Real-time detection of glomeruli in renal pathology

Robin Heckenauer
IRIMAS
Universit de Haute-Alsace
robin.heckenauer@uha.fr

Jonathan Weber
IRIMAS
Universit de Haute-Alsace
jonathan.weber@uha.fr

Cédric Wemmert
ICube
Universit de Strasbourg
wemmert@unistra.fr

Friedrich Feuerhake
Institute for Pathology
MHH, Hannover Medical School
Feuerhake.Friedrich@mh-hannover.de

Michel Hassenforder
IRIMAS
Universit de Haute-Alsace
michel.hassenforder@uha.fr

Pierre-Alain Muller
IRIMAS
Universit de Haute-Alsace
pierre-alain.muller@uha.fr

Germain Forestier
IRIMAS
Universit de Haute-Alsace
germain.forestier@uha.fr

***Abstract*—The field of digital pathology emerged with the introduction of whole slide imaging scanners and lead to the development of new tools for analyzing histopathological slides. The availability of digital representation of the slides has motivated the development of artificial intelligence methods to automatically identify microscopic structures in order to support pathologists in their diagnosis. Unlike many existing approaches targeting the detection of microscopic structures on static images at a given and fixed magnification level, our work focuses on the real-time detection of the structures at different scales. Indeed, real-time detection at different scales brings additional challenges but also better mimics the way pathologists work as they continuously move the slides and change the magnification level during their analysis. In this paper, we focus on renal pathology and more specifically on the real-time detection of glomeruli at different scales. Our method is based on the deep learning object detection model YOLOv3 pre-trained on the COCO dataset and fine tuned to detect glomeruli. We investigate the benefits of using multi-scale images to improve the network ability to detect glomeruli at variable magnification levels in real time.**

***Index Terms*—real-time glomerulus detection, digital pathology, deep convolutional networks**

I. Introduction

According to the Organ Procurement and Transplantation Network (OPTN) [1], 113,021 patients were on a waiting list for organs in 2019. Among them, 94,715 ($83,8\%$) needed a kidney transplantation while at the same time only 13,408 donations of kidney graft were made. As a consequence, the median waiting time is between a year and a half, and three years. In this process, kidney graft rejection occurred in $21,5\%$ cases after five years following the transplantation.

In order to early detect and prevent graft rejection, a common procedure is to perform a kidney biopsy and observe the tissue through a microscope. Pathologists generally focus on observing glomeruli, whose main function is the filtration of urine. Glomeruli identification and examination in the slides inform the pathologists about the current state of the kidney and can help detecting rejection at early stage.

In recent years, the field of pathology evolved with the advent of digital pathology and the introduction of whole slide imaging scanners which lead to the development of new tools for analyzing histopathology slides. The availability of digital representation of the slides has motivated the development of artificial intelligence methods to automatically identify microscopic structures in order to support pathologists in their diagnosis. However, analyzing Whole Slide Images (WSIs) has two significant challenges. First, WSIs are difficult to handle due to their large size (up to 70,000 pixels per side). Second, during the WSI creation process, the tissues are manually sliced which can lead to tissue deformation. Those geometric transformations can affect glomeruli size, shape and texture. Figure 1 shows examples of different glomeruli as they appear in WSI.

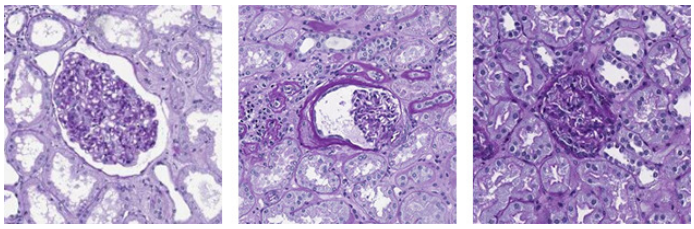

Fig. 1. Examples of glomeruli as they appear in WSI with variety of size, shape and texture (PAS staining).

Unlike many existing approaches that target the detection of microscopic structures on static images at a given and fixed magnification level [2]–[4], our work focuses on the real-time detection of the structures at different scales. Real-time detection at different scales brings additional challenges but also better mimics the way pathologists work as they continuously move the slides and change the magnification level during their analysis. Our method is based on the deep learning object detection model YOLOv3 pre-trained on the COCO dataset [5] and fine tuned to detect glomeruli. As shown in [6], pathologists assisted by AI improved their overall performance during a diagnosis. Based on this observation, this work is a proof-of-concept of what could be done in the future to help pathologists in their daily work as in [7].

The paper is organized as follows: in Section II our method for real-time multi-scale glomeruli detection is presented. Then, some experiments on renal WSIs are described in

This work was supported by the SysMIFTA project (FKZ:031L-0085A).

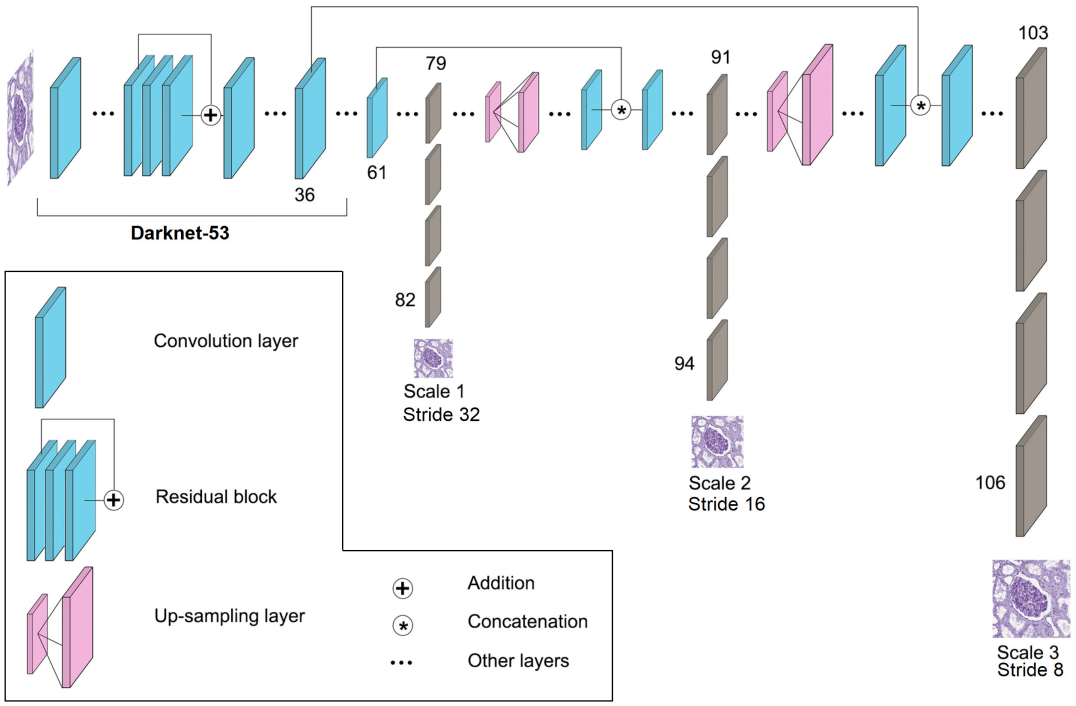

Fig. 2. The YOLOv3 architecture with Darknet-53 as backbone. The architecture is composed of 106 fully convolutional layers.

Section III and discussed in Section IV to highlight the potential benefits of this approach in pathology.

II. Method

A. Real-time object detection

Numerous methods have been proposed for real-time object detection: the real-time object detector SSD [8] and RetinaNet [9] detect multiple objects in images by taking information at one single shot. Whereas SSD, R-CNN [10] uses a two-stage architecture to extract 2,000 region proposals and detect object in each proposal. R-CNN authors use a CNN followed by a SVM to detect and classify objects. To overcome the long time required to train the CNN, R-CNN authors remove the 2,000 region proposals and directly feed the CNN with the entire image. This improvement leads to the state-of-the-art real-time object detection architecture called Faster-RCNN [11]. Furthermore, Kawazoe et al. [12] show the superiority of Faster-RCNN for glomeruli detection task. Due to its two-stage detector, Faster-RCNN is one of the slowest real-time method compared to SSD, RetinaNet and YOLO [13] as shown in [14]. Our goal being to perform detection in real-time on images displayed by a microscope, we want a fast architecture. We made a trade-off between quality of the detection and speed. Therefore, we privilege YOLO for its performance close to Faster-RCNN and its speed superior to Faster-RCNN.

B. YOLOv3 architecture

To perform detection, YOLO uses features from 3 different scales as shown in Figure 2. For each input image, the architecture uses multiple downsampling layers to reduce the image size to scale 1. Objects of interest are detected at this level with stride 32. Then, the image is upsampled to scale 2 and the network performs detection with stride 16. Finally the image is upsampled again and detection is performed with stride 8 at scale 3. The feature maps from the three different scales are extracted and concatenated. At each scale, the network predicts three bounding boxes locations x, y, w, h along with the class probability. To filter all those predictions, a confidence score is defined (see Equation 2) as the product of the class probability multiplied by the Intersection Over Union (IOU) between the predictions and the ground truth:

$$IOU = \frac{pred \cap truth}{pred \cup truth} \tag{1}$$

$$ConfidenceScore = Pr(Class_i) * IOU_{pred}^{truth} \tag{2}$$

In other words, a high confidence score means a high probability for a bounding box to belong to an object given by the network. Hence, the detections with a low confidence score are often low quality detections. A threshold is established to remove those detections with a low confidence. After this step, it remains multiple detections to the same object. To resolve this problem, non-maximum suppression (NMS) [15] algorithm is used to remove the redundant detection belonging to the same object. In this way, it remains only one single bounding box per object in the image.

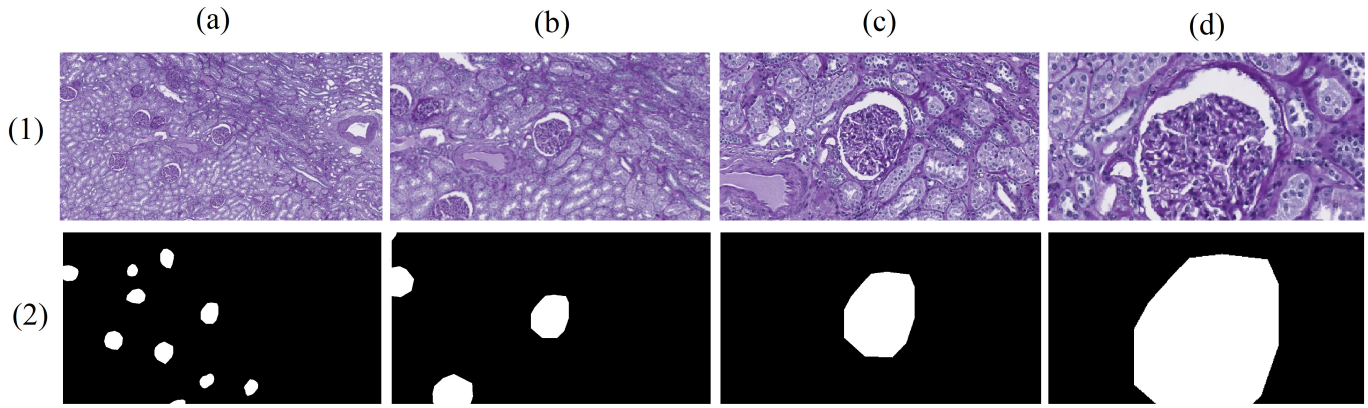

Fig. 3. The first row (1) shows examples of patches with a size of 736x1280 pixels took at different scales. The second row (2) shows the associated ground truth with in black the tissue and in white the glomeruli (columns (a) to (d) are respectively at 40X, 20X, 10X and 5X magnification).

C. Implementation

We use YOLOv3 implementation provided by [16] which is programmed in Python using Keras [17]. We employ a transfer learning approach called fine tuning to transfer the neural network weights from the model pre-trained on COCO [5] dataset to our dataset (COCO is one of the largest images dataset for the detection task, regrouping 123,287 images which contains 886,284 instances). Then, the model is trained on glomeruli images to be adapted to our task. Fine tuning is very useful in our context where data are hard to obtain. Thus, it allow us to use a modest size dataset and still have an efficient and relevant model as in [18].

In our implementation, we train our network with field-of-view images as it better mimics the way a pathologist see a part of a slide under a virtual microscope during a diagnosis. We take a standard monitor size of 720p as reference to feed our network. Hence, each WSI is divided in patches of the same size (736[1]x1280 pixels). In order to analyze the network performance and the impact of each scale, we create 5 distinct datasets to lead our experiments. Four datasets contain images at one single scale (40X, 20X, 10X and 5X magnification) and the last dataset contains images from all those scales. An example of patches and their associated ground truths are presented in Figure 3.

We expect that using multi-scale images will help the network to generalize its ability to detect glomeruli at different magnification levels in real time as shown in [19] as we want our model to be able to detect glomeruli at different scales.

D. Data augmentation

To mitigate the lack of data and increase the dataset diversity, we used the Augmentor [20] library to perform data augmentation on our dataset. We privilege transformations which could appear in reality due to human or mechanical manipulations as in [21]:

- affine: random rotation and horizontal/vertical flip;
- contrast: random variation of shadow and light.

We apply the chosen data augmentation methods to every images in our datasets and we remove the images produced without glomerulus. This process increase our total number of images by 6 for each dataset as shown in Table I.

III. Experiments and Results

A. Data

WSIs used in experiments were collected by Hannover Medical School (MHH) from 10 patients who received a kidney graft which failed. Kidney were extracted by nephrectomy after complete loss of function. The collected tissues were retained in paraffin, then samples were cut in 3 µm slices. Staining instrument (Ventana Benchmark Ultra) were used to stain slides with PAS staining. This process highlights the tissue structures. A digital whole slide scanner (Aperio AT2) created 40X magnification images from slides. Largest image size is 113,543 x 76,898 pixels. Every WSI has been annotated by pathology experts thanks to Cytomine [22].

For our experiments, we use 10 WSIs containing between 86 to 443 glomeruli as presented in Table II. To tackle overfitting, we use cross-validation with 5 folds ($K = 5$). The Table II also shows the number of glomeruli by cross-validation fold which vary between 400 and 718 glomeruli. During the cross-validation process, each WSI is presented to the network as a test data once as shown in Table III.

Next, we create 5 distinct datasets to feed our networks. Four of them contain images (736x1280 pixels) of one single scale (40X, 20X, 10X and 5X magnification). The last dataset contains multi-scale images from all the previous datasets.

B. Experiments

We train a network with YOLOv3 architecture on each dataset (40X, 20X, 10X, 5X, multi-scale) to observe the impacts of using multi-scale images. We set the same parameters and hyper-parameters during the training part to compare the networks performance. The COCO model provide by YOLO

[1]The YOLOv3 network downsamples input images by 32. Hence, the input images must be multiple of 32. So we choose 736x1280 pixels instead of 720x1280 pixels

	Dataset				
Number of images	40X	20X	10X	5X	Total
Before data augmentation	5,701	2,768	1,130	412	10,011
After data augmentation	35,185	16,851	7,372	2,883	62,272

TABLE I
NUMBER OF IMAGES IN EACH DATASET BEFORE AND AFTER DATA AUGMENTATION.

Patient ID	#glomeruli per WSI	Fold number	#glomeruli per fold
10	192	1	400
11	208		
12	86	2	506
13	420		
14	208	3	436
15	228		
16	443	4	718
17	275		
18	360	5	461
19	101		

TABLE II
THE NUMBER OF GLOMERULI PER WSI AND FOLD.

Split number	Train	Validation	Test
First split	3, 4, 5	2	1
Second split	1, 4, 5	3	2
Third split	1, 2, 5	4	3
Fourth split	1, 2, 3	5	4
Fifth split	2, 3, 4	1	5

TABLE III
THE DISTRIBUTION OF FOLDS IN PATIENTS IN TRAIN, VALIDATION AND TEST SPLIT.

authors is used as pre-trained model. To get a stable loss, we freeze all the layers except 3. Then, we train our networks with a learning rate of 0.001, a batch of 1 and 50 epochs. We then fine tune our network by unfreezing the layers and retrain it with a lower learning rate of 0.0001, a batch of 1 and 50 epochs. To force YOLOv3 to learn the correct information, we experimentally choose an IOU of 0.3 and a confidence score of 0.2 during the training process. Those values are the best to filter bad predictions and keep the majority of best detections as shown in [23] and [24]. We evaluate the model with the same IOU and confidence values. Adam optimizer and YOLO loss proposed by YOLO authors are used.

The entire process is performed on Nvidia GTX 1080 GPU cards provided by the computer center Mesocentre and last 4 to 72 hours depending of the dataset used. To evaluate the model, we compute Precision and Recall from True positive (TP), False positive (FP) and False negative (FN) instances. We show an example of detections made by the model on several patches in Figure 4.

C. Results

Once training is over, each network is tested on each dataset (40X, 20X, 10X, 5X, multi-scale). We use the same IOU and confidence score as training to evaluate the networks. Then as a first step, the TP, FP and FN obtained by our models on the test images are drawn to visually evaluate the results

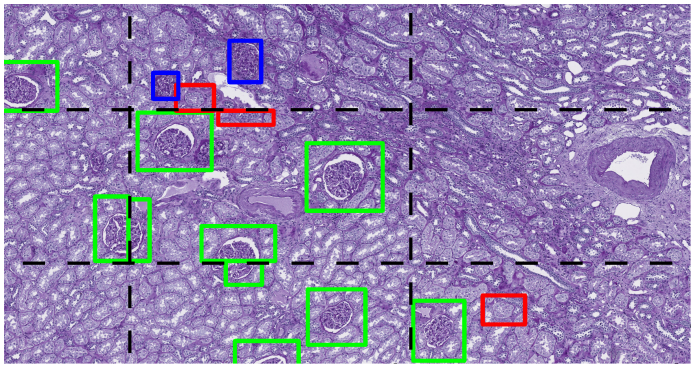

Fig. 4. Example of detection made by one of our model on field-of-view images from a WSI stained in PAS. We draw a black dotted line to show the bound of images. Green bounding boxes show the correct predictions made by the network, red are incorrect detection and blue missed glomeruli.

as shown in Figure 5. Secondly, we compute F1-scores (five-fold cross validation) shown in Table IV. An example of real-time detection performed by our network trained with 5X magnification images on a video is available here[2].

IV. DISCUSSION

The results show that the network trained on 20X magnification images gives the best F1-score on 40X magnification images. According to these results, 20X magnification images have more relevant information to improve the understanding ability of the network at 40X magnification. In a different way, the network trained on 5X magnification images is the best for datasets containing 20X, 10X and 5X magnification images. Again, this observation shows that some magnification levels are more relevant for the network and it seems a good idea to train the network using images with a lower magnification.

Unlike the results found by Song et al. [19] where the multi-scale approach got the best F1-scores, in our experiments, the 5X and multi-scale networks have the same F1-score when they are used on multi-scale dataset. The multi-scale network seems to fail to generalize on several magnification level.

Furthermore, some F1-scores in Table IV are close to or equal 0. That happens when we test a model with a dataset which contains images with a much lower magnification. This observation show our networks can easily detect glomeruli with a higher magnification level rather than a lower magnification level. As a consequence, to properly detect objects of interest with an AI assisted microscope, it seems interesting to use images with an low magnification level for the train the network. One last thing to note, the best performance are given

[2] http://jonathan-weber.eu/cp/cbms2020/

		Datasets used for testing				
	Magnification	40X	20X	10X	5X	40X,20X,10X,5X
Datasets used for training	40X	0.38	0.18	0	0	0.32
	20X	**0.44**	0.40	0.03	0	0.35
	10X	0.35	0.44	0.56	0.18	0.43
	5X	0.18	**0.53**	**0.78**	**0.77**	**0.48**
	40X,20X,10X,5X	0.43	0.49	0.61	0.46	**0.48**

TABLE IV
F1-SCORES RESULTS FOR YOLOV3 TRAINED AND TESTED ON DATASET OF DIFFERENT SCALES.

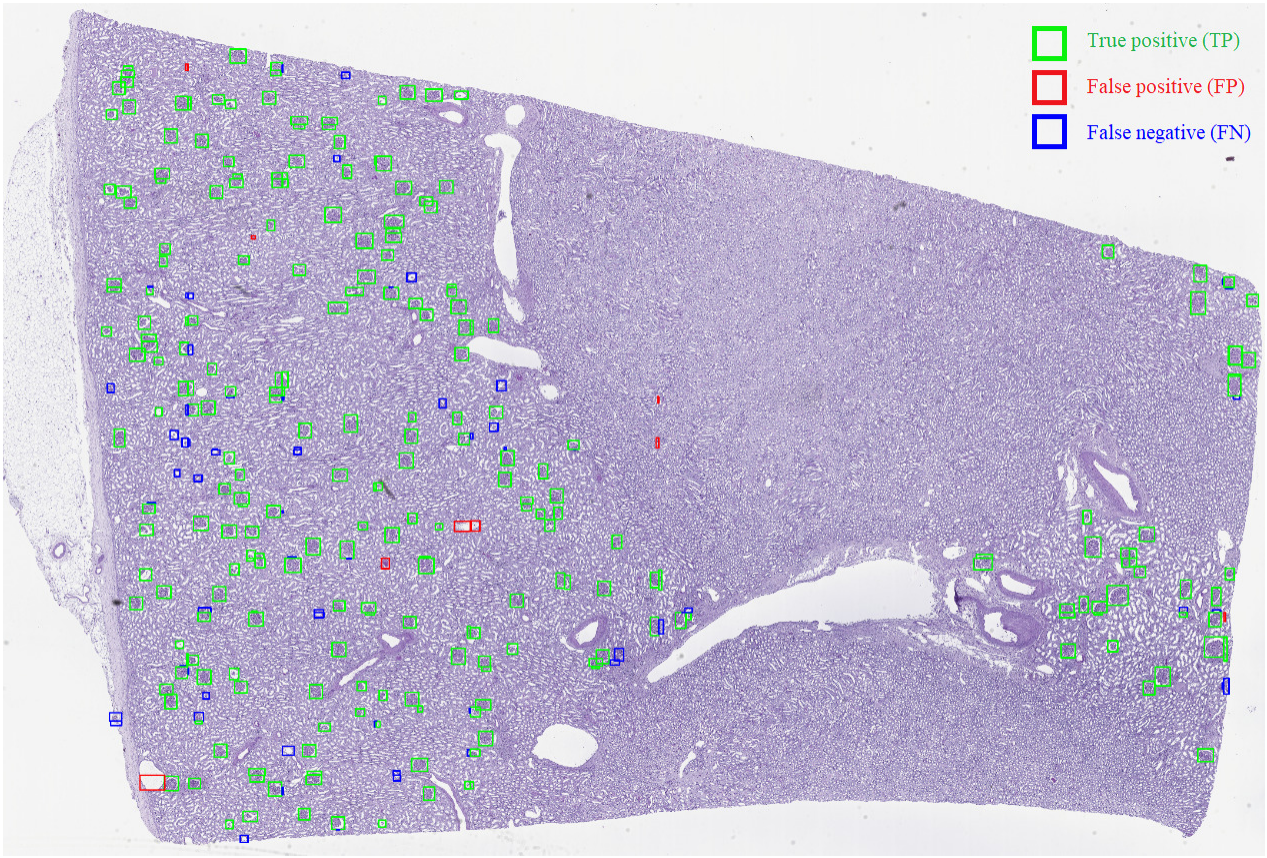

Fig. 5. Example of detection made by the network trained and tested on 5X magnification images. The images are merged together to reconstruct the original WSI of the patient 15. Green bounding boxes show the correct predictions made by the network, red are incorrect detection and blue missed glomeruli.

by 5X network which is trained with less than 500 original images (a low magnification level produces less images).

In Figure 5, we observe many missed glomeruli. This is due to the fact that the detections performed by the network are filtered by the confidence and the IOU threshold. Filtering the detections decreases the number of glomeruli found, but it also decreases the number of false detections found by the network. So there is a trade-off to get the best F1-Score possible.

V. CONCLUSION

The recent research in analysis of WSI in digital pathology has shown good results for computer vision task in medical images. However, to assist the pathologist in his/her daily work, those tasks, like glomeruli detection, must be realized in real-time directly on microscope. In this paper, we proposed a real-time detection method for glomeruli in renal pathology. Considering that in a real case the pathologist continuously changes the magnification level, we explore multi-scale method to enhance the network performance. Our experimental results indicate that the choice of magnification level for training is crucial to obtain good results. To improve detection performance, future works could investigate ensemble learning to combine networks output at different scales. Moreover, we can imagine to deal with additional challenges such as multi-stain WSI and multiclass glomeruli (healthy, sclerotic, partially sclerotic) in order to determine in real-time the state of a kidney according to Banff classification. Finally, it would be interesting to deploy the approach in clinical research setting and to assess its influence through interviews with pathologists.

VI. Acknowledgments

ERACoSysMed project "SysMIFTA", co-funded by EU H2020 and the national funding agencies German Ministry of Education and Research (BMBF) project management PTJ (FKZ: 031L-0085A), and Agence National de la Recherche (ANR), project number ANR-15-CMED-0004.

The authors would like to acknowledge the High Performance Computing center of the University of Strasbourg for supporting this work by providing scientific support and access to computing resources. Part of the computing resources were funded by the Equipex Equip@Meso project (Programme Investissements d'Avenir) and the CPER Alsacalcul/Big Data.

References

[1] M. D. Ellison, M. A. McBride, S. E. Taranto, F. L. Delmonico, and H. M. Kauffman, "Living kidney donors in need of kidney transplants: a report from the organ procurement and transplantation network," *Transplantation*, vol. 74, no. 9, pp. 1349–1351, 2002.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[3] S. Kannan, L. A. Morgan, B. Liang, M. G. Cheung, C. Q. Lin, D. Mun, R. G. Nader, M. E. Belghasem, J. M. Henderson, J. M. Francis *et al.*, "Segmentation of glomeruli within trichrome images using deep learning," *Kidney international reports*, vol. 4, no. 7, pp. 955–962, 2019.

[4] J. N. Marsh, M. K. Matlock, S. Kudose, T.-C. Liu, T. S. Stappenbeck, J. P. Gaut, and S. J. Swamidass, "Deep learning global glomerulosclerosis in transplant kidney frozen sections," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2718–2728, 2018.

[5] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[6] A. Kiani, B. Uyumazturk, P. Rajpurkar, A. Wang, R. Gao, E. Jones, Y. Yu, C. P. Langlotz, R. L. Ball, T. J. Montine *et al.*, "Impact of a deep learning assistant on the histopathologic classification of liver cancer," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–8, 2020.

[7] V. Andreoli Petrolini, E. Beckhauser, A. Savaris, M. Ines Meurer, A. von Wangenheim, and D. Krechel, "Collaborative telepathology in a statewide telemedicine environment-first tests in the context of the brazilian public healthcare system," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems*, 2019, pp. 684–689.

[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[16] qqwweee, "keras-yolo3: A keras implementation of yolov3 (tensorflow backend)," 2019.

[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[11] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[12] Y. Kawazoe, K. Shimamoto, R. Yamaguchi, Y. Shintani-Domoto, H. Uozaki, M. Fukayama, and K. Ohe, "Faster R-CNN-based glomerular detection in multistained human whole slide images," *Journal of Imaging*, vol. 4, no. 7, p. 91, 2018.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[14] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[15] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-nms–improving object detection with one line of code," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5561–5569.

[17] F. Chollet *et al.*, "Keras," 2015.

[18] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[19] Y. Song, E.-L. Tan, X. Jiang, J.-Z. Cheng, D. Ni, S. Chen, B. Lei, and T. Wang, "Accurate cervical cell segmentation from overlapping clumps in pap smear images," *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 288–300, 2016.

[20] M. D. Bloice, P. M. Roth, and A. Holzinger, "Biomedical image augmentation using Augmentor," *Bioinformatics*, 04 2019, btz259.

[21] T. Lampert, O. Merveille, J. Schmitz, G. Forestier, F. Feuerhake, and C. Wemmert, "Strategies for training stain invariant cnns," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 905–909.

[22] R. Marée, L. Rollus, B. Stévens, R. Hoyoux, G. Louppe, R. Vandaele, J.-M. Begon, P. Kainz, P. Geurts, and L. Wehenkel, "Collaborative analysis of multi-gigapixel imaging data using cytomine," *Bioinformatics*, vol. 32, no. 9, pp. 1395–1401, 2016.

[23] Q. Peng, W. Luo, G. Hong, M. Feng, Y. Xia, L. Yu, X. Hao, X. Wang, and M. Li, "Pedestrian detection for transformer substation based on gaussian mixture model and yolo," in *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 2. IEEE, 2016, pp. 562–565.

[24] M. A. Al-Masni, M. A. Al-Antari, J.-M. Park, G. Gi, T.-Y. Kim, P. Rivera, E. Valarezo, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system," *Computer methods and programs in biomedicine*, vol. 157, pp. 85–94, 2018.