GLO-YOLO: A Dynamic Glomerular Detecting and Slicing Model in Whole Slide Images

Xueyu Liu
College of Data Science,
Taiyuan University of Technology
Taiyuan, Shanxi, China
liuxueyu1122@gmail.com

Ming Li
College of Data Science,
Taiyuan University of Technology
Taiyuan, Shanxi, China
lm13653600949@126.com

Fang Hao
College of Data Science,
Taiyuan University of Technology
Taiyuan, Shanxi, China
haofang12358@163.com

Guoye Zhang
Information Center,
Shanxi Provincial Science and Technology Department
Taiyuan, Shanxi, China
1963766@qq.com

Chen Wang
Department of Pathology,
The Second Hospital of Shanxi Medical University
Taiyuan, Shanxi, China
wangchen8877322@163.com

Xiaoshuang Zhou*
Department of Nephrology,
Shanxi Provincial People's Hospital
Taiyuan, Shanxi, China
xueyuliu1122@163.com

ABSTRACT

Renal pathology is the gold standard in renal clinical diagnosis. The glomerulus is the most important region of interest in the renal biopsy microscope slides. However, the development of renal pathology is limited by the doctor's laborious, time-consuming and repetitive detection of glomeruli. In this paper, we established GLO-YOLO to detect and slice glomerulus in whole slide image (WSI) of human renal biopsy microscope slides. In the training phase, GLO-YOLO improved the YOLO-v3 to improve the speed and efficiency of training. In the testing phase, we proposed dynamic scale evaluation, which can help models find the most appropriate evaluation scale for each high-resolution WSI. We collected and digitalized 1204 renal biopsy microscope slides to WSIs contained more than 42000 annotated glomeruli. The experimental results show the validity of dynamic-scale evaluation and improved YOLO-v3. GLO-YOLO has obtained state-of-the-art performance and outperformed the recently published approaches. Furthermore, GLO-YOLO can deploy in the renal intelligent auxiliary diagnosis system for renal clinical diagnosis by detecting glomeruli in high-resolution WSIs very quickly and effectively.

CCS CONCEPTS

• **Applied computing** → **Imaging**.

KEYWORDS

Glomerulus, Detecting and Slicing, Dynamic-scale Evaluation, GLO-YOLO, Whole Slide Image

CAIH2020, October 23–25, 2020, Taiyuan, China

ACM ISBN 978-1-4503-8864-1/20/10…$15.00
https://doi.org/10.1145/3433996.3434038

ACM Reference Format:
Xueyu Liu, Ming Li, Fang Hao, Guoye Zhang, Chen Wang, and Xiaoshuang Zhou. 2020. GLO-YOLO: A Dynamic Glomerular Detecting and Slicing Model in Whole Slide Images. In *2020 Conference on Artificial Intelligence and Healthcare (CAIH2020), October 23–25, 2020, Taiyuan, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3433996.3434038

1 INTRODUCTION

Chronic renal disease in clinical medicine has the feature of substantial morbidity and high fatality, but it has a low awareness rate and prevention rate of society. Patients with chronic renal disease are growing at an annual rate of 13 percent. However, there are few qualified pathologists for the reasons of lacking pathologists' standardized training system, low wages, and high risk. What is worse, the subtle pathological changes and complicated pathogenesis of renal disease contribute to a serious shortage of renal expert pathologists. In addition, the diagnostic process has a heavy workload and the diagnosis results greatly depend on the renal pathologists' professional knowledge and experience. Medical staff urgently need relevant intelligent diagnostic technology to diagnose renal disease since the backward traditional medical-seeking pattern and a large population of chronic renal disease patients.

It is an essential step in the diagnosis of chronic renal disease to observe pathological changes in the renal biopsy microscope slides by the optical microscope. In addition, the glomerulus is the most important region of interest in the renal biopsy microscope slides. And the number of glomeruli is a significant medical diagnostic indicator of many chronic renal diseases. However, the statistical work of glomeruli in the pathological biopsy slides is extremely complicated and the location of the glomeruli is not fixed, the glomeruli are easy to be missed or repeated. The development of renal pathology is greatly limited by these inefficient heavy works and current medical situation. The rapid development of digital pathology has effectively solved the similar problems in some areas.

Digital pathology aims to intelligent diagnose and quantify disease based on image data obtained by scanning pathological tissue samples. Great progress were made in the retinal fundus image [1], skin image [2]. While in the pathology, the researches are mainly

about the segmentation and recognition of cancer [3, 4]. The implementation of these technologies in the clinical diagnosis and their impact on medical staff aroused great interest [5]

In this paper, we built renal pathological datasets by preliminary works such as digitalize slides, data annotated. We proposed dynamic-scale evaluation to process high quality WSIs efficiently. And we established the glomerular detecting model GLO-YOLO by developing an improved CNN model and dynamic-scale evaluation. GLO-YOLO can detect all glomeruli in renal pathological 400× WSIs according to image sizes of different WSIs, and it can slice them respectively at a lossless resolution of 0.25μm/pixel. The speed of processing a WSI is around 10 seconds. This fast and accurate model is a valuable application in online clinical auxiliary diagnosis.

2 RELATED WORK

In recent years, computer version technology achieved some gratifying results. Several studies also applied computer version technology to process renal pathological images. Most researches firstly scanned renal biopsy microscope slides to WSIs by digital slide scanner. Then they constructed models to process WSIs, such as object detecting on the region of interest (glomeruli, renal tubules, interstitial cells, etc.). On the one hand, the feature extraction methods in images are defined manually by the machine learning algorithms in the computer version. Kato et al. [11, 12] introduced a method of detecting glomeruli. They used the histogram of oriented gradient (HOG) [6] and segmental-HOG for increasing robustness in the different glomerular morphology to extract features. HOG is an operator of focusing on the structure or shape of the object in the image, which is widely used in object detection. The HOG divides the image into rectangular areas of appropriate size (called cells) and calculates the intensity gradient in each cell through the histogram. The HOG feature of a cell is obtained by normalizing the value of the histogram in the block, which composed of adjacent cells. Simon et al. [13] extracted glomerular features by local binary pattern (LBP) [7]. LBP is a kind of operator to describe the local texture feature of the image. In the window of 3×3, LBP takes the pixel in the center of the window as the threshold, and compare it with the gray value of eight adjacent pixels to obtain the texture feature of this region. Then they both used the support vector machine (SVM) [16] as classifier to train models. The traditional methods do not have good generalization for detecting heterogeneous glomeruli, and they are easily influenced by different colors and object sizes.

On the other hand, deep learning is an effective technique to simulate the signal transduction process of nerve cells. Besides, the convolutional neural network (CNN) is specially designed to process images, which has a high correlation with adjacent data. Utilizing the automatic feature extraction method, SPPnet [8], Faster R-CNN [9], You Only Look Once (YOLO) [10] and other deep learning methods extract feature through a large number of data to have a better generalization of object deformation and shrinkage. Gallego et al. [14] proposed a method of slicing WSIs to patches by using a fixed-size sliding window. Then they classified the patches which were glomeruli and background by convolutional neural networks (CNN). Yoshimasa et al. [15] firstly used object detecting model, Faster-RCNN [9], to detect the glomeruli in WSIs with multiple stains of human renal biopsy slices, which improved performance significantly. Liu et al.[21] improved the Faster-RCNN algorithm to achieve better results. However, in these researches, the performance of glomerular detecting has not reached a relatively ideal level yet, and the algorithms are both slow to process WSIs at very high resolution. Moreover, there is not research about automatic slicing glomeruli in 400× WSIs since it is difficult and time-consuming to process it above 100 MB in size. There is not a suitable method for renal clinical auxiliary diagnosis up to now.

3 MATERIALS AND METHODS

3.1 Data Collection

1250 renal biopsy microscope slides with multiple stains were collected in the Second Hospital of Shanxi Medical University (SHSXMU) from 2014 to 2019 and in the Shanxi Provincial People's Hospital (SXPPH) from 2017 to 2019. In this research, we used periodic acid silver methenamine (PASM) stains. It can show more details of mesangial and basement membranes. Slides were digitized to WSIs by using a KF-PRO-005-EX digital slide scanner (KFbio, Ningbo, China) with a 40× objective and 10× eyepiece at a resolution of 0.25μm/pixel and stored in KFB format. Then they were converted to SVS format at full resolution. 1204 well preserved and the complete digital slides were finally used after evaluating and selecting.

3.2 YOLO-V3

You Only Look Once (YOLO) [10] is an efficient object detection algorithm, which handles the problem with a regression analysis based on darknet deep learning framework. Its source code is written in Program C and CUDA, which can give full play to the multi-core processor and GPU parallel computing function. In YOLO's feature extraction, it uses the global region of the image to process directly, which could better distinguish the target from the background, instead of using region proposal in two-step algorithm represented by Faster R-CNN [9]. This method is easy undetected and location error since the weakly regression loss function. YOLO-v3 [17] has a better performance through adding feature fusion of multiple scales and changing the backbone to darknet53. The performance of darknet53 is similar to ResNet [18] but much faster. It also works more effective for smaller objects by introducing anchors and small receptive fields. In this research, the processing speed is important since the model of this study would be deployed in the hospital auxiliary diagnosis system. Otherwise, detecting small objects is an essential ability for our model because the glomerulus is very small compared to the size of the entire image. Therefore, our research is based on YOLO-v3, which has high performance in small objects and high speed. However, high-resolution WSIs and complex background make YOLO-v3 difficult to process pathological images. Even though some researches validated YOLO-v3's effectiveness of detecting tissues in ultrasound, CT images, microscopic images, etc. [19, 20], it has not been applied to pathological images yet.

3.3 Dynamic Scale Evaluation

Medical images such as WSI mostly have large image sizes, so they cannot be processed directly by computers. In the training phase, we

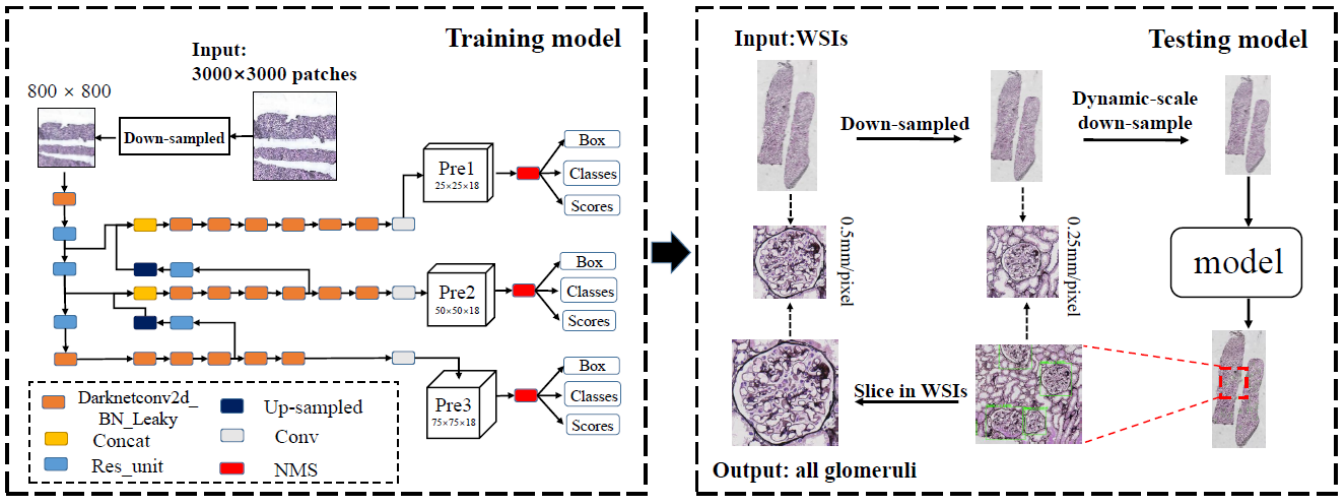

Figure 1: Training and testing process of GLO-YOLO.

solved this problem by cutting WSIs into patches. The image size of the 400× WSIs are presented by $S = w \times h$. Then the 400× WSIs were down-sampled to the 200× WSIs which image size can be presented by $S' = w' \times h'$. Then 200× WISs were cut by overlapping steps to the patches for the image size can be presented by $S_{patch} = n^2$.The training dataset of neural network is composed of patches. Since the input image is required to be multiple of 32 in YOLO-v3, we down-sampled patches to $S_{train} = (32m)^2$ for training. Obviously, the model has a good ability to detect glomerulus on the scale of S_{train}. However, in the evaluation phase, glomeruli in patches are easily duplicate detected in over-lapping portions. The model also confuses about incomplete glomeruli caused by morphological and caused by cutting patches (Figure 2).

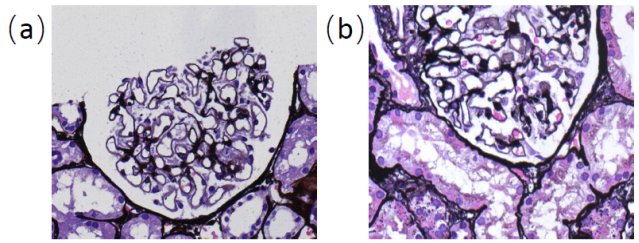

Figure 2: (a) is Glomerular morphological incompletion. (b) is incompletion caused by cutting.

Incomplete glomeruli caused by cutting patches need to be abandoned although they are annotated in the training dataset. Therefore, we would like to evaluate the entire WSI rather than patches. However, we could not find a proper constant scale to directly down-sample the entire WSIs because the image size of each WSI is very different. Therefore, we proposed a dynamic-scale evaluation method to process high-resolution images such as WSIs. This method ensures the accuracy and speed of detection.

$$32 \cdot (\alpha - 1) < \frac{w' \times 32m}{n} < 32 \cdot (\alpha + 1) \; \alpha \in \mathbb{Z} \tag{1}$$

$$w^i = 32 \cdot \alpha \tag{2}$$

$$32 \cdot (\beta - 1) < \frac{w' \times 32m}{n} < 32 \cdot (\beta + 1) \; \beta \in \mathbb{Z} \tag{3}$$

$$w^i = 32 \cdot \beta \tag{4}$$

3.4 Glomerular Detecting Model

In this section, we construct a deep learning model to detect glomeruli fast and accurately.

Training Phase. The features are extracted by darknet53, which uses the ResNet [18] residual structure can make the network structure deeper. Three different scale feature maps (25, 50, 75) are generated after feature extraction according to the size of the input image (800).

Testing phase. 400× WSIs at the resolution of 0.25μm/pixel are down-sampled to the 200× WSIs at the resolution of 0.50μm/pixel. Then 200× WSIs are dynamic-scale down-sampled to input testing images. The dynamic-scale coefficients are calculated from the input image size (800) in training phase. Input testing images have the best performance of the model since they have same scale as the training images. Test results are dynamic-scale up-sampled to 200× WSIs with glomerular bounding box by same dynamic-scale coefficients.

Slicing glomeruli in WSIs. Lossless glomeruli are needed in our detection. We sliced glomeruli from 400× WSIs to preserve significant image information. The glomeruli are sliced by corresponding bounding box in 200× WSIs and relative positions in 400× WSIs rather than up-sampled 200× WSIs directly. It can reduce up-sampled calculation in the other areas without glomeruli to speed up the algorithm. The process of model is shown in Figure 1.

4 EXPERIMENTS AND RESULTS

4.1 Dataset

955 slides from two hospitals and in different years were used as a train/validation dataset. Multiple sources ensure the universally applicable of training model. 400× WSIs are used have large image sizes for their high quality. The average image size is about 20000 × 20000 pixels. They are too large to be directly trained by the convolutional neural network (CNN). Therefore, we down-sampled the 400× WSIs at the resolution of 0.25μm/pixel to 200× WSIs at the resolution of 0.50μm/pixel, the average image size is about 8000 × 8000 pixels, to ensure the processing efficiency by openslide[23]. Then 200× WSIs were cut by overlapping steps to 5646 patches with 3000 × 3000 pixels. Patches with the resolution of 0.50μm/pixel are clear enough to identify the glomerulus in renal biopsy for doctors. All glomeruli, including sclerosing glomeruli, were boxed in the patches by using labelme [22] with three pathologists. Annotations were checked with two chief pathologists by deleting wrong annotations, adding missing annotations and adjusting the scope of the annotations boxes. The whole dataset includes 35491 glomeruli finally. Process of making training dataset is shown in Figure 3.

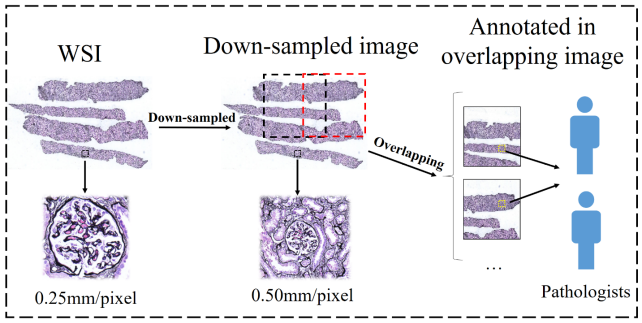

Figure 3: Process of making training dataset.

4.2 Training Parameter Settings

We built the training model of glomerular detecting using a Linux server with Single NVIDIA Tesla K80. Two models have the same training parameters. Anchors were renewedly designed to use prior knowledge. The anchor boxes are set to (10, 13), (29, 25), (40, 35), (50, 39), (64, 66), (55, 47), (116, 90), (156, 198), (373, 326) based on k-means. They were clustered on the width and height of the ground truth on the training dataset after down-sampled to 800 × 800. Carefully designed smaller boxes could achieve great performance since the glomerulus is generally smaller. Then we down-sampled the 3000×3000 patches in the training dataset to the image size of 800×800. It ensures the speed of training and not affects the general glomerular morphological characteristics. We applied data augmentation by changing saturation, exposure and hue for training the network. With regard to hyperparameter setting, the maximum batch is set to 52000. The initial learning rate is set to 7.5^{-3}, decrease to 7.5^{-4} after 40000 batches and decrease to 7.5^{-5} after 45000 batches.The batch size is assigned as 8, and subdivisions is set to 4 to save memory. Weight decay and momentum is set to 2.5^{-4} and 0.9 respectively. The loss curves of the model during training are shown in Figure 4.

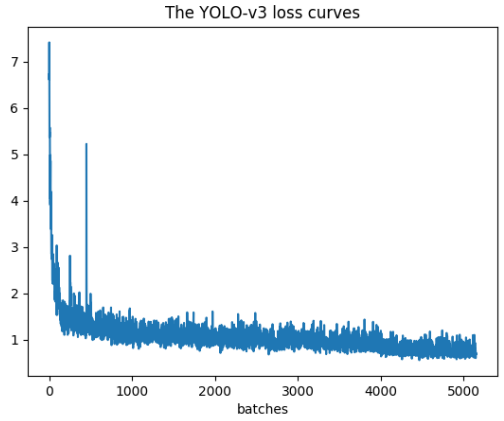

Figure 4: Loss curves of the training model.

4.3 Performance with Different Structure

We set up controlled experiments which evaluated 250 WSIs in testing dataset by using 2 kinds of models to verify the performance of GLO-YOLO. The image size was defined to 800 × 800 in models without dynamic-scale evaluation. GLO-YOLO has been evaluated in testing datasets. The performance of the model is shown in Table 1. F_1 is F-measures which is the harmonic average of precision and recall rates. The definition of precision, recall and F-measures are as follows,

$$Precision = \frac{TP}{TP + FP}, \tag{5}$$

$$Recall = \frac{TP}{TP + FN}, \tag{6}$$

$$F_1 = \frac{2Recall \times Precision}{Recall + Precision}, \tag{7}$$

which TP, FP and FN are true positive, false positive and false negative respectively. True positive means that detecting the glomerulus correctly detect into the glomerulus. False positive means that detecting other object such as background mistakenly detect into the glomerulus. We validated the effective of dynamic-scale evaluation. The performance improves significantly with appropriate evaluation scale. Furthermore, we compared our research to recently published approaches. GLO-YOLO outperformed relevant proposed methods and basic models with F1-measures of 0.966.

4.4 Processing Speed

GLO-YOLO is deployed in intelligent auxiliary diagnosis system. Processing speed is also one of the important evaluation indexes of the model. The minimum, maximum and average time of processing one WSI in Yoshima's work were 2s, 364s and 64s, respectively. GLO-YOLO has a faster speed to process high resolution WSIs, the minimum, maximum and average time of processing one WSI are 1.93s, 23.00s and 8.58s respectively. We also analyze the processing

Table 1: Performance with Different Models

Author	Model	Recall	Precision	F_1
Kato et al.[12]	RHOG+SVM	0.911	0.777	0.838
Simon et al.[13]	mrcLBP+SVM	0.761	0.917	0.832
Kawazoe et al.[15]	Faster R-CNN with inception v2	0.929	0.977	0.952
Liu et al.[21]	Faster R-CNN with ResNet	0.918	0.939	0.928
Our experiments	YOLOv3	0.712	0.983	0.823
	GLO-YOLO	**0.944**	**0.991**	**0.966**

speed of other models. Detailed results are shown in Table 2. Models with dynamic-evaluation are apparently more efficient since models can find the most appropriate scale for each WSI. Recall rate of models without dynamic-evaluation declined obviously by improper compression for lager size of WSIs, although the maximum processing time is shorter.

Table 2: Results of Processing Speed

Model	Average	Max	Min	Per glomerulus
Faster R-CNN	64s	364s	2s	N/A
YOLO-v3	12.46s	**20.78s**	8.69s	0.74s
GLO-YOLO	**8.58s**	23.00s	**1.93s**	**0.37s**

5 CONCLUSION

In this paper, we proposed an efficient glomerular detection and slicing model, GLO-YOLO, which combines improved YOLO-v3 and dynamic-scale evaluation method. We constructed a dataset of 1204 high-quality renal pathological WSIs with PASM stain containing 41440 glomeruli. GLO-YOLO has higher average F-measures in PASM stains compared with other recently published approaches. What's more, GLO-YOLO processes one WSI and slices the all glomeruli only in an average of 8.58 seconds. It detects and slices one glomerulus only in an average of 0.374 seconds. The high-speed and high-precision method provide the foundation for the deployment in the intelligent auxiliary system for clinical detecting. By the way, we have sliced all glomeruli in 1204 WSIs by using GLO-YOLO to do further research. It is about intelligent diagnosis of glomerulopathy.

ACKNOWLEDGMENTS

The authors sincerely acknowledge the doctors who offering great assistance in the Second Hospital of Shanxi Medical University and in the Shanxi Provincial People's Hospital. We also sincerely acknowledge the editor and any reviewers for their valuable advice.

REFERENCES

[1] Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathyin Retinal Fundus Photographs. JAMA. 2016;316:2402-2410.
[2] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer deep neural networks. Nature. 2017;542:115-118.
[3] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J Pathol Inform. 2016;7:29.
[4] Cabitza F, Rasoini R, Gensini G. Unintended Consequences of Machine Learning in Medicine. JAMA. 2017;318:517-518.
[5] Coudray, N., Ocampo, P.S., Sakellaropoulos, T. *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* **24,** 1559-1567 (2018)
[6] 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, San Diego, CA, USA, 20-25 June 2005; Volume 1.
[7] Zhang B, Gao Y, Zhao S, et al. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor[J]. IEEE transactions on image processing, 2009, 19(2): 533-544.
[8] He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. arXiv 2014, arXiv:1406.4729.
[9] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE T rans. Pattern Anal. Mach. Intell. 2017, 39, 1137-1149.
[10] Joseph, R.; Santosh, D.; Ross, G.; Ali, F. You Only Look Once: Unified, Real-Time Object Detection. arXiv 2015, arXiv:1506.02640.
[11] Kakimoto, T.; Okada, K.; Hirohashi, Y.; Relator, R.; Kawai, M.; Iguchi, T.; Fujitaka, K.; Nishio, M.; Kato, T.; Fukunari, A.; et al. Automated image analysis of a glomerular injury marker desmin in spontaneously diabetic Torii rats treated with losartan. J. Endocrinol. 2014, 222, 43-51.
[12] Kato, T.; Relator, R.; Ngouv, H.; Hirohashi, Y.; Takaki, O.; Kakimoto, T.; Okada, K. Segmental HOG: New descriptor for glomerulus detection in kidney microscopy image. BMC Bioinform. 2015, 16, 316.
[13] 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.
[14] 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.
[15] Kawazoe Y , Shimamoto K , Yamaguchi R , et al. Faster R-CNN-Based Glomerular Detection in Multistained Human Whole Slide Images[J]. 2018.
[16] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural processing letters, 1999, 9(3): 293-300.
[17] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[18] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Thirty-first AAAI conference on artificial intelligence. 2017.
[19] George J, Skaria S, Varun V V. Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans[C]//Medical Imaging 2018: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2018, 10575: 105751I.
[20] Wang X, Xu T, Zhang J, et al. SO-YOLO based WBC detection with Fourier ptychographic microscopy[J]. IEEE Access, 2018, 6: 51566-51576.
[21] Liu R, Wang L, He J, et al. Towards Staining Independent Segmentation of Glomerulus from Histopathological Images of Kidney[J]. bioRxiv, 2019: 821181.
[22] Russell B C, Torralba A, Murphy K, et al. LabelMe: A Database and Web-Based Tool for Image Annotation[J]. International Journal of Computer Vision, 2008, 77(1): 157-173.
[23] Goode, Adam Gilbert, Benjamin Harkes, Jan Jukic, Drazen Satyanarayanan, Mahadev. (2013). OpenSlide: A vendor-neutral software foundation for digital pathology. Journal of pathology informatics. 4. 27. 10.4103/2153-3539.119005.