a train-validation set composed of 19 biopsies and a test set composed of 7 biopsies. The train-validation set has been exploited for model fitting and hyperparameters tuning, whereas the final estimation of the results has been computed on the test set. The whole train-validation set contains 1852 non-sclerotic glomeruli and 341 sclerotic glomeruli; the test set contains 492 non-sclerotic glomeruli and 87 sclerotic glomeruli.

## 2.2. Object Detection with Deep Learning

Deep learning refers to the adoption of architectural processing models, composed by different layers, at the purpose of learning structured representation of the input data. The role of deep learning has been pivotal in different sectors, including visual object recognition and object detection [27]. Starting from the breakthrough obtained by AlexNet [28], CNNs have become widely used for almost every kind of computer vision problem. In this work, we will focus on CNNs for object detection, a problem which consists of finding the bounding boxes for all the objects of interest present in the image, and for instance segmentation, in which it is also required to delineate precise masks for the objects.

Among the CNN-based methods for object detection, a particular mention is devoted to the Region-Based Convolutional Neural Networks (R-CNN) family of models. The original R-CNN was proposed in 2014 by R. Girshick et al. from UC Berkeley [29]. The method fuses region proposals with CNNs at the purpose of performing object detection. The first part of the R-CNN algorithm is devoted in generating region proposals which are category-agnostic and that may contain objects. Then, those regions are fed to a CNN which extracts a vector representing the features for each region. Finally, the feature vector is given as input to a set of class-specific linear support vector machines (SVMs).

In 2015, R. Girshick improved the R-CNN method, creating a new object detection network named Fast R-CNN [30]. In Fast R-CNN, the whole input image is fed to the CNN to generate convolutional feature maps. Then, region proposals are discovered from the convolutional feature maps, and are warped into squares. An RoI-pooling layer (RoI stands for Region of Interest) is adopted to reshape the proposals to a fixed size, so that they can be forwarded to fully connected layers. The inclusion of region proposals based on selective search causes performance issues in Fast R-CNN.

This concern was solved in 2016 with a further evolution of the R-CNN architecture, Faster R-CNN, proposed by S. Ren, K. He, R. Girshick, and J. Sun [31]. The team of Microsoft Research discovered that feature maps computed in the first part of Fast R-CNN can be used to generate region proposals instead of slower and not-learnable algorithms as selective search. The big evolution in Faster R-CNN is the introduction of a Region Proposal Network (RPN) after the feature maps extraction of Fast R-CNN. RPN exploits a novel concept, namely anchor boxes, instead of previous architectures which adopted pyramids of images or pyramids or filters. In order to generate anchor boxes, it is possible to employ a small network which input is an  $n \times n$  spatial window of the feature map; the resulting anchor boxes are a collection of the rectangular bounding boxes proposals, with the related scores. The scale and aspect ratio of anchor boxes are parameters that can be decided from the architecture designer. In order to identify objects at different resolutions, it is required to make use of anchor boxes with different shapes.

A further improvement from the R-CNN family of detectors is Mask R-CNN, developed by a team of Facebook AI Research (FAIR) in 2017 [26]. Mask R-CNN allows to solve instance segmentation tasks, whereas Faster R-CNN and previous approaches were only able to perform object detection. The overall Mask R-CNN architecture is composed by two parts: the backbone architecture, which performs feature extraction, and the head architecture, which performs classification, bounding box regression and mask prediction.

## 2.3. Object Detection Definitions and Metrics

Reference metrics used for evaluating object detection models are based on object detection challenges as PASCAL VOC (<http://host.robots.ox.ac.uk/pascal/VOC/>), Google Open Images (<https://opensource.google/projects/open-images-dataset>), and COCO (<https://cocodataset.org/>). In general, the performance metrics used in these challenges offer a global level evaluation, estimating the performances of the model in the whole dataset. The adoption of global metrics makes benchmarking much simpler, but it does not provide insights on how and why the mistakes have been made.

In order to define object detection metrics, we have to outline what we intend with a detection first. For this purpose, we introduce Intersection over Union (*IoU*) and Intersection over Minimum (*IoM*) Given two bounding boxes  $A$  and  $B$  we can define the *IoU* as the ratio between the intersection of their areas and the union of their areas:

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

In (1),  $|\cdot|$  denotes the set cardinality operator. *IoU* values lie in the range  $[0, 1]$ , where 1 indicates a perfect match. We say that a predicted object matches with a ground truth object when *IoU* between them is above a certain threshold (a common choice for the threshold is 0.5). Another concept related to *IoU* is *IoM*, which can be quite useful for defining detections in post processing algorithms. The *IoM* between two bounding boxes  $A$  and  $B$  is the ratio between the intersection of their areas and the minimum of their areas:

$$IoM(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)} \quad (2)$$

*IoM* values lie in the range  $[0, 1]$ , where 1 indicates a perfect match. Note that in these definitions we referred to bounding boxes, but *IoU* and *IoM* can be calculated between any finite sample sets. Widespread evaluation metrics are Average Precision (*AP*), which can be mainly defined as the area under the precision–recall curve, and mean Average Precision (*mAP*), that is *AP* averaged over all classes. A naive implementation of *AP* is described by the following equation:

$$AP = \int_0^1 p(r) dr \quad (3)$$

Anyway, we have to note that *AP* is usually calculated (e.g., PASCAL VOC) by adopting the average interpolated precision value of the positive examples [32]. We can explicate the dependence of precision and recall from confidence  $c$  using the notation  $p = P(c)$  and  $r = R(c)$ . Recall  $R(c)$  is the fraction of objects detected with confidence of at least  $c$ . Precision  $P(c)$  is the fraction of detections that are correct:

$$P(c) = \frac{R(c) \cdot N_j}{R(c) \cdot N_j + F(c)} \quad (4)$$

In (4),  $N_j$  is the number of objects in class  $j$  and  $F(c)$  is the number of incorrect detections with at least confidence  $c$ .

Mean Average Precision (*mAP*) for  $K$  classes can be calculated as reported in (5):

$$mAP = \frac{1}{K} \sum_{j=1}^K AP_j \quad (5)$$

#### 2.4. Non-Maximum Suppression

The NMS algorithm is a fundamental post-processing step for object detection when it is required to remove overlapped bounding boxes for avoiding duplicate detections. Object detection and instance segmentation architectures from the R-CNN family discussed before adopt NMS to reduce the number of proposals, since many of them are overlapped. NMS is reported in Algorithm 1. Different improvements of the NMS algorithm have been proposed, as Soft-NMS by N. Bodla et al. [33]. In NMS, we pick the detection box  $B$  with the maximum score, and then we suppress all other detection boxes that overlap more than a predefined threshold. We continue with this procedure in a recursive way until all boxes have been processed. The NMS algorithm is designed so that objects lying within the predefined overlap threshold lead to misses. Soft-NMS attempts to solve this problem by decaying the detection scores of all other objects as a continuous function of their overlap with  $B$ . Therefore, no object is discarded in this procedure.

**Algorithm 1:** Non-Maximum Suppression (NMS) [33].

---

**input :**  $B_i = b_1, \dots, b_{N_i}$ , the  $N_i$  initial detections  
 $b_j = (x_j, y_j, w_j, h_j), j = 1, \dots, N_i$   
 $S_i = s_{i_1}, \dots, s_{i_{N_i}}$ , the  $N_i$  initial scores  
 $T_{iou}$ , the NMS threshold on  $IoU$

**output:**  $B_o = b_1, \dots, b_{N_o}$ , the  $N_o \leq N_i$  final detections  
 $S_o = s_{o_1}, \dots, s_{o_{N_o}}$ , the  $N_o \leq N_i$  final scores

- 1  $B_o = \{\}$ ;
- 2  $S_o = S_i$ ;
- 3 **while**  $B_i$  is not empty **do**
- 4      $m = \text{argmax}(S_o)$ ;
- 5      $B_o = B_o \cup \{b_m\}$ ;
- 6      $B_i = B_i \setminus \{b_m\}$ ;
- 7     **while**  $b_j \in B_i$  **do**
- 8         **if**  $iou(b_m, b_j) \geq T_{iou}$  **then**
- 9              $B_i = B_i \setminus \{b_j\}$ ;
- 10             $S_o = S_o \setminus \{s_j\}$ ;
- 11         **end**
- 12     **end**
- 13 **end**

---

NMS can be used also when applying object detectors in a sliding window fashion, to remove duplicate detections at the boundaries of adjacent windows. Anyway, both NMS and Soft-NMS suffer from the problem of not considering the area of the detected objects. This means that, if for an object there are two detected bounding boxes, one inside the other, the algorithm can choose the smaller box even if it has only a very slightly higher confidence score. In this paper, we define an algorithm, similar to NMS, but better suited for the purpose of handling overlapped bounding boxes in sliding window approaches. We called it NMAS, since it is a modification of NMS which considers also the area of the bounding box and not only its confidence  $s_j$ . We introduced a new parameter  $f_j = w_j h_j s_j^2$ , which incorporates also the area of the bounding box ( $w_j h_j$ ) and the square of the confidence ( $s_j^2$ ). Since  $s_j$  falls in the range from 0 to 1, we used the square of the confidence to penalize lower values. NMAS is reported in Algorithm 2. Another improvement of NMAS is the usage of  $IoM$  together with  $IoU$  to detect overlapping boxes.  $IoM$  easily allows to recognize bounding boxes mainly contained in other ones, a common case in overlapping sliding window approaches.

**Algorithm 2:** Non-Maximum-Area Suppression (NMAS).

---

**input :**  $B_i = b_1, \dots, b_{N_i}$ , the  $N_i$  initial detections  
 $b_j = (x_j, y_j, w_j, h_j), j = 1, \dots, N_i$   
 $S_i = s_{i_1}, \dots, s_{i_{N_i}}$ , the  $N_i$  initial scores  
 $T_{iou}$ , the NMAS threshold on  $Iou$   
 $T_{iom}$ , the NMAS threshold on  $Iom$

**output:**  $B_o = b_1, \dots, b_{N_o}$ , the  $N_o \leq N_i$  final detections  
 $S_o = s_{o_1}, \dots, s_{o_{N_o}}$ , the  $N_o \leq N_i$  final scores

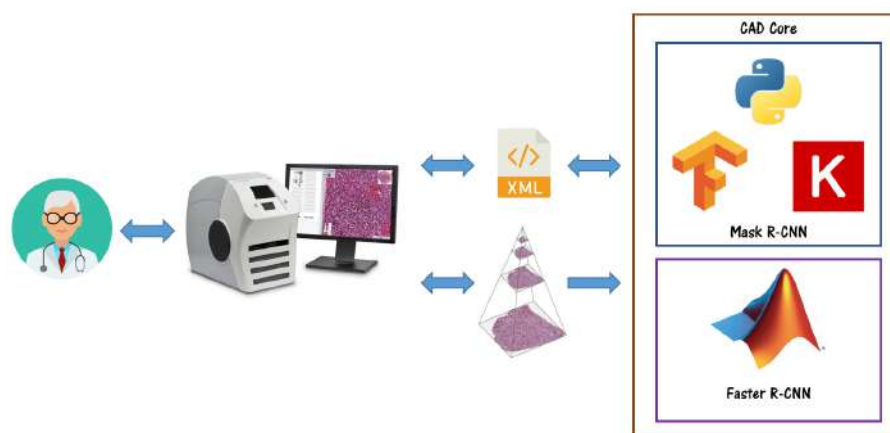
- 1  $B_o = \{\}$ ;
- 2  $S_o = (w_1 h_1 s_{i_1}^2, w_2 h_2 s_{i_2}^2, \dots, w_{N_i} h_{N_i} s_{i_{N_i}}^2)$ ;
- 3 **while**  $B_i$  is not empty **do**
- 4      $m = \text{argmax}(S_o)$ ;
- 5      $B_o = B_o \cup \{b_m\}$ ;
- 6      $B_i = B_i \setminus \{b_m\}$ ;
- 7     **while**  $b_j \in B_i$  **do**
- 8         **if**  $iou(b_m, b_j) \geq T_{iou} \vee iom(b_m, b_j) \geq T_{iom}$  **then**
- 9              $B_i = B_i \setminus \{b_j\}$ ;
- 10             $S_o = S_o \setminus \{s_j\}$ ;
- 11         **end**
- 12     **end**
- 13 **end**

---

## 2.5. Workflow

The methods and algorithms proposed for Faster R-CNN are based on MATLAB R2019a, whereas a Python implementation, with TensorFlow and Keras libraries, developed by Waleed Abdulla from Matterport, Inc. (Sunnyvale, CA, USA) [34] and published under the MIT License (<https://opensource.org/licenses/MIT>), has been exploited for Mask R-CNN.

A high-level overview of the proposed CAD system is depicted in Figure 2.



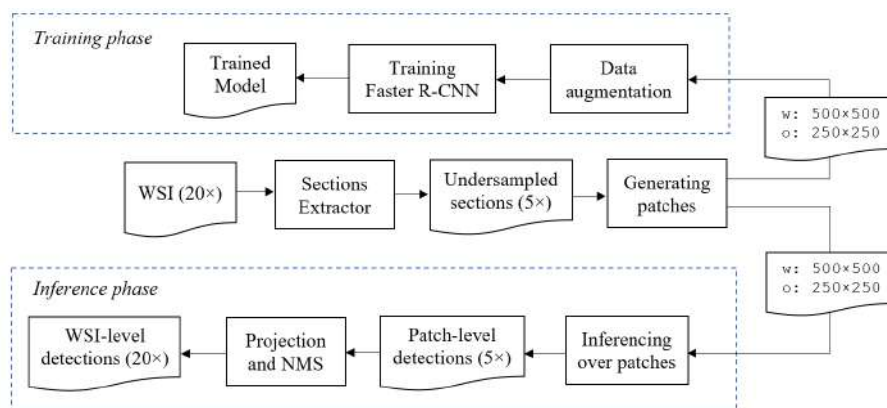
**Figure 2.** Computer-Aided Diagnosis (CAD) System overview. The pathologist can visualize network results and annotate images through ImageScope, exploiting the XML interface implemented by our CAD.

The pathologists can visualize and annotate whole slide images (WSIs) using Aperio ImageScope. An XML interface has been implemented for both the MATLAB and Python environments. This allows to create the training set and also to make the network predictions available to the clinicians, with a very

smooth integration. To accomplish the task of calculating the Karpinski histological score, we have to make a careful choice for the architecture of the network. In this work, we compare an object detection framework with an instance segmentation one. For a semantic segmentation approach, consider our previous work [15]. All the models have been trained and validated on the same machine of [15]. We used a dual boot system; the MATLAB implementation has been tested on Windows, whereas Ubuntu has been exploited for the Python implementation.

### 2.5.1. Faster R-CNN

The implemented baseline is based on Faster R-CNN, with the workflow depicted in Figure 3. Starting from a WSI, we segmented its sections using Section Extractor [15]; then we got kidney sections undersampled by a factor of 4. These undersampled biopsy sections are divided into patches of size  $500 \times 500$ , with stride of  $250 \times 250$ . The stride has been chosen to guarantee an overlap of  $250 \times 250$ , so that there is at least one patch in which each glomerulus is fully contained. Since the dimensions of glomeruli in images at full resolution ( $20\times$ ) are lesser than  $800 \times 800$ , at undersampled resolution ( $5\times$ ) they are lesser than  $200 \times 200$ , thus the claimed condition is easily obtained. In this way we did not discard any glomerulus from training data. Note that in inferencing phase we can apply again this procedure, reducing the eventuality of missing glomeruli. Dividing the original image into patches poses the problem on how the partially contained glomeruli should be considered in the training patch (compare Figures 4 and 5). At the purpose of solving this issue, a hyperparameter has been introduced, the *tolerance*, indicating the maximum allowed percentage of glomerulus size that can be out of patch to consider that glomerulus as positive example for training. For example, if we set *tolerance* = 0, then only glomeruli fully contained in each patch are considered as positive examples for training. This means that, even if a glomerulus is out of 1 px, it will not be used as positive example (again, compare Figures 4 and 5). Optimal values for this parameter are approximately in the range [0.2, 0.4]. These values have been found in empirical way. Our final proposed Faster R-CNN detector has been trained with *tolerance* = 0.3.



**Figure 3.** Baseline approach based on Faster Region-Based Convolutional Neural Network (R-CNN). The top part describes how to perform the training of the model, exploiting the train-validation set (19 whole slide images). The bottom part explains how to use the trained model for performing inference on the test set sections (7 whole slide images).



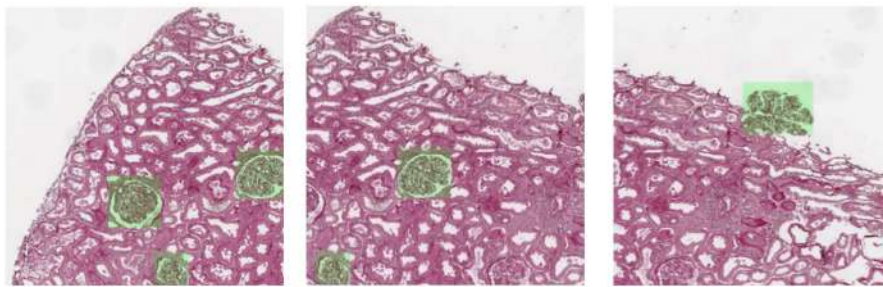


Figure 4. Patches with *tolerance* = 0.3.

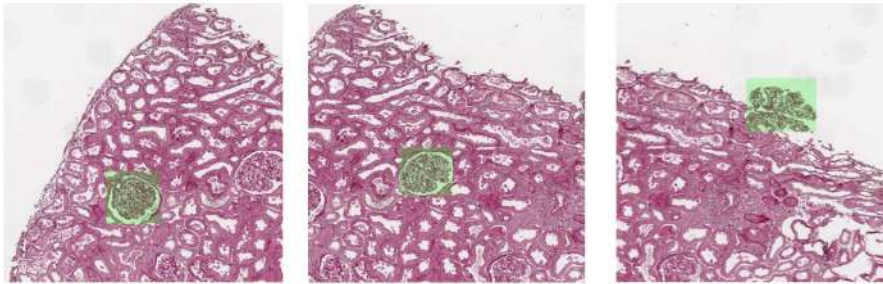


Figure 5. Patches with *tolerance* = 0.

Due to the small dataset sample size, composed of 26 WSIs which contain 101 sections, we exploited oversampling as data rebalancing methodology. In particular, for each training patch that has at least a sclerotic glomerulus inside (underrepresented class), we performed data augmentation by rotating this patch by  $90^\circ$ ,  $180^\circ$  and  $270^\circ$ . In this way, we roughly quadruplicated the number of sclerotic glomeruli (note that also the number of non-sclerotic glomeruli is increased by this operation).

Since our model has been trained on small patches (with size of  $500 \times 500$ ), it is not advisable to directly adopt it for performing inferences on images of full sections (up to  $2500 \times 2500$ ). Moreover, some sections can be too large to fit in memory. The proposed solution is straightforward: we divided in patches also the images used for inferences. Again, we used patches of  $500 \times 500$  with stride of  $250 \times 250$ , thus reducing the probability of a glomerulus miss (since, as stated before, we have at least a full glomerulus in each patch). The use of overlaid windows for patches posed the problem of overlapped detections in full image (when we projected patch-level detections on original image), as can be seen in Figure 6. For suppressing duplicated bounding boxes, we used two iterations of NMS (Algorithm 1): standard NMS and NMS with matches computed on *IoM* instead of *IoU*. We exploited MATLAB `selectStrongestBboxMulticlass` function <https://www.mathworks.com/help/vision/ref/selectstrongestbboxmulticlass.html>. The result of applying NMS with threshold for *IoU* set to 0.3 to bounding boxes in Figure 6 is depicted in Figure 7. Since in some cases there are small bounding boxes mainly contained inside larger ones, we performed also NMS on *IoM* with the threshold set to 0.5 (i.e., we performed NMS on all the boxes that overlaps with *IoM* greater or equal than 0.5). In the case of Figure 7, this step did not result in further suppression. Further details about hyperparameters configuration of Faster R-CNN approach can be found in Appendix A.1.

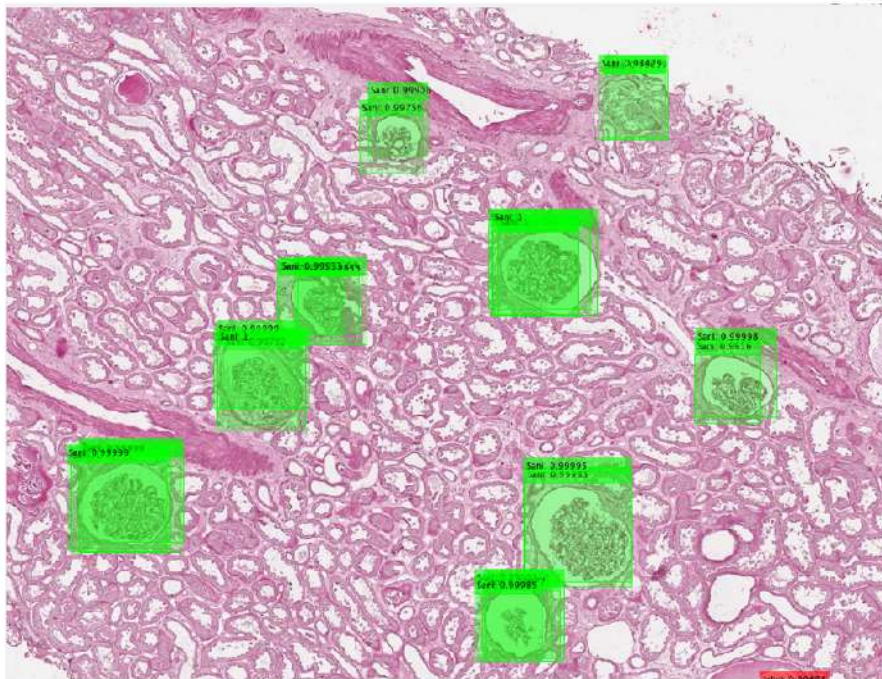


Figure 6. Overlapped bounding boxes after projection in full image.

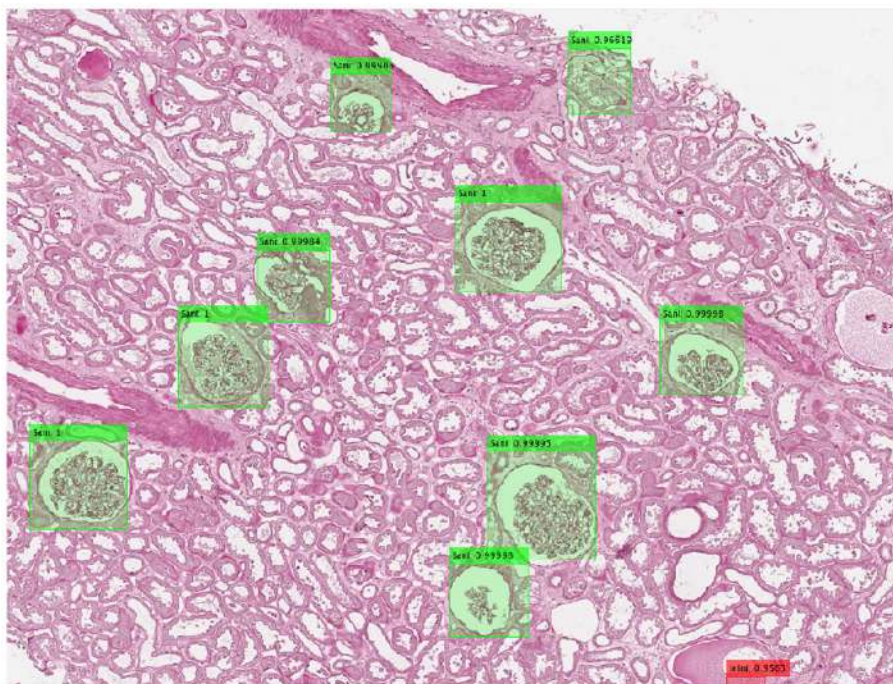
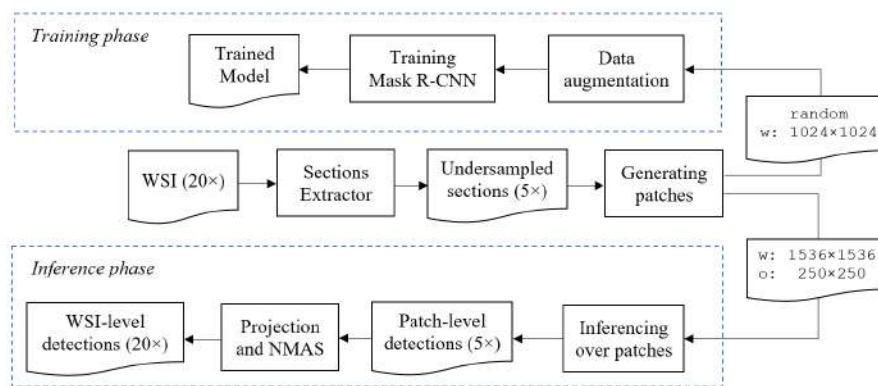


Figure 7. Detection on section after Non-Maximum Suppression based on Intersection over Union.

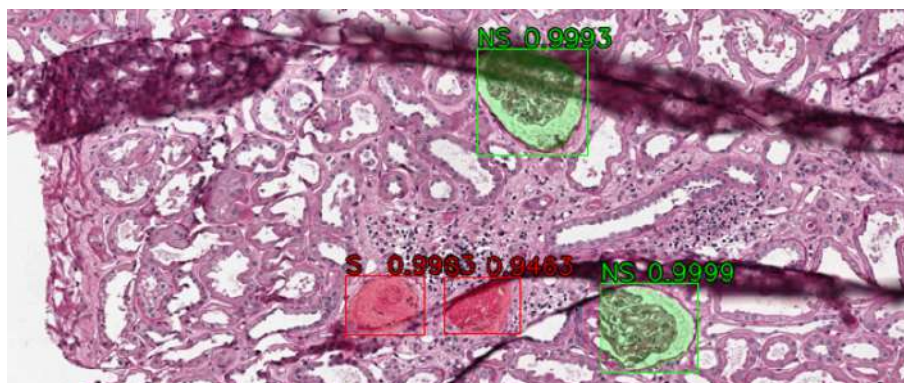
### 2.5.2. Mask R-CNN

The general schema that we used for the instance segmentation approach is depicted in Figure 8.

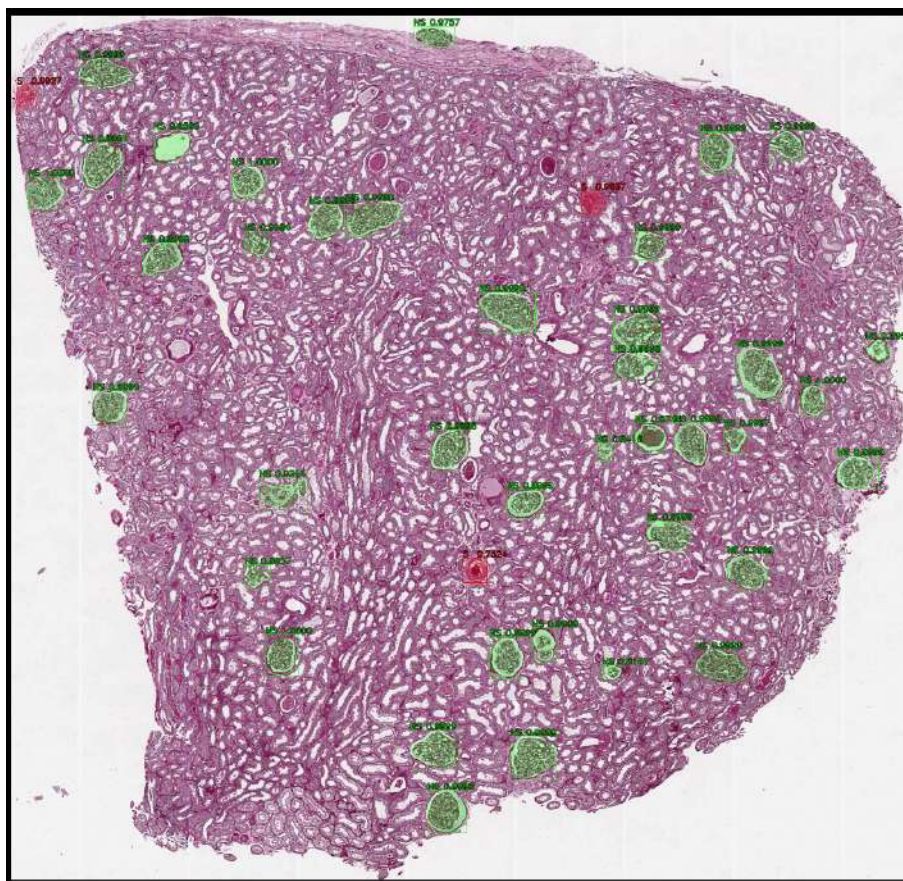


**Figure 8.** Proposed approach based on Mask R-CNN. The top part describes how to perform the training of the model, exploiting the train-validation set (19 whole slide images). The bottom part explains how to use the trained model for performing inference on the test set sections (7 whole slide images). Note that in the inference phase we take advantage of the proposed Non-Maximum-Area Suppression (NMAS) algorithm.

In the training phase we sampled from each section (obtained using the Sections Extractor algorithm already employed in [15]) random patches of  $1024 \times 1024$  pixels, then we performed random data augmentations on-the-fly, so that the network processes different data for each epoch. In the inferencing phase we used larger windows, since the memory requirements are less restrictive. We selected patches with size  $1536 \times 1536$ , with an overlap between adjacent patches of  $250 \times 250$ , for the same reason we explained in Faster R-CNN based detector. We performed zero padding for the missing information. When we project back the patch-level detections to WSI-level detections, we perform NMAS described in Algorithm 2, which results in an improvement over NMS. Examples of patch-level and WSI-level detections can be seen in Figures 9 and 10, respectively.

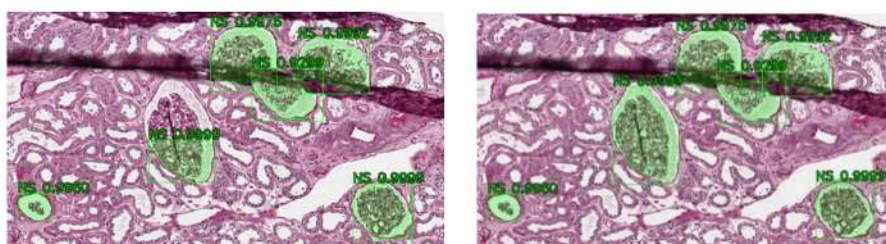


**Figure 9.** Patch-level detection with Mask R-CNN.



**Figure 10.** Whole slide image (WSI)-level detection with Mask R-CNN. Overlapping bounding boxes have been eliminated with the proposed Non-Maximum-Area Suppression algorithm.

Note that, compared with Faster R-CNN, we have also a mask besides the bounding box, since Mask R-CNN purpose is to solve instance segmentation task and not only object detection task. Using NMAS proved to be very useful in sliding window approaches. An example is depicted in Figure 11. We can see that using simple NMS, the chosen bounding box in one case is not the most suitable, since it does not overlay the whole glomerulus. NMAS solves this problem by considering also the areas of involved bounding boxes and not only their confidence scores.



**Figure 11.** Mask R-CNN predictions, after removal of overlapping bounding boxes with the two considered algorithms: Non-Maximum Suppression (**Left**) and Non-Maximum-Area Suppression (**Right**).

We used ResNet-50 as backbone, since it allows quality feature extraction but is lighter than ResNet-101 [35]. In the training process, we used a pretrained model on the COCO dataset. In order to exploit in the best way the pretraining, we trained only the network heads for the first 20 epochs. Then, for the subsequent 40 epochs, we fine-tuned ResNet stage 4 and layers above. For the last 40 epochs,

we trained all the layers of the network, and we lowered the learning rate to 0.0001. Further details about hyperparameters configuration of Mask R-CNN approach can be found in Appendix A.2.

### 3. Results

#### 3.1. Baseline: Faster R-CNN

With the Faster R-CNN-based approach, we get the results reported in Tables 1 and 2. The *mAP* for the Faster R-CNN approach is 0.803.

**Table 1.** Object detection confusion matrix with the baseline Faster R-CNN workflow. NS stands for non-sclerotic glomeruli, S stands for sclerotic glomeruli, and B for background.

		Prediction		
		NS	S	B
Ground Truth	NS	463	0	29
	S	7	61	19
	B	62	35	–

**Table 2.** Detection metrics with the baseline Faster R-CNN workflow.

Class	Recall	Precision	F-Score
Non-sclerotic	0.941	0.870	0.904
Sclerotic	0.701	0.635	0.667

#### 3.2. Mask R-CNN

The results obtained with the Mask R-CNN-based approach are reported in Tables 3 and 4. Using NMAS instead of NMS for suppressing overlapped bounding boxes leads to an improvement of *mAP* from 0.881 to 0.902, and of F-measure for non-sclerotic glomeruli from 0.917 to 0.925.

**Table 3.** Object detection confusion matrix with the proposed Mask R-CNN workflow. NS stands for non-sclerotic glomeruli, S stands for sclerotic glomeruli, and B for background.

		Prediction		
		NS	S	B
Ground Truth	NS	470	0	22
	S	8	61	18
	B	46	9	–

**Table 4.** Detection metrics with the proposed Mask R-CNN workflow.

Class	Recall	Precision	F-Score
Non-sclerotic	0.955	0.897	0.925
Sclerotic	0.701	0.871	0.777

#### 3.3. Karpinski Score Assessment

In order to assess the clinical validity of the obtained results, we compared the Karpinski score computed by our CNN with that of expert pathologists.

The comparison between the baseline Faster R-CNN and Mask R-CNN is shown in Table 5. Ratio refers to number of sclerosed glomeruli divided by the overall number of glomeruli:  $Ratio = \frac{S}{S+NS}$ . The corresponding Karpinski score for the glomerular compart is determined according to the

following: 0, if there are no globally sclerosed glomeruli; 1, if there is <20% global glomerulosclerosis; 2, if there is 20-50% global glomerulosclerosis; 3, if there is >50% global glomerulosclerosis [24]. We note that the Faster R-CNN approach makes five errors in assessing the Karpinski score: four times it gives a score of 1 instead of a score of 2; one time it gives a score of 2 instead of a score of 1. The Mask R-CNN approach makes only three errors in assessing the Karpinski score: one time it gives a score of a score of 0 instead of a score of 1; two times it gives a score of 1 instead of a score of 2.

**Table 5.** Karpinski score, results on hold-out test set. Comparison between Faster R-CNN, Mask R-CNN, and ground truth annotations. NS stands for non-sclerotic, S stands for sclerotic. Score belongs to the range [0–3].

Donor	Kidney	Section	Mask R-CNN				Faster R-CNN				Ground Truth			
			NS	S	Ratio	Score	NS	S	Ratio	Score	NS	S	Ratio	Score
1	Left	1	30	3	0.09	1	31	3	0.09	1	30	3	0.09	1
		2	30	2	0.06	1	32	2	0.06	1	30	2	0.06	1
		3	31	4	0.11	1	29	5	0.15	1	28	4	0.13	1
		4	29	5	0.15	1	31	5	0.14	1	25	4	0.14	1
		5	32	0	0.00	0	30	1	0.03	1	31	1	0.03	1
		6	31	1	0.03	1	35	3	0.08	1	31	1	0.03	1
2	Right	1	11	5	0.31	2	9	8	0.47	2	10	5	0.33	2
3	Right	1	41	1	0.02	1	40	8	0.17	1	38	2	0.05	1
	Left	1	39	3	0.07	1	38	3	0.07	1	41	4	0.09	1
4	Right	1	19	4	0.17	1	23	7	0.23	2	17	5	0.23	2
		2	26	3	0.10	1	29	4	0.12	1	25	3	0.11	1
		3	30	2	0.06	1	29	5	0.15	1	25	3	0.11	1
		4	29	5	0.15	1	28	9	0.24	2	25	5	0.17	1
5	Right	1	22	4	0.15	1	23	3	0.12	1	22	4	0.15	1
		2	30	5	0.14	1	27	3	0.10	1	28	5	0.15	1
6	Right	1	14	4	0.22	2	14	3	0.18	1	13	6	0.32	2
		2	14	4	0.22	2	13	3	0.19	1	13	6	0.32	2
		3	13	4	0.24	2	13	3	0.19	1	14	5	0.26	2
		4	14	2	0.13	1	12	1	0.08	1	12	2	0.14	1
		5	17	5	0.23	2	16	4	0.20	2	14	6	0.30	2
		6	19	4	0.17	1	20	4	0.17	1	17	10	0.37	2

#### 4. Discussion

Recent studies tried to accomplish glomerular detection in kidney biopsies, using a wealth of techniques, most of which based on deep learning. Nonetheless, many of these approaches did not consider the task of classifying between non-sclerotic glomeruli and sclerotic ones. A full comparison of our approach with the recent research works in the task of glomerular detection is in Table 6, extending the one proposed by Kawazoe et al. [13] to the glomerulosclerosis classification case when available. We note that our model performs well in the detection of non-sclerotic glomeruli, with very high recall and precision values, but metrics for sclerotic glomeruli suffer from a higher number of false negatives.

**Table 6.** Comparison with literature, extending the one proposed by Kawazoe et al. [13]. Stain acronyms: HE stands for Hematoxylin and Eosin, PAS stands for Periodic acid–Schiff, PAM stands for periodic acid-methenamine silver, D stands for Desmin, JS stands for Jones Silver, TRI stands for Gömöri’s Trichrome, CR stands for Congo Red, SR stands for Sirius Red, M1 stands for HE/PAS/JS/TRI/CR, M2 stands for HE/PAS/CD10/SR. Species (Sp) acronyms: H stands for human, R stands for rat, M stands for mouse. Method acronyms: R-HOG stands for Rectangle-Histogram of Oriented Gradients, S-HOG stands for Segmental-Histogram of Oriented Gradients, SVM stands for support vector machine, mrcLBP stands for multi-radial color local binary patterns, CNN stands for Convolutional Neural Network, R-CNN stands for Region-based Convolutional Neural Network, FCN stands for fully convolutional network. Class acronyms: A stands for all (no distinction between non-sclerotic and sclerotic glomeruli), NS stands for non-sclerotic glomeruli, and S stands for sclerotic glomeruli.

Author	Sp	Stain	WSIs	Method	Class	Performances		
						Recall	Precision	F-Measure
Kato et al. [19]	R	D	20	R-HOG + SVM	A	0.911	0.777	0.838
				S-HOG + SVM	A	0.897	0.874	0.866
Temerinac-Ott et al. [21]	H	M2	80	R-HOG + SVM	A	N/A	N/A	0.405–0.551
				CNN	A	N/A	N/A	0.522–0.716
Gallego et al. [16]	H	PAS	108	CNN	A	1.000	0.881	0.937
	M	HE	15		A	0.800	0.900	0.850
Simon et al. [20]	R	M1	25	mrcLBP + SVM	A	0.560–0.730	0.750–0.914	0.680–0.801
	H	PAS	25		A	0.761	0.917	0.832
		PAS	200		A	0.919	0.931	0.925
Kawazoe et al. [13]	H	PAM	200	Faster R-CNN	A	0.918	0.939	0.928
		MT	200		A	0.878	0.915	0.896
		Azan	200		A	0.849	0.904	0.876
Marsh et al. [17]	H	HE	48	FCN + BLOB	NS	0.885	0.813	0.848
					S	0.698	0.607	0.649
Altini et al. [15]	H	PAS	26	SegNet	A	0.855	0.832	0.843
					NS	0.886	0.834	0.859
					S	0.667	0.806	0.730
				DeepLab v3+	A	0.858	0.952	0.903
					NS	0.913	0.935	0.924
					S	0.471	0.976	0.636
Proposed	H	PAS	26	Faster R-CNN	A	0.917	0.846	0.880
					NS	0.941	0.870	0.904
				Mask R-CNN	S	0.701	0.635	0.667
					A	0.931	0.907	0.919
					NS	0.955	0.897	0.925
S	0.701	0.871	0.777					

From the tests performed in this paper, it is possible to observe that glomerular detection and classification tasks should be approached as an instance segmentation tasks. Even if object detection approaches can guarantee respectable results, they do not exploit the mask information in the dataset. Semantic segmentation approaches allow to obtain decent results too, but they are slightly worse than instance segmentation ones. Indeed, training a CNN which classifies at pixel-level in a detection task is a less powerful method. Difficulties that occur with semantic segmentation networks include presence of noisy points in the output and lack of distinction between touching objects. Semantic segmentation networks can principally exploit texture information but are less capable to understand concepts as shapes, thus working worse on detection task comparing to specialized architectures. Nonetheless, in [17], Marsh et al. used fully convolutional network (FCN) (together with BLOB detection as

post-processing of semantic segmentation network output) to measure global glomerulosclerosis from kidney biopsies. The proposed Mask R-CNN approach outperforms their FCN-based one, improving F-score for healthy glomeruli from 0.848 to 0.925 and F-score for sclerosed glomeruli from 0.649 to 0.777. An important reason for these better performances may lie in the choice of the better model, relying on an instance segmentation network instead of a semantic segmentation one. Anyway, it has to be noted that Marsh et al. dealt with HE stained biopsies, whereas the dataset adopted for our experiments is made up by Periodic acid–Schiff (PAS) stained biopsies, which can be a better staining for glomerular recognition tasks. Although, it has to be demonstrated that CNNs work consistently better on PAS compared to HE. It is also worth noting that in [17] the unbalancing ratio is less than ours, being 3.44:1 compared to 5.48 : 1, thus allowing a smoother training process for the underrepresented class. Other works do not address the task of determining glomerulosclerosis, but focus only on glomerular detection. Though this is a simpler task, we confront our work also with them, considering healthy and sclerotic glomeruli as a single class. In [19,20] the authors used classical machine learning approaches, obtaining worse results than us. In [21], Temerinac-Ott et al. compared a machine learning approach, based on Histogram of Oriented Gradients (HOG) [36]) feature extraction and a support vector machine (SVM) classifier, and a deep learning one with CNN. Anyway, both obtained lower performances than our end-to-end instance segmentation framework. Gallego et al. exploited a CNN for classifying if each patch is a glomerulus or not in a sliding window fashion [16]. Although this may look like a more naive approach, compared to adopting a detector from the R-CNN family (which can also reduce the problem of redundant computation across neighboring patches), the results obtained in the paper are quite impressive, with a recall of 1. However, it is worth noting that Gallego et al. considered only glomeruli with area of at least  $200 \times 200$  pixel ( $>100 \mu\text{m}$  of diameter), whereas we consider glomeruli of all sizes in the metrics, and many of our false negatives are among the small glomeruli. Furthermore, we provide a precise mask for each glomerulus found, while Gallego et al. can only determine coarse masks composed by the union of rectangular patches they considered. Kawazoe et al. used Faster R-CNN for the glomerular detection task, obtaining results comparable with the proposed Mask R-CNN approach, with an F-score of 0.925 (ours is 0.919) [13]. We believe that the possibility to use a larger training dataset (200 WSIs instead of 26) can explain why they can get comparable (or even slightly better) results even with a less powerful model. As already noted, our Faster R-CNN model performs worse than our Mask R-CNN one.

## 5. Conclusions

In this paper, we develop a framework that could aid pathologists in the process of automatically detecting and classifying non-sclerotic and sclerotic glomeruli from sections of kidney biopsies. The proposed approach relies on Mask R-CNN, which proved to be a very sensible choice for a glomerular detection and classification task, improving over the baseline Faster R-CNN method and our previous works based on semantic segmentation approaches [15]. The proposed method allows to train an end-to-end instance segmentation neural network, therefore strongly reducing the need for post processing operations and allowing to learn all the required features in a unified process. An interesting novelty concerning post processing is the development of the Non-Maximum-Area suppression algorithm, that with seemingly minor changes compared to standard NMS algorithm, led to an improvement of the performances in our sliding window approaches. Note that NMAS, like NMS, is a general purpose algorithm and can be useful also for other detection tasks. The best model we trained is based on Mask R-CNN, and exploits NMAS for projection on full images. It outperforms related works in the field of the determination of global glomerulosclerosis, as [15,17]. The methods we used for evaluating the validity of our detection models are more specific than widespread global metrics (as *mAP*) used in benchmark datasets as PASCAL VOC or COCO. The analysis of object detection confusion matrices allows a better understanding of the model performance, bringing an insight on the model response for each problem class. At the moment, the proposed framework allows to get a reliable estimate of global glomerulosclerosis; the pathologists



can benefit from glomeruli annotations provided by our CAD through an XML interface with the commonly used Aperio ImageScope software, easing the burden of the manual annotation. In the future, it could be extended to other kidney biopsies analysis tasks, consenting to define the complete Karpinski histological score.

**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, L.G. and V.B.; validation, M.R., F.P., and L.G.; writing—original draft, N.A.; writing—review and editing, G.D.C., A.B., I.D.F., D.B., M.R., F.P., L.G., and V.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

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

## Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Meaning
AP	Average Precision
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
CR	Congo Red
DETO	Department of Emergency and Organ transplantation
FCN	Fully convolutional network
HE	Hematoxylin and Eosin
HOG	Histogram of Oriented Gradients
<i>IoM</i>	Intersection over Minimum
<i>IoU</i>	Intersection over Union
JS	Jones Silver
<i>mAP</i>	mean average precision
mrcLBP	multi-radial color local binary patterns
NMAS	Non-Maximum-Area Suppression
NMS	Non-Maximum Suppression
PAM	Periodic acid-methenamine silver
PAS	Periodic acid–Schiff
R-CNN	Region-based Convolutional Neural Network
R-HOG	Rectangle-Histogram of Oriented Gradients
RoI	Region of Interest
RPN	Region Proposal Network
S-HOG	Segmental-Histogram of Oriented Gradients
SR	Sirius Red
SVM	Support vector machine
TRI	Gömöri's Trichrome
VOC	Visual object classes
WSI	Whole slide image

## Appendix A. Implementation Details

### Appendix A.1. Faster R-CNN-Based Detector

For training a Faster R-CNN detector, we use the patches obtained using the procedures described so far, including data augmentation techniques for rebalancing the dataset. Here we describe how we tuned the most important hyperparameters for the Faster R-CNN approach. Faster R-CNN hyperparameters are in Table A1. Since the training of a Faster R-CNN involves 4 different stages, we have to specify training options for each of these stages. We used Adam Optimizer for all the stages. We tuned hyperparameters for each of these stages according to Table A2. We trained 10 epochs for each stage, and we lowered the learning rate in the 3rd and

4th stage. Further details regarding hyperparameters used in each stage can be found in MATLAB documentation of `trainfastrcnnobjectdetector` method (<https://www.mathworks.com/help/vision/ref/trainfastrcnnobjectdetector.html>).

**Table A1.** Faster R-CNN hyperparameters.

Faster R-CNN	
Hyperparameter	Value
CNN	resnet50
NegativeOverlapRange	[0 0.3]
PositiveOverlapRange	[0.6 1]
NumRegionsToSample	256
BoxPyramidScale	1.2
NumStrongestRegions	512

**Table A2.** Hyperparameters per stages of Faster R-CNN.

Hyperparameters per stages of Faster R-CNN	
Hyperparameter	Value
<b>All stages</b>	
Optimizer	ADAM
MaxEpochs	10
MiniBatchSize	1
<b>Stage 1</b>	
InitialLearnRate	0.0001
<b>Stage 2</b>	
InitialLearnRate	0.0001
<b>Stage 3</b>	
InitialLearnRate	0.000001
<b>Stage 4</b>	
InitialLearnRate	0.000001

#### Appendix A.2. Mask R-CNN Based Detector

We tuned training hyperparameters according to Table A3. These hyperparameters refer to Mask R-CNN implementation realized by Matterport Inc., so more details are available in the documentation [34]. The inference configuration is slightly different, with `IMAGE_RESIZE_MODE = "pad64"` and `RPN_NMS_THRESHOLD = 0.7`.

**Table A3.** Hyperparameter tuning for Mask R-CNN-based detector.

Training Configuration	
Hyperparameter	Value
BACKBONE	resnet50
RPN_ANCHOR_SCALES	(32, 96, 160, 200, 256)
RPN_ANCHOR_RATIOS	[0.5, 1, 2]
POST_NMS_ROIS_TRAINING	800
POST_NMS_ROIS_INFERENCE	1600
RPN_NMS_THRESHOLD	0.8
RPN_TRAIN_ANCHORS_PER_IMAGE	64
MEAN_PIXEL	[218.85, 198.25, 207.18]
MINI_MASK_SHAPE	(56, 56)
TRAIN_ROIS_PER_IMAGE	128
IMAGE_RESIZE_MODE	crop
IMAGE_MIN_DIM	1024
IMAGE_MAX_DIM	1024
LEARNING_RATE	0.001
LEARNING_MOMENTUM	0.9
WEIGHT_DECAY	0.0001
GRADIENT_CLIP_NORM	5.0

We performed data augmentation, exploiting the *imgaug* library (<https://imgaug.readthedocs.io/en/latest/>) [37], as reported in Table A4. In particular, of the augmentations listed there, we randomly performed none, one, or two augmentations.

**Table A4.** Augmentations for Mask R-CNN approach.

Data Augmentation	
Type	Details
Flip upside-down	$P(\text{flip}_{ud}) = 0.5$
Flip left-right	$P(\text{flip}_{lr}) = 0.5$
Rotate	$\theta \in \{90^\circ, 180^\circ, 270^\circ\}$
Multiply	$\alpha \in [0.8, 1.1]$
Gaussian Blur	$\sigma \in [0, 0.1]$

## References

1. Bevilacqua, V.; Cariello, L.; Columbo, D.; Daleno, D.; Dellisanti Fabiano, M.; Giannini, M.; Mastronardi, G.; Castellano, M. Retinal Fundus Biometric Analysis for Personal Identifications. In Proceedings of the Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence (ICIC 2008), Shanghai, China, 15–18 September 2008; doi:10.1007/978-3-540-85984-0\_147.
2. Menolascina, F.; Bellomo, D.; Maiwald, T.; Bevilacqua, V.; Ciminelli, C.; Paradiso, A.; Tommasi, S. Developing optimal input design strategies in cancer systems biology with applications to microfluidic device engineering. *BMC Bioinform.* **2009**, *10*, S4, doi:10.1186/1471-2105-10-S12-S4.
3. Bevilacqua, V.; Dimauro, G.; Marino, F.; Brunetti, A.; Cassano, F.; Di Maio, A.; Nasca, E.; Trotta, G.F.; Girardi, F.; Ostuni, A.; et al. A novel approach to evaluate blood parameters using computer vision techniques. In Proceedings of the 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Benevento, Italy, 15–18 May 2016; pp. 1–6, doi:10.1109/MeMeA.2016.7533760.
4. Bevilacqua, V.; Mastronardi, G.; Menolascina, F.; Pannarale, P.; Pedone, A. A Novel Multi-Objective Genetic Algorithm Approach to Artificial Neural Network Topology Optimisation: The Breast Cancer Classification Problem. In Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 16–21 July 2006; pp. 1958–1965, doi:10.1109/IJCNN.2006.246940.
5. Bevilacqua, V.; D’Ambruso, D.; Mandolino, G.; Suma, M. A new tool to support diagnosis of neurological disorders by means of facial expressions. In Proceedings of the 2011 IEEE International Symposium on Medical Measurements and Applications, Bari, Italy, 30–31 May 2011; pp. 544–549, doi:10.1109/MeMeA.2011.5966766.
6. Dimauro, G.; Girardi, F.; Gelardi, M.; Bevilacqua, V.; Caivano, D. Rhino-Cyt: A System for Supporting the Rhinologist in the Analysis of Nasal Cytology. In Proceedings of ICIC 2018: Intelligent Computing Theories and Application, Wuhan, China, 15–18 August 2018; doi:10.1007/978-3-319-95933-7\_71.
7. 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.
8. Dimauro, G.; Simone, L. Novel biased normalized cuts approach for the automatic segmentation of the conjunctiva. *Electronics* **2020**, *9*, 997.
9. Dimauro, G.; Colagrande, P.; Carlucci, R.; Ventura, M.; Bevilacqua, V.; Caivano, D. CRISPRLearner: A deep learning-based system to predict CRISPR/Cas9 sgRNA on-target cleavage efficiency. *Electronics* **2019**, *8*, 1478.
10. Zhao, Z.Q.; Zheng, P.; Xu, S.-t.; Wu, X. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232.
11. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv* **2017**, arXiv:1704.06857.
12. Ledbetter, D.; Ho, L.; Lemley, K.V. Prediction of Kidney Function from Biopsy Images Using Convolutional Neural Networks. *arXiv* **2017**, arXiv:1702.01816.
13. 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, doi:10.3390/jimaging4070091.

14. 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. In Proceedings of ICIC 2019: Intelligent Computing Methodologies, Nanchang, China, 3–6 August 2019; doi:10.1007/978-3-030-26766-7\_66.
15. Altini, N.; Cascarano, G.D.; Brunetti, A.; Marino, F.; Rocchetti, M.T.; Martino, 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.
16. 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, doi:10.3390/jimaging4010020.
17. 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, doi:10.1109/TMI.2018.2851150.
18. 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.
19. Kato, T.; Relator, R.; Ngouv, H.; Hirohashi, Y.; Kakimoto, T.; Okada, K. New Descriptor for Glomerulus Detection in Kidney Microscopy Image. *arXiv* **2015**, arXiv:1506.05920.
20. Simon, O.; Yacoub, R.; Jain, S.; Tomaszewski, J.E.; Sarder, P. Multi-radial LBP Features as a Tool for Rapid Glomerular Detection and Assessment in Whole Slide Histopathology Images. *Sci. Rep.* **2018**, *8*, 2032, doi:10.1038/s41598-018-20453-7.
21. 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 10th International Symposium on Image and Signal Processing and Analysis, Ljubljana, Slovenia, 18–20 September 2017; pp. 19–24.
22. Bueno, G.; Fernandez-Carrobles, M.M.; Gonzalez-Lopez, L.; Deniz, O. Glomerulosclerosis identification in whole slide images using semantic segmentation. *Comput. Methods Prog. Biomed.* **2020**, *184*, 105273, doi:10.1016/j.cmpb.2019.105273.
23. Jha, A.; Yang, H.; Deng, R.; Kapp, M.E.; Fogo, A.B.; Huo, Y. Instance Segmentation for Whole Slide Imaging: End-to-End or Detect-Then-Segment. *arXiv* **2020**, arXiv:2007.03593.
24. 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, doi:10.1097/00007890-199904270-00013.
25. Remuzzi, G.; Grinyò, J.; Ruggenti, 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.
26. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
27. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
28. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. 2012 AlexNet. In Proceedings of the 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, NV, USA; 3–6 December 2012; doi:10.1016/j.protcy.2014.09.007.
29. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
30. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
31. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *arXiv* **2016**, arXiv:1506.01497v3.
32. Hoiem, D.; Chodpathumwan, Y.; Dai, Q. Diagnosing error in object detectors. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2012; pp. 340–353.

33. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS—improving object detection with one line of code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5561–5569.
34. Abdulla, W. Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow. 2017. Available online: [https://github.com/matterport/Mask\\_RCNN](https://github.com/matterport/Mask_RCNN) (accessed on 15 September 2020).
35. 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, doi:10.1109/CVPR.2016.90.
36. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
37. Jung, A.B.; Wada, K.; Crall, J.; Tanaka, S.; Graving, J.; Reinders, C.; Yadav, S.; Banerjee, J.; Vecsei, G.; Kraft, A.; et al. *Imgaug*. 2020. Available online: <https://github.com/aleju/imgaug> (accessed on 1 February 2020).

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).