

Détection en temps réel des glomérules en pathologie rénale

R. Heckenauer¹ J. Weber¹ C. Wemmert² F. Feuerhake³
M. Hassenforder¹ P-A. Muller¹ G. Forestier¹

¹ IRIMAS, Université de Haute-Alsace, France

² ICube, Université de Strasbourg, France

³ Institut de pathologie, MHH, École de Médecine de Hanovre, Allemagne

robin.heckenauer@uha.fr

Résumé

Le domaine de la pathologie numérique a émergé avec l'introduction des scanners de lames et a conduit au développement de nouveaux outils d'analyse de lames histopathologiques. La disponibilité de la représentation numérique des lames a motivé le développement de méthodes issues de l'intelligence artificielle pour identifier automatiquement les structures microscopiques afin d'accompagner les pathologistes dans leur diagnostic. Contrairement à de nombreuses approches existantes visant la détection de structures microscopiques sur des images statiques à un niveau de grossissement donné et fixe, nos travaux se concentrent sur la détection en temps réel des structures à différentes échelles. En effet, la détection en temps réel à différentes échelles apporte des défis supplémentaires, mais imite également mieux la façon dont les pathologistes travaillent, car ils déplacent continuellement les lames et changent le niveau de grossissement pendant leur analyse. Dans cet article, nous nous concentrons sur la pathologie rénale et plus particulièrement sur la détection en temps réel des glomérules à différentes échelles. Notre méthode est basée sur le modèle de détection d'objets d'apprentissage profond YOLOv3 pré-entraîné sur le jeu de données COCO et affiné pour détecter les glomérules. Nous étudions les avantages de l'utilisation d'images multi-échelles pour améliorer la capacité du réseau à détecter les glomérules à des niveaux de grossissement variables en temps réel.

Mots-Clés

Détection de glomérule en temps réel, pathologie numérique, réseaux convolutifs profonds.

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.

Keywords

Real-time glomerulus detection, digital pathology, deep convolutional networks.

1 Introduction

Selon le Réseau d'Approvisionnement et de Transplantation d'Organes (OPTN) [6], 113 021 patients étaient sur liste d'attente pour un organe en 2019. Parmi eux, 94 715 (83,8%) ont eu besoin d'une transplantation rénale alors que dans le même temps seulement 13 408 dons de greffe rénale ont été effectués. En conséquence, le temps d'attente médian se situe entre un an et demi et trois ans. Dans ce processus, le rejet de greffe rénale s'est produit dans 21,5% des cas 5 ans après la transplantation.

Afin de détecter et de prévenir le rejet de greffe, une procédure courante consiste à effectuer une biopsie rénale et à observer le tissu au microscope. Les pathologistes se concentrent généralement sur l'observation des glomérules, dont la fonction principale est la filtration de l'urine. L'identification et l'examen des glomérules sur les lames informent les pathologistes de l'état actuel du rein et peuvent aider à détecter le rejet à un stade précoce.

Ces dernières années, le domaine de la pathologie a évo-

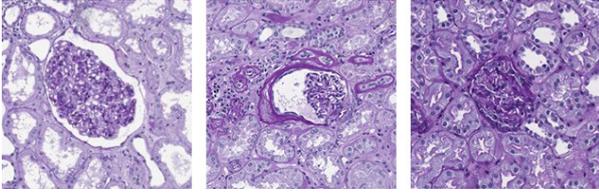


FIGURE 1 – Exemples de glomérules tels qu’ils apparaissent dans une WSI avec une variété de taille, forme et texture (coloration PAS).

lué avec l’avènement de la pathologie numérique et l’introduction de scanners de lames qui ont conduit au développement de nouveaux outils d’analyse de lames histopathologiques. La disponibilité de la représentation numérique des lames a motivé le développement de méthodes d’intelligence artificielle pour identifier automatiquement les structures microscopiques afin d’accompagner les pathologistes dans leur diagnostic. Cependant, l’analyse des images de lames (WSI) présente deux défis importants. Premièrement, les WSI sont difficiles à gérer en raison de leur grande taille (jusqu’à 70 000 pixels par côté). Deuxièmement, pendant le processus de création de WSI, les tissus sont découpés manuellement, ce qui peut entraîner une déformation des tissus. Ces transformations géométriques peuvent affecter la taille, la forme et la texture des glomérules. La figure 1 montre des exemples de différents glomérules tels qu’ils apparaissent dans une WSI.

Contrairement à de nombreuses approches existantes qui ciblent la détection de structures microscopiques sur des images statiques à un niveau de grossissement donné et fixe [22] [9] [17], nos travaux se concentrent sur la détection en temps réel des structures à différentes échelles. La détection en temps réel à différentes échelles apporte des défis supplémentaires, mais imite également mieux la façon dont les pathologistes travaillent, car ils déplacent continuellement les lames et changent le niveau de grossissement au cours de leur analyse. Notre méthode est basée sur le modèle de détection d’objets d’apprentissage profond YOLOv3 pré-entraîné sur le jeu de données COCO [13] et affiné pour détecter les glomérules. Comme le montre [11], les pathologistes assistés par l’IA ont amélioré leurs performances globales lors d’un diagnostic. Partant de ce constat, ce travail est une preuve de concept de ce qui pourrait être fait à l’avenir pour aider les pathologistes dans leur travail quotidien comme dans [2].

L’article est organisé comme suit : dans la section 2, notre méthode de détection des glomérules multi-échelles en temps réel est présentée. Ensuite, quelques expériences sur les WSI de reins sont décrites dans la section 3 et discutées dans la section 4 pour mettre en évidence les avantages potentiels de cette approche en pathologie.

2 Méthode

2.1 Détection d’objets en temps réel

De nombreuses méthodes ont été proposées pour la détection d’objets en temps réel : les détecteurs d’objets en temps réel SSD [15] et RetinaNet [14] détectent plusieurs objets dans les images en prenant des informations en une seule fois. Alors que SSD, R-CNN [7] utilisent une architecture en deux étapes pour extraire 2000 propositions de région et détecter un objet dans chaque proposition. Les auteurs de R-CNN utilisent un CNN suivi d’un SVM pour détecter et classer les objets. Pour surmonter le long temps nécessaire à l’entraînement du CNN, les auteurs de R-CNN retirent les 2000 propositions de régions et alimentent directement le CNN avec l’image entière. Cette amélioration conduit à l’architecture de détection d’objets en temps réel de pointe appelée Faster-RCNN [8]. De plus, Kawazoe et al. [10] montrent la supériorité de Faster-RCNN pour la tâche de détection des glomérules. En raison de son détecteur en deux étapes, Faster-RCNN est une méthode en temps réel lente par rapport aux SSD, RetinaNet et YOLO [20] comme indiqué dans [21]. Notre objectif étant d’effectuer une détection en temps réel sur des images affichées par un microscope, nous souhaitons une architecture rapide. Nous avons fait un compromis entre la qualité et la vitesse de détection. Par conséquent, nous privilégions YOLO pour ses performances proches de Faster-RCNN et sa vitesse supérieure à Faster-RCNN.

2.2 L’architecture YOLOv3

Pour effectuer la détection, YOLO utilise des caractéristiques de 3 échelles différentes comme le montre la Figure 2. Pour chaque image d’entrée, l’architecture utilise plusieurs couches de sous-échantillonnage pour réduire la taille de l’image à l’échelle 1. Les objets d’intérêts sont détectés à ce niveau avec un pas de 32. Ensuite, l’image est suréchantillonnée à l’échelle 2 et le réseau effectue la détection avec un pas de 16. Enfin, l’image est à nouveau suréchantillonnée et la détection est effectuée avec un pas de 8 à l’échelle 3. Les cartes des caractéristiques des trois échelles différentes sont extraites et concaténées. À chaque échelle, le réseau prédit trois emplacements de boîtes englobantes x, y, w, h avec une probabilité de classe associée. Pour filtrer toutes ces prédictions, un score de confiance est défini (voir l’équation 2) comme le produit de la probabilité de classe multipliée par l’intersection de l’union (*IOU*) entre les prédictions et la vérité terrain :

$$IOU = \frac{\text{prédictions} \cap \text{vérité terrain}}{\text{prédictions} \cup \text{vérité terrain}} \quad (1)$$

$$\text{Score de Confiance} = Pr(\text{Classe}_i) * IOU \quad (2)$$

En d’autres termes, un score de confiance élevé signifie une forte probabilité qu’une boîte englobante appartienne à un objet donné par le réseau. Par conséquent, les détections avec un faible score de confiance sont souvent des détections de faible qualité. Un seuil est établi pour supprimer

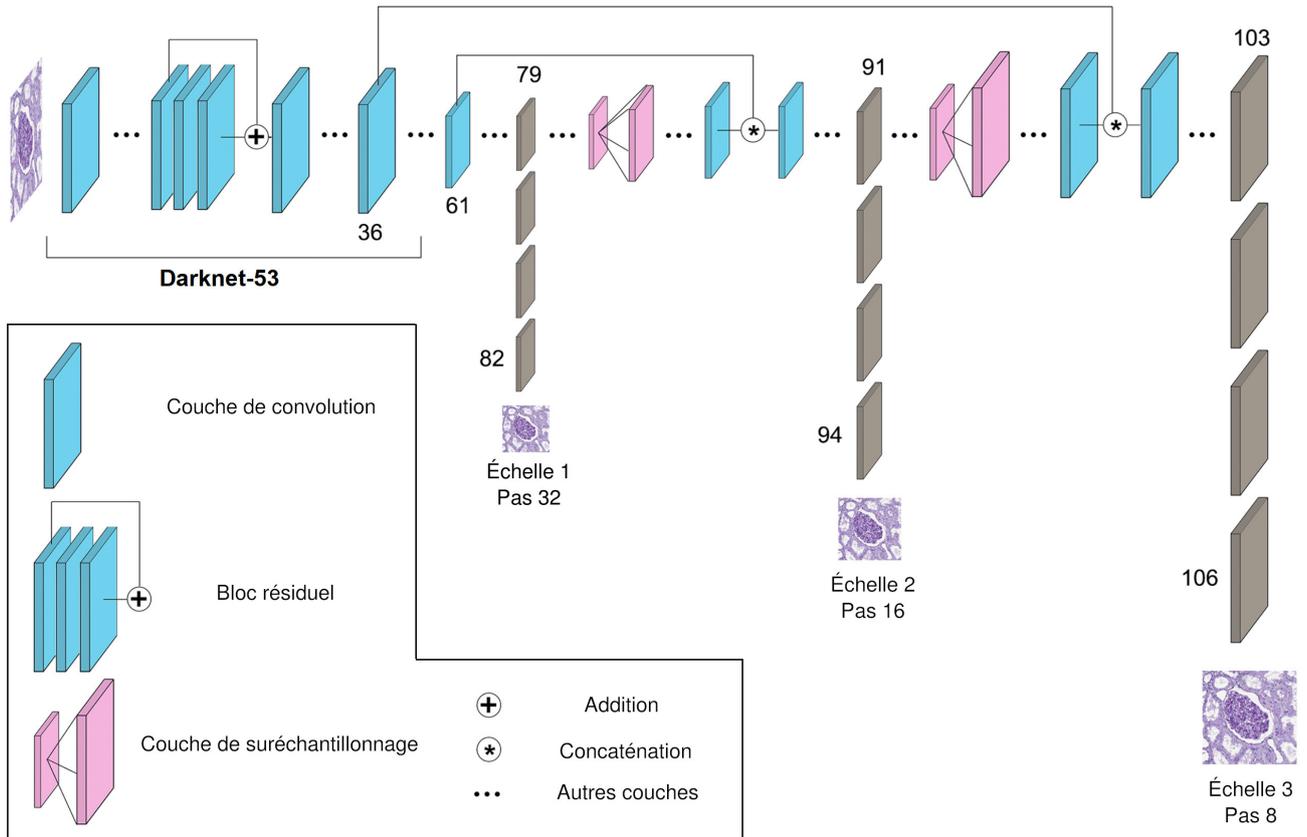


FIGURE 2 – L’architecture YOLOv3 avec Darknet-53 comme *backbone* (i.e. réalise l’extraction de caractéristiques). L’architecture est composée de 106 couches entièrement convolutives.

ces détections avec une faible confiance. Après cette étape, il reste plusieurs détections sur le même objet. Pour résoudre ce problème, l’algorithme de suppression non maximale (NMS) [4] est utilisé pour supprimer la détection redondante appartenant au même objet. De cette manière, il ne reste qu’une seule boîte englobante par objet dans l’image.

2.3 Mise en œuvre

Nous utilisons l’implémentation YOLOv3 fournie par [19] qui est programmée en Python en utilisant Keras [5]. Nous utilisons une approche d’apprentissage par transfert appelée *fine – tuning* pour transférer les poids du réseau neuronal du modèle pré-entraîné sur le jeu de données COCO[13] à notre jeu de données (COCO est l’un des plus grands jeu de données d’images pour la tâche de détection, regroupant 123 287 images qui contient 886 284 instances). Ensuite, le modèle est entraîné sur des images de glomérules pour être adapté à notre tâche. Affiner notre réseau est très utile dans notre contexte où les données sont difficiles à obtenir. Ainsi, cela nous permet d’utiliser un jeu de données de taille modeste tout en conservant un modèle efficace et pertinent comme dans [24].

Dans notre implémentation, nous formons notre réseau avec des extraits de lames entières scannées car cela imite

mieux la façon dont un/une pathologiste voit une partie d’une lame sous un microscope virtuel lors d’un diagnostic. Nous prenons une taille de moniteur standard de 720p comme référence pour alimenter notre réseau. Ainsi, chaque WSI est divisée en patches de même taille (736¹ x 1280 pixels). Afin d’analyser les performances du réseau et l’impact de chaque échelle, nous créons 5 jeux de données distincts pour mener nos expériences. Quatre jeux de données contiennent des images à une seule échelle (grossissement 40X, 20X, 10X et 5X) et le dernier jeu de données contient des images de toutes ces échelles. Un exemple de patches et de leurs vérités terrain associées est présenté dans la Figure 3.

Nous prévoyons que l’utilisation d’images multi-échelles aidera le réseau à généraliser sa capacité à détecter les glomérules à différents niveaux de grossissement en temps réel, comme indiqué dans [23] car nous voulons que notre modèle soit capable de détecter les glomérules à différentes échelles.

1. Le réseau YOLOv3 sous-échantillonne les images d’entrée par 32. Par conséquent, les images d’entrée doivent être multiples de 32. Nous choisissons donc 736 x 1280 pixels au lieu de 720 x 1280 pixels

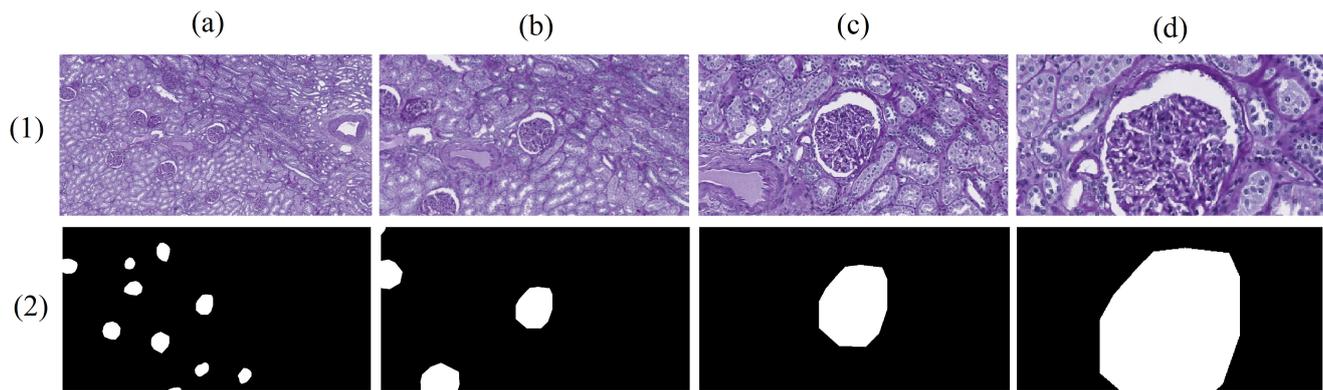


FIGURE 3 – La première ligne (1) montre des exemples de patches d’une taille de 736 x 1280 pixels pris à différentes échelles. La deuxième ligne (2) montre la vérité terrain associée avec en noir le tissu et en blanc les glomérules (les colonnes (a) à (d) sont respectivement à un grossissement de 40X, 20X, 10X et 5X).

2.4 Augmentation des données

Pour atténuer le manque de données et augmenter la diversité du jeu de données, nous avons utilisé la bibliothèque Augmentor [3] pour effectuer l’augmentation des données sur notre jeu de données. Nous privilégions les transformations qui pourraient apparaître dans la réalité du fait de manipulations humaines ou mécaniques comme dans [12] :

- affine : rotation aléatoire et retournement horizontal/vertical ;
- contraste : variation aléatoire de l’ombre et de la lumière.

Nous appliquons les méthodes d’augmentation de données choisies à toutes les images de nos jeux de données et nous supprimons les images produites sans glomérule. Ce processus multiplie notre nombre total d’images par 6 pour chaque jeu de données, comme indiqué dans le tableau 1.

3 Expériences et résultats

3.1 Données

Les WSI utilisées dans les expériences ont été collectées par la faculté de médecine de Hanovre (MHH) auprès de 10 patients ayant reçu une greffe de rein qui a échoué. Les reins ont été extraits par néphrectomie après perte complète de fonction. Les tissus collectés ont été conservés dans de la paraffine, puis les échantillons ont été coupés en tranches de 3 μm . Un instrument de coloration (Ventana Benchmark Ultra) a été utilisé pour colorer les lames avec une coloration PAS. Ce processus met en évidence les structures tissulaires. Un scanner numérique de diapositives entières (Aperio AT2) a créé des images avec un grossissement de 40X à partir de lames. La plus grande taille d’image est 113 543 x 76 898 pixels. Chaque WSI a été annoté par des experts en pathologie grâce à Cytomine [16].

Pour nos expériences, nous utilisons 10 WSI contenant entre 86 et 443 glomérules comme présenté dans le tableau 2. Pour lutter contre le surapprentissage, nous utilisons la validation croisée avec 5 blocs ($K = 5$). Le ta-

bleau 2 montre également le nombre de glomérules par bloc de validation croisée qui varie entre 400 et 718 glomérules. Pendant le processus de validation croisée, chaque WSI est présenté au réseau comme une donnée de test une fois, comme indiqué dans le tableau 3.

Ensuite, nous créons 5 jeux de données distincts pour alimenter nos réseaux. Quatre d’entre eux contiennent des images (736 x 1280 pixels) d’une seule échelle (grossissement 40X, 20X, 10X et 5X). Le dernier jeu de données contient des images multi-échelles de tous les jeux de données précédents.

3.2 Expériences

Nous entraînons un réseau avec l’architecture YOLOv3 sur chaque jeu de données (40X, 20X, 10X, 5X, multi-échelle) pour observer les impacts de l’utilisation d’images multi-échelles. Nous définissons les mêmes paramètres et hyperparamètres lors de la partie entraînement pour comparer les performances des réseaux. Le modèle COCO fourni par les auteurs YOLO est utilisé comme modèle pré-entraîné. Pour obtenir une perte stable, nous gèlons toutes les couches sauf les 3 dernières. Ensuite, nous entraînons nos réseaux avec un taux d’apprentissage de 0,001, un *batch* de 1 et 50 époques. Nous affinons ensuite notre réseau en dégelant les couches et en le recyclant avec un taux d’apprentissage inférieur de 0,0001, un *batch* de 1 et 50 époques. Pour forcer YOLOv3 à apprendre les informations correctes, nous choisissons expérimentalement une IOU de 0,3 et un score de confiance de 0,2 pendant le processus d’entraînement. Ces valeurs sont les meilleures pour filtrer les mauvaises prédictions et conserver la majorité des meilleures détections, comme indiqué dans [18] et [1]. Nous évaluons le modèle avec les mêmes valeurs d’IOU et de confiance. L’optimiseur Adam et la perte YOLO proposés par les auteurs YOLO sont utilisés.

L’ensemble du processus est effectué sur des cartes GPU Nvidia GTX 1080 fournies par le Centre de Calcul de l’Université de Strasbourg et dure de 4 à 72 heures selon

Nombre d'images	Jeux de données				
	40X	20X	10X	5X	Total
Avant l'augmentation des données	5,701	2,768	1,130	412	10,011
Après l'augmentation des données	35,185	16,851	7,372	2,883	62,272

TABLE 1 – Nombre d'images dans chaque jeu de données avant et après l'augmentation des données.

ID du patient	#glomérules par WSI	Numéro de bloc	#glomérules par bloc
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 2 – Le nombre de glomérules par WSI et par bloc.

Numéro de division	Entraînement	Validation	Test
Première division	3, 4, 5	2	1
Deuxième division	1, 4, 5	3	2
Troisième division	1, 2, 5	4	3
Quatrième division	1, 2, 3	5	4
Cinquième division	2, 3, 4	1	5

TABLE 3 – La distribution des blocs dans les divisions.

le jeu de données utilisé. Pour évaluer le modèle, nous calculons la précision et le rappel à partir des instances Vrai positif (TP), Faux positif (FP) et Faux négatif (FN). Nous montrons un exemple de détections effectuées par le modèle sur plusieurs patches dans la figure 4.

3.3 Résultats

Une fois la formation terminée, chaque réseau est testé sur chaque ensemble de données (40X, 20X, 10X, 5X, multi-échelle). Pour l'évaluation, nous utilisons l'IOU et le score de confiance utilisé pendant l'entraînement des réseaux. Ensuite, dans un premier temps, les TP, FP et FN obtenus par nos modèles sur les images de test sont dessinés pour évaluer visuellement les résultats comme le montre la figure 5. Deuxièmement, nous calculons les $F1 - scores$ (validation croisée $K=5$) indiqués dans le tableau 4. Un exemple de détection en temps réel effectuée par notre réseau entraîné avec des images agrandies 5X sur une vidéo est disponible ici ².

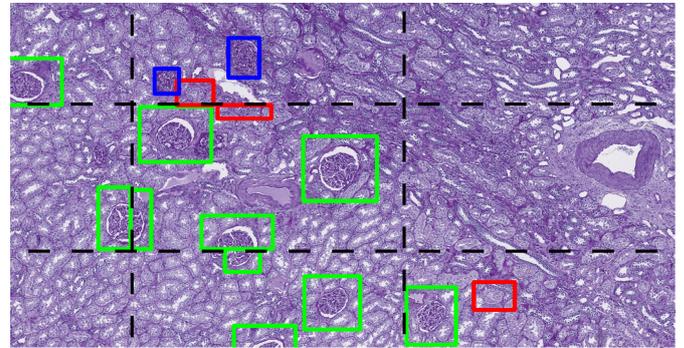


FIGURE 4 – Exemple de détection réalisée par l'un de nos modèles sur des patches à partir d'une WSI colorée en PAS. Nous dessinons une ligne pointillée noire pour montrer la limite des images. Les encadrés verts indiquent les prévisions correctes faites par le réseau, les rouges sont une détection incorrecte et les glomérules manqués sont bleus.

4 Discussion

Les résultats montrent que le réseau entraîné sur des images à grossissement 20X donne le meilleur $F1 - score$ sur des images à grossissement 40X. Selon ces résultats, les images à grossissement 20X ont des informations plus pertinentes pour améliorer la capacité de compréhension du réseau à un grossissement 40X. D'une manière différente, le réseau entraîné sur des images à grossissement 5X est le meilleur pour les jeux de données contenant des images à grossissement 20X, 10X et 5X. Encore une fois, cette observation montre que certains niveaux de grossissement sont plus pertinents pour le réseau et il semble judicieux de former le réseau en utilisant des images avec un grossissement plus faible.

Contrairement aux résultats trouvés par Song et al. [23] où l'approche multi-échelle a obtenu les meilleurs $F1 - scores$, dans nos expériences, les réseaux 5X et multi-échelles ont le même $F1 - score$ lorsqu'ils sont utilisés sur ensemble de données multi-échelles. Le réseau multi-échelles semble ne pas se généraliser à plusieurs niveaux de grossissement.

De plus, certains $F1 - score$ du tableau 4 sont proches ou égaux à 0. Cela se produit lorsque nous testons un modèle avec un ensemble de données contenant des images avec un grossissement beaucoup plus faible. Cette observation montre que nos réseaux peuvent facilement détecter les glomérules avec un niveau de grossissement plus élevé plutôt qu'un niveau de grossissement inférieur. En consé-

2. <http://jonathan-weber.eu/cp/cbms2020/>

		Ensembles de données utilisés pour les tests					
		Grossissement	40X	20X	10X	5X	40X,20X,10X,5X
Ensembles de données utilisés pour entraîner	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 4 – Résultats des $F1$ – scores pour YOLOv3 entraîné et testé sur un jeu de données à différentes échelles.

quence, pour bien détecter les objets d'intérêt avec un microscope assisté par l'IA, il semble intéressant d'utiliser des images à faible grossissement pour l'entraînement du réseau. Une dernière chose à noter, les meilleures performances sont données par le réseau 5X qui est entraîné avec moins de 500 images originales (un faible niveau de grossissement produit moins d'images).

Dans la figure 5, nous observons de nombreux glomérules manqués. Ceci est dû au fait que les détections effectuées par le réseau sont filtrées par les seuils de confiance et d'IOU. Le filtrage des détections diminue le nombre de glomérules trouvés, mais il diminue également le nombre de fausses détections trouvées par le réseau. Il y a donc un compromis pour obtenir le meilleur $F1$ – score possible.

5 Conclusion

La recherche récente dans l'analyse de WSI en pathologie numérique a montré de bons résultats pour la tâche de vision par ordinateur dans les images médicales. Cependant, pour assister le/la pathologiste dans son travail quotidien, ces tâches, comme la détection des glomérules, doivent être réalisées en temps réel directement au microscope. Dans cet article, nous avons proposé une méthode de détection en temps réel des glomérules en pathologie rénale. Considérant que dans un cas réel, le/la pathologiste change continuellement le niveau de grossissement, nous explorons une méthode multi-échelle pour améliorer les performances du réseau. Nos résultats expérimentaux indiquent que le choix du niveau de grossissement pour l'entraînement est crucial pour obtenir de bons résultats. Pour améliorer les performances de détection, les travaux futurs pourraient étudier l'apprentissage d'ensemble pour combiner la sortie des réseaux à différentes échelles. De plus, on peut imaginer faire face à des défis supplémentaires tels que les WSI multi-domaines et les glomérules multiclasse (sains, sclérosés, partiellement sclérosés) afin de déterminer en temps réel l'état d'un rein selon la classification de Banff. Enfin, il serait intéressant de déployer la démarche en recherche clinique et d'évaluer son influence à travers des entretiens avec des pathologistes.

6 Remerciements

Projet ERACoSysMed "SysMIFTA", cofinancé par EU H2020 et les agences nationales de financement du minis-

tère allemand de l'Éducation et de la Recherche (BMBF) gestion de projet PTJ (FKZ : 031L-0085A), et Agence nationale de la recherche (ANR), numéro de projet ANR-15-CMED-0004.

Les auteurs tiennent à remercier le Centre de Calcul Haute Performance de l'Université de Strasbourg pour avoir soutenu ce travail en fournissant un soutien scientifique et un accès aux ressources informatiques. Une partie des ressources informatiques ont été financées par le projet Equipex Equip@Meso (Programme Investissements d'Avenir) et le CPER Alsacalcul / Big Data.

Références

- [1] 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, 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.
- [2] V. Andreoli Petrolini, E. Beckhauser, A. Savaris, M. Ines Meurer, A. von Wangenheim, D. Kreche, Collaborative Telepathology in a Statewide Telemedicine Environment-First Tests in the Context of the Brazilian Public Healthcare System, *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems*, pp. 684-689, 2019.
- [3] M. D. Bloice, P. M. Roth, A. Holzinger, Biomedical image augmentation using Augmentor, *Bioinformatics*, pp. 1367-4803, 2019.
- [4] N. Bodla, B. Singh, R. Chellappa, L. S. Davis, Soft-NMS-improving object detection with one line of code, *Proceedings of the IEEE international conference on computer vision*, pp. 5561-5569, 2017.
- [5] F. Chollet et al., Keras, 2015.
- [6] M. D. Ellison, M. A. McBride, S. E. Taranto, F. L. Delmonico, H. M. Kauffman, Living kidney donors in need of kidney transplants : a report from the organ procurement and transplantation network, *Transplantation*, Vol. 74, pp. 1349-1351, 2002.
- [7] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE confe-*

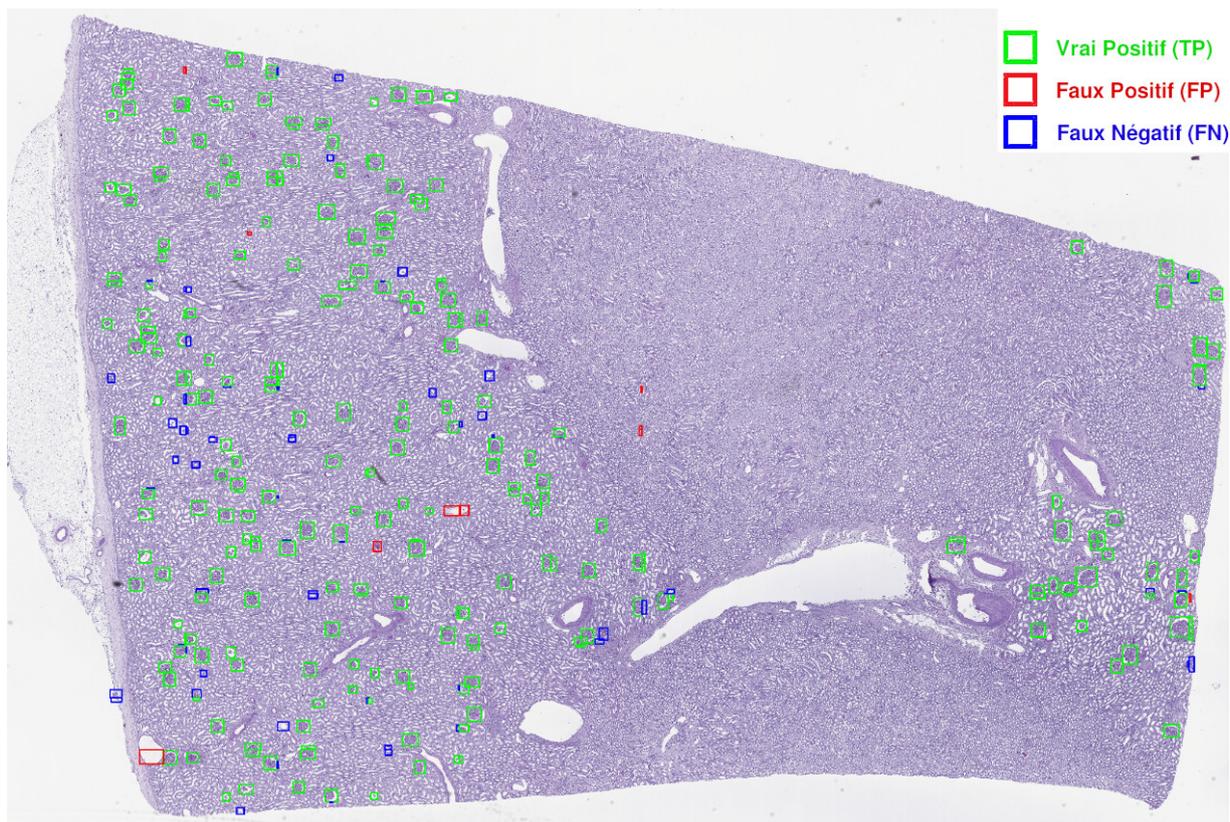


FIGURE 5 – Exemple de détection réalisée par le réseau entraîné et testé sur des images grossies 5X. Les images sont fusionnées pour reconstruire la WSI originale du patient 15. Les boîtes de délimitation vertes montrent les prédictions correctes faites par le réseau, les rouges sont une détection incorrecte et les glomérules manqués sont bleus.

rence on computer vision and pattern recognition, pp. 580-587, 2014.

- [8] R. Girshick, Fast R-CNN, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440-1448, 2015.
- [9] 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, pp. 955-962, 2019.
- [10] Y. Kawazoe, K. Shimamoto, R. Yamaguchi, Y. Shintani-Domoto, H. Uozaki, M. Fukayama, K. Ohe, Faster R-CNN-Based Glomerular Detection in Multistained Human Whole Slide Images, *Journal of Imaging*, Vol. 4, pp. 91, 2018.
- [11] 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, pp. 1-8, 2020.
- [12] T. Lampert, O. Merveille, J. Schmitz, G. Forestier, F. Feuerhake, C. Wemmert, Strategies for training stain invariant CNNs, *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 905-909, 2019.
- [13] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco : Common objects in context, *European conference on computer vision*, pp. 740-755, 2014.
- [14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980-2988, 2017.
- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, "Ssd : Single shot multibox detector, Ssd : Single shot multibox detector, *European conference on computer vision*, pp. 21-37, 2016.
- [16] R. Marée, L. Rollus, B. St évens, R. Hoyoux, G. Louppe, R. Vandaele, J.-M. Begon, P. Kainz, P. Geurts, L. Wehenkel, Collaborative analysis of multi-gigapixel imaging data using Cytomine, *Bioinformatics*, Vol. 32, pp. 1395-1401, 2016.
- [17] J. N. Marsh, M. K. Matlock, S. Kudose, T.-C. Liu, T. S. Stappenbeck, J. P. Gaut, S. J. Swamidas, Deep learning global glomerulosclerosis in transplant kidney frozen sections, *IEEE transactions on medical imaging*, Vol. 37, pp. 2718-2728, 2018.

- [18] Q. Peng, W. Luo, G. Hong, M. Feng, Y. Xia, L. Yu, X. Hao, X. Wang, M. Li, Pedestrian detection for transformer substation based on Gaussian mixture model and YOLO, *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, pp. 562-565, 2016.
- [19] qqwweee, keras-yolo3 : A Keras implementation of YOLOv3 (Tensorflow backend), 2019.
- [20] J. Redmon, S. Divvala, R. Girshick, A. Farhad, You only look once : Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.
- [21] J. Redmon, A. Farhad, Yolov3 : An incremental improvement, *arXiv preprint arXiv :1804.02767*, 2018.
- [22] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN : Towards real-time object detection with region proposal networks, *Advances in neural information processing systems*, pp. 91-99, 2015.
- [23] 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, pp. 288-300, 2016.
- [24] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, J. Liang, Convolutional neural networks for medical image analysis : Full training or fine tuning?, *IEEE transactions on medical imaging*, Vol. 35, pp. 1299-1312, 2016.