HRGC-YOLO for Urine Sediment Particle Detection in High-Resolution Microscopic Images

1st Yunqi Zhu
Department of Biomedical Engineering
Zhejiang University
Hangzhou, China
yunqizhu@zju.edu.cn

2nd Haixu Yang
Department of Biomedical Engineering
Zhejiang University
Hangzhou, China
yanghaixu@zju.edu.cn

3rd Luhong Jin
Department of Biomedical Engineering
Zhejiang University
Hangzhou, China
lhjin@zju.edu.cn

4th Dagan Yang
Department of Laboratory Medicine
Zhejiang University School of Medicine
Hangzhou, China
yangdagan@zju.edu.cn

5th Yu Chen
Department of Laboratory Medicine
Zhejiang University School of Medicine
Hangzhou, China
chenyuzy@zju.edu.cn

6th Xianfei Ye
Department of Laboratory Medicine
Zhejiang University School of Medicine
Hangzhou, China
yeahxf@zju.edu.cn

7th Sergey Ablameyko
Mechanics-mathematical faculty
Belarusian State University
Minsk, Republic of Belarus
ablameyko@bsu.by

8th Yingke Xu
Department of Biomedical Engineering
Zhejiang University
Hangzhou, China
yingkexu@zju.edu.cn

***Abstract*—The automatic detection of urine sediment particle (USP) in microscopy images plays a vital role in evaluating renal and urinary tract diseases. Convolutional neural networks (CNN)-based object detectors have demonstrated remarkable precision in end-to-end detection. However, directly applying CNN-based detectors to high-resolution USP microscopic images poses two major challenges: classification confusion and underutilization of fine-grained information. To address these problems, we present a novel High-Resolution Global Context (HRGC)-YOLO model, which based on YOLOv5m structure and incorporates a global context (GC) block to capture long-range dependencies. Meanwhile, we employ a tile-based detection approach to leverage the uncompressed fine-grained information in high-resolution images. We evaluated the performance of HRGC-YOLO on high-resolution USP datasets from clinic. Compared to YOLOv5m, our HRGC-YOLO network achieved a 4.5% improvement in mAP and outperformed all tested YOLO series models. Our results demonstrate the effectiveness of the proposed method in accurately detecting USPs in high-resolution images.**

Keywords—Deep learning, Object detection, Urine sediment, Global context, Tile-based image processing

I. Introduction

The microscopy imaging and image analysis of visible urine sediment components play a pivotal role in diagnosis of renal and urinary tract diseases [1]. With the increasing demand from the clinic, the need for automated and efficient detection of particle instances from microscopic images has become urgent. Vast quantities of microscopic images are generated in hospitals on a daily basis, necessitating advanced methods to accurately identify and categorize urine sediment particles (USPs).

Computer vision assisted USP detection has transitioned from multistage methodologies to end-to-end approaches. Prior to the widespread adoption of Convolutional Neural Networks (CNNs) for object detection, the detection processes for USPs constitute discrete steps, such as the region of interests proposal [2, 3], feature extraction [4], and classification [5]. More recently, CNN-based object detection has witnessed rapid development, enabling swift and accurate detection outcomes. For example, Liang et al. [6] proposed improvements to the Faster Region CNN (R-CNN) [7] framework to make accurate detection results in USP images. Additionally, Derya et al. [8] merged Faster R-CNN with super-resolution reconstruction methods and image denoising techniques to accurately recognize USPs in low-resolution medical images. Besides the two-stage object detection models exemplified by Faster R-CNN, one-stage object detection models such as the You Only Look Once (YOLO) series [9–11] have also been extensively utilized in USP detection [12–15]. The YOLO networks have efficient network architecture that eliminates the need of a region proposal network and converts object detection into a single regression problem.

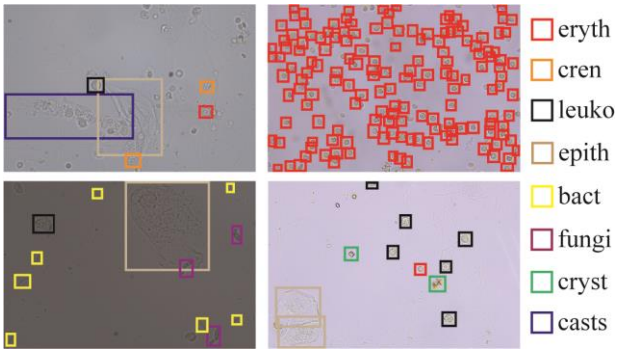

Fig. 1. Illustration of the classes and distribution of USPs labeled in high resolution images. This study specifically examined eight distinct types of USPs, namely: erythrocyte (*eryth*), crenated erythrocyte (*cren*), leukocyte (*leuko*), epithelial cells (*epith*), bacteria (*bact*), *fungi*, crystals (*cryst*) and *casts*.

Current methods for detecting USP images face two major challenges. Firstly, USPs exhibit intra-class variation and inter-class similarity [16], leading to classification confusion [6]. Secondly, fine-grained information in high-resolution (HR) USP images is inevitably underutilized. This is due to the constraints of computational complexity that requires compression before feeding the HR images into network. Previous researches have shown that attention modules are capable to ensure networks to focus only on the pertinent information [17]. It is an effective strategy to enhance the performance of network. However, conventional approaches mainly paid attention to the prospect of attentional modules in extracting local information or combining channel

Corresponding author: Yingke Xu

Y. Zhu and H. Yang contributed equally to this work

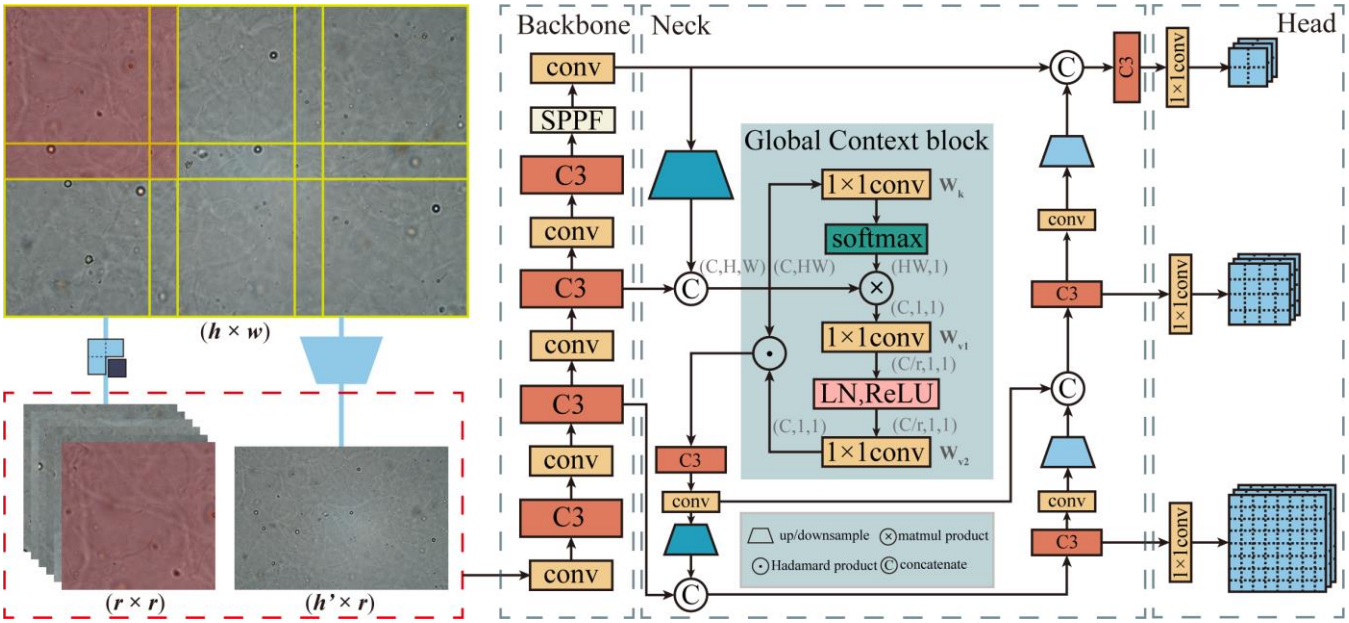

Fig. 2. Schematic of the tile-based approach (left) and the HRGC-YOLO model structure (right). The input image will be cropped into several square patches, and fed into network along with the compressed original image. The network was modified from YOLOv5m. We added a GC block in the neck network to help capture long-range dependencies.

relationship (later in this article we will refer to them as local attention modules for convenience). Neglecting the importance of long-range dependencies will cause the detection performance degeneration. Moreover, existing approaches primarily rely on public datasets consisting of only low-resolution USP images in JPEG format, lacking evaluation on uncompressed images directly captured from HR cameras.

To address these two issues, we propose a novel High-Resolution Global Context (HRGC)-YOLO model. We integrated the long-range dependency capturing module, GC block [18], into the original YOLOv5m architecture. Additionally, we introduced a tile-based image processing approach to effectively exploit the fine-grained information in HR images. To evaluate its performance, we built a comprehensive dataset comprising 1,278 HR, high-fidelity images with eight classes of USPs (Fig. 1). The proposed method achieved a 4.5% improvement than YOLOv5m in mean Average Precision (mAP), and significantly surpassed other tested YOLO series models.

II. HRGC-YOLO Network

The HRGC-YOLO model is a modified version of the YOLOv5m model. It is specifically designed to accommodate the unique characteristics of HR USP images. By incorporating the GC blocks in the neck network, the HRGC-YOLO is capable of capturing the global dependencies of the images. To retain fine-grained information, we employed a tile-based approach for data processing. In this section, we will provide a description of the classic YOLOv5 model and detailed information of the two integrated approaches in our developed model.

A. YOLOv5 architechture

The architecture of YOLOv5 is shown in Fig. 2 and can be divided into three parts: backbone, neck, and detection head. The backbone is a deep CNN and serves as the feature extractor. It plays an important role in capturing hierarchical features of different scales. The neck is responsible for combining features from different depths of the backbone network. In YOLOv5, the neck utilizes pyramidal feature hierarchies to aggregate features with various resolutions and to enhance capability of the model to detect objects at different scales. The detection head is responsible for predicting bounding boxes and class probabilities.

B. Global context block

As mentioned earlier, the attention modules have been proven its effectiveness in alleviating classification confusion caused by intra-class variation and inter-class similarity. However, the local attention modules in recently proposed USP detectors, such as the Convolutional Block Attention Module (CBAM) [19], predominantly focus on local spatial information and overlook the significance of long-range dependencies. We observed that by incorporating non-local attention-like modules into the network, it will improve the detection accuracy more effectively. This is because the non-local module has the ability to integrate and analyze regions of interest across the entire image. Therefore, it enables the model to compare confusing objects with other similar objects in the image, resulting in more reliable and reasonable results.

In the HRGC-YOLO, we introduced a global context (GC) block [18] (Fig. 2), which can be defined as a combination of simplified non-local block and squeeze-excitation (SE) block [20]. The mathematic mechanism of GC block is illustrated as follows:

$$z_i = x_i \times W_{v2} ReLU\left(LN\left(W_{v1}\sum_{j=1}^{N_p}\frac{exp\left(W_{kx_j}\right)}{\sum_m exp\left(W_{kx_m}\right)}x_j\right)\right) \quad (1)$$

where $\frac{\exp(W_{kx_j})}{\sum_m \exp(W_{kx_m})}$ represents the weight of global attention pooling, which corresponds to the softmax function in Fig. 2. z_i and x_i are the output and input of the GC block, respectively. $ReLU$ stands for Rectified Linear Unit, and LN stands for Layer Normalization. W denotes a 1×1 convolution, and the scaling factor r of W_{v1} is empirically set to 8 in this

work. N_p represents the number of positions in the feature map.

Since GC block is a relatively lightweight and powerful module, we applied this module solely at the front end of the neck network in our HRGC-YOLO model. As a result, the computational requirements of HRGC-YOLO only have increased from around 47.95 GFLOPs (Giga Floating Point of Operations) to about 48.07 GFLOPs, corresponding to a slightly increase of 0.25%. Moreover, the introduction of this module results in merely 0.65% increase in the number of model parameters (refer to Table II for more information).

C. Tile-based image process

One challenge in training the HRGC-YOLO model is to design a procedure to handle the HR images. Directly input HR images into a CNN would cause memory overflow, as well as slow down the training process. Since inputs with high resolution will significantly increase the time requirement of some complex data augmentation procedures. Meanwhile, simple and rough images rescaling may lose the essential fine-grained information.

Therefore, we adopt a tile-based image processing approach (Fig. 2). It allows the model to perceive both the entire image and local information. In addition, it can minimize the computational resource and preserve the fine-gained information at the same time. As illustrated in Fig. 2, we crop the original image with size (h, w) into several square patches, each with a width of r. During the training procedure, these patches, along with the compressed original image, are fed into the network. During the inference phase, the detection algorithm summarizes the prediction boxes of the patches and the original image, followed by uniform non-maximum suppression to obtain the final detection result. In this way, our HRGC-YOLO is able to practically use the fine-grained details and maintaining a high computational efficiency.

III. Experiments and Results

In this section, we present a series of experiments conducted on manually built dataset to evaluate the performance of the proposed HRGC-YOLO model. Firstly, to show the performance of the non-local attention module, we compared the GC block with several other attention modules. Furthermore, we introduced the tile-based method and conducted a comparative analysis with other YOLO models to demonstrate the performance of HRGC-YOLO in HR USP image detection tasks.

A. Dataset preparation

The current public USP dataset contains only low-resolution JPEG format data, which may lose a lot of detailed information about the objects. Besides, previous work [21] mentioned that although CNNs are resilient to low level JPEG compression, but a high compression rate can still lead to a sudden decrease in their performance.

To conduct an evaluation of the HRGC-YOLO model, we manually built a dataset consisting of 1,278 HR USP images. These images were captured using a Leica DM500 microscope equipped with a Leica ICC50 camera at a magnification of 400. The majority of the images have dimensions of 5,440×3,648 and 4,000×3,000 in pixels[1] and were saved in TIFF format. Our dataset consists of 8 classes of particles, with a total of 32,968 particles manually labelled by clinical experts. Table I provides a summary of the dataset, illustrating the number of instances in each category. We divided the dataset into two sets: 903 images for training and 375 images for testing.

TABLE I. The Number, Size and Test Set Percentage of Each Class in Self-built Dataset

categories	number	size	test percentage
eryth	12931	30.1	32.0%
cren	1505	27.7	21.1%
leuko	5242	44.2	28.2%
epith	1087	179.1	32.0%
bact	5767	18.1	24.7%
mold	1669	36.6	34.2%
cryst	3363	40.1	28.9%
casts	800	341.2	31.0%

B. Experimental Settings

The HRGC-YOLO model was developed on the YOLOv5 open-source project with the help of PyTorch framework. Adam optimizer was used to optimize the parameters of all the models tested, and they were trained on a RTX 3090 GPU. We set the maximum number of training epochs to 300, and the size of the minibatch used for each experiment was determined by the maximum size the GPU can handle. The remaining hyperparameter were set to the default values of the official YOLOv5 project. During the training phase, the input images were scaled isometrically to a width of 1,280 pixels while maintaining the original aspect ratio. To ensure shape uniformity within each minibatch, gray padding was applied to the resized images.

For evaluation metrics, we adopt the mAP at intersection over union (IoU) threshold 0.5 (mAP50), consistent with previous studies in the field. The mAP50 and the Average Precision (AP) values for each specific type of USP in Table II and III present the optimal results obtained from three separate experiments, except for the results obtained from the tile-based method owing to the training time limitation. Fig. 3 visualizes the comparison of the detection results of our proposed method with the base model.

C. The importance of GC block

Next, we investigated the effectiveness of GC block, and compared it with four classic local attention modules (CBAM [19], BAM [22], SE [20], and EffectiveSE (ESE) [23]). All these five modules were plugged separately into the front end of the neck network of the YOLOv5m. For convenience, we named the YOLOv5m model that incorporates GC blocks as GC-YOLO. To be clear, the dataset used in this step was our original HR images without preprocessing with our tile-based image method. The detailed detection results were presented in Table II. In summary, the utilization of attention blocks consistently enhanced the detection accuracy of the model. Compared to other modules, the GC block demonstrated superior performance (+1.6% mAP50 to the base YOLOv5m model), especially the AP value of small target bacteria (*bact*) gained the most significant improvement (+9.1%).

[1] To clearly present the morphology and number of USPs, the images in Fig. 1 and Fig. 3 only show 1/4 from the original ones.

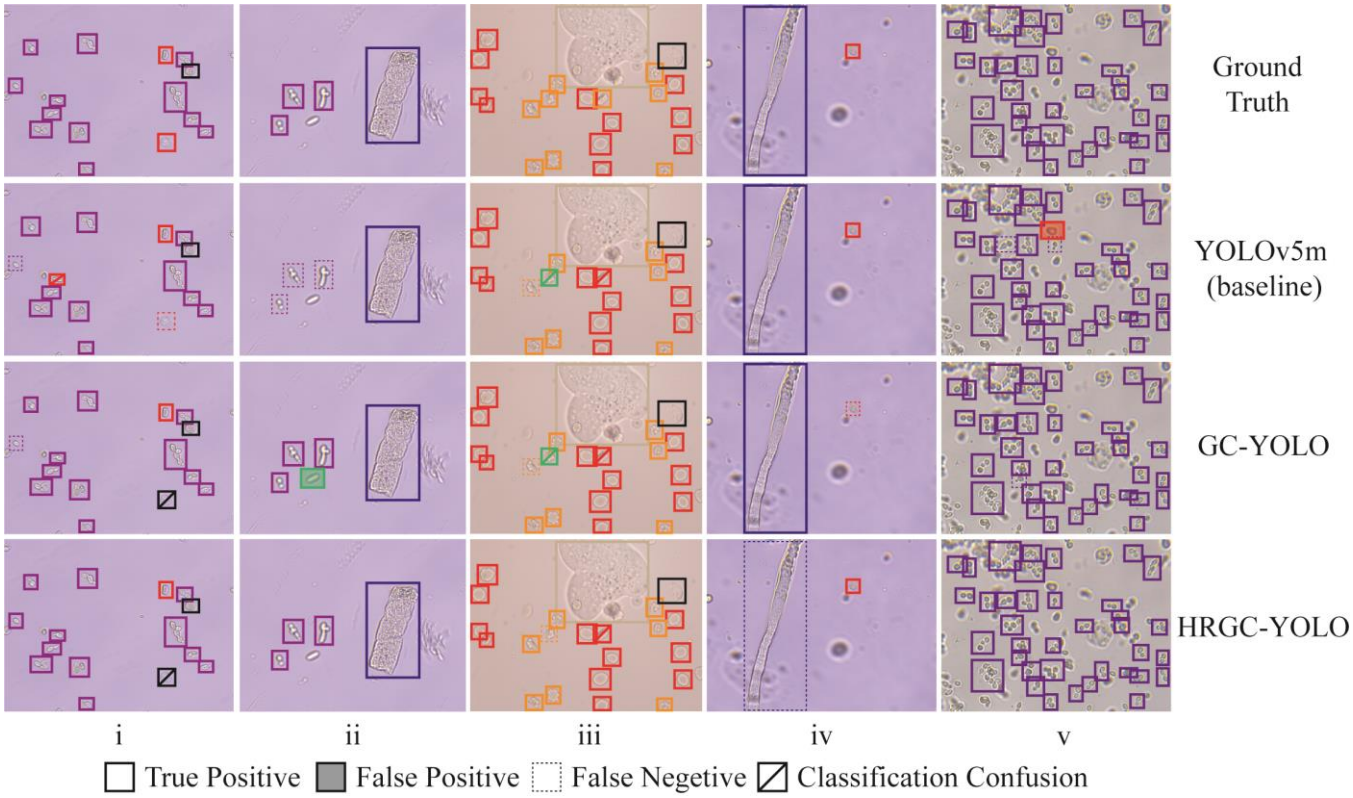

Fig. 3. The performance of our developed HRGC-YOLO, GC-YOLO and YOLOv5m in USPs detection task. The HRGC-YOLO outperforms the others, showing great ability in detecting small and dense objects. But its accuracy on large-sized objects is slightly lower than the GC-YOLO.

These results highlighted the advantage of the GC block in capturing global contextual information, particularly in the detection of large-size USP images. Moreover, the GC block introduced only a marginal number of additional parameters and GFLOPs, ensuring that the computational burden remains manageable in practical applications.

D. The outstanding detection performance of HRGC-YOLO

In this section we compared the detection performance of HRGC-YOLO with other detectors that commonly used in natural images. Here we chose three recently released updated YOLO models. The results of YOLO series are shown in Table III.

To further improve the detection accuracy of HRGC-YOLO, we additionally used the Focal-GIoU bounding box regression loss and SimOTA label assignment strategy during the training procedure. Compared to ordinary losses, GIoU loss [24] can provide greater robustness in processing changes in scale, rotation, and tilt of visible components in microscopic images. In addition, GIoU loss is also more effective in penalizing the occurrence of overlapping target boxes, particularly in situations involving dense targets. The GIoU loss can be represented as:

$$Loss_{GIoU} = 1 - IoU + \frac{|A_c - U|}{|A_c|} \quad (2)$$

Where A_c is the area of smallest enclosing rectangle of the ground truth (GT) and the prediction result. U is the overlap area between them. Inspired by [25], we further changed GIoU to Focal-GIoU:

$$Loss_{Focal-GIoU} = IoU^{\gamma} \times Loss_{GIoU} \quad (3)$$

γ is a parameter to control the degree of inhibition of outliers, which was empirically set to 0.5 in this study. Moreover, we applied the SimOTA label assignment strategy that was used in YOLOX [9] to the training process. It analyzes the label assignment task from a global perspective, and provides more accurate and effective assignment results. The experimental results showed that models combined with the Focal-GIoU loss and SimOTA strategy did promote the detection performance. In Table III, for the experiments did not use Focal-GIoU and SimOTA, we used the default bounding box regression loss function and label assignment strategy of the respective model.

Afterward, we tested the effectiveness of the tile-based strategy and presented the results of ablation study and comparison with the YOLO series. The basic tile-based preprocessing procedure increased the mAP50 values from the base YOLOv5m model by 2.1%. The most exciting conclusion is that our HRGC-Yolo outperformed all the other models in HR USP image detection. The detection performance reflected by the mAP50 value was improved by 4.5% compared to the base YOLOv5m model.

We noticed that the inclusion of the tile-based strategy in YOLOv5m and GC-YOLO reduced the detection accuracy for large-sized objects (*casts* and *epith* as shown in Fig. 3 iv). This phenomenon can be attributed to the slicing process, which diminished the field of view of the image. As a result, many large-sized objects were only partially retained in the cropped patches, and our algorithm did not consider them as positive samples. However, the tile-based strategy still exhibited a notable impact on the detection of small targets that are inherently challenging to detect. Additionally, it proved to be beneficial when dealing with images containing a high density of targets. The experimental results supported the notion that

TABLE II. THE COMPARISON OF GC BLOCK WITH OTHER ATTENTION MODULES

Methods	mAP (%)	AP for each USP category (%)								GFLOPs	Params (million)
		eryth	*cren*	*leuko*	*epith*	*bact*	*mold*	*cryst*	*casts*		
Base (YOLOv5m)	72.9	83.4	65.0	82.8	86.2	38.6	64.7	**84.5**	78.0	47.95	23.06
Base + CBAM [19]	73.7	84.0	63.1	84.1	**87.0**	45.8	64.7	83.4	77.7	48.07	23.13
Base + BAM [22]	73.6	84.1	**66.2**	82.4	85.4	44.0	65.4	84.1	77.3	48.34	23.24
Base + SE [20]	74.0	83.9	64.2	**84.7**	86.3	38.6	**75.1**	82.5	76.4	48.01	23.13
Base + ESE [23]	73.6	83.6	64.9	83.3	85.4	40.1	73.2	83.9	74.7	48.42	23.65
GC-YOLO (ours)	**74.5**	**84.6**	64.0	82.7	85.5	**47.7**	70.1	83.0	**78.6**	48.07	23.21

TABLE III. THE PERFORMANCE COMPARISON OF HRGC-YOLO WITH OTHER YOLO SERIES AND ABLATION STUDY

Methods	mAP (%)	AP for each USP category (%)							
		eryth	*cren*	*leuko*	*epith*	*bact*	*mold*	*cryst*	*casts*
Base (YOLOv5m)	72.9	83.4	65.0	82.8	86.2	38.6	64.7	84.5	78.0
YOLOv6m [10]	65.0	66.4	59.7	83.3	86.1	26.6	54.2	61.8	81.7
YOLOv7 [11]	72.1	83.8	59.2	84.5	86.7	43.5	66.9	82.7	69.2
YOLOv8m	66.1	59.1	33.7	82.4	63.3	37.7	50.5	76.7	62.8
Base + SimOTA [13] & Focal-GIoU [24, 25]	74.3	86.2	62.7	84.9	86.6	44.5	67.9	84.7	77.2
Base + Tile-based	75.0	85.4	62.6	83.0	85.2	48.5	70.2	89.9	75.3
GC-YOLO	74.5	84.6	64.0	82.7	85.5	47.7	70.1	83.0	78.6
+ SimOTA & Focal-GIoU	75.6 **(+1.1)**	86.0	**68.4**	**85.2**	**87.3**	44.4	68.3	84.4	**80.7**
+ Tile-based (HRGC-YOLO)	**77.4 (+1.8)**	**88.1**	65.4	85.0	86.7	**52.1**	**74.5**	**91.3**	75.7

by dividing the image into smaller patches, the model can effectively capture fine-grained details and improve detection performance in challenging scenarios such as HR USP images. Nonetheless, it is important to note that this approach also introduces a significant increase in computational burden, which needs to be further improved in the future.

IV. CONCLUTION

In this paper, we propose a novel object detection model HRGC-YOLO. It is specifically designed to address the challenges posed by HR USP image data. We employed a tile-based approach to preserve and utilize fine-grained information in large scaled images to ensure the detection accuracy for small objects. Additionally, the GC module integrated effectively captured long-range dependencies and enabled the model to handle complex spatial relationships within the image. The evaluation results on clinical collected dataset showed that HRGC-YOLO outperforms other object detectors. The ablation study further confirmed the significance of the tile-based procedure and the GC block we integrated in this model. In conclusion, our proposed method, HRGC-YOLO, demonstrates exceptional performance in accurately identifying and categorizing USPs. With its remarkable capabilities, HRGC-YOLO holds great promise as an indispensable diagnostic tool for renal and urinary tract diseases.

ACKNOWLEDGMENT

This work was supported by Zhejiang Provincial Natural Science Foundation (LZ23H180002 and LQ22F050018), National Key Research and Development Program of China (2021YFF0700305), Zhejiang University K.P.Chao's High Technology Development Foundation (2022RC009), and the Fundamental Research Funds for the Central Universities (226-2023-00091). We would like to thank S. Wang, Y. Liu and X. Fu for their assistance in conducting several experiments throughout this research.

REFERENCES

[1] C. Cavanaugh and M. A. Perazella, "Urine Sediment Examination in the Diagnosis and Management of Kidney Disease: Core Curriculum 2019," *American Journal of Kidney Diseases*, vol. 73, no. 2, pp. 258–272, Feb. 2019,

[2] Y.-M. Li and X.-P. Zeng, "A new strategy for urinary sediment segmentation based on wavelet, morphology and combination method," *Computer Methods and Programs in Biomedicine*, vol. 84, no. 2, pp. 162–173, Dec. 2006,

[3] Q. Wang, Q. Sun, and Y. Wang, "A two-stage urine sediment detection method," in *2020 International Conference on Image, Video Processing and Artificial Intelligence*, Nov. 2020, vol. 11584, pp. 15–21.

[4] X. Zhou, X. Xiao, and C. Ma, "A study of automatic recognition and counting system of urine-sediment visual components," in *2010 3rd International Conference on Biomedical Engineering and Informatics*, Oct. 2010, vol. 1, pp. 78–81.

[5] M. Shen and R. Zhang, "Urine Sediment Recognition Method Based on SVM and AdaBoost," in *2009 International Conference on Computational Intelligence and Software Engineering*, Dec. 2009, pp. 1–4.

[6] Y. Liang, Z. Tang, M. Yan, and J. Liu, "Object detection based on deep learning for urine sediment examination," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 3, pp. 661–670, Jan. 2018,

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017,

[8] D. Avci, E. Sert, E. Dogantekin, O. Yildirim, R. Tadeusiewicz, and P. Plawiak, "A new super resolution Faster R-CNN model based detection and classification of urine sediments," *Biocybernetics and Biomedical Engineering*, vol. 43, no. 1, pp. 58–68, Jan. 2023,

[9] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," arXiv, Aug. 05, 2021.

[10] C. Li *et al.*, "YOLOv6 v3.0: A Full-Scale Reloading," arXiv, Jan. 13, 2023.

[11] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pp. 7464–7475.

[12] Z. Chen *et al.*, "An Efficient Particle YOLO Detector for Urine Sediment Detection," in *Machine Learning for Cyber Security*, Cham, 2023, pp. 294–308.

[13] M. Yu, Y. Lei, W. Shi, Y. Xu, and S. Chan, "An Improved YOLOX for Detection in Urine Sediment Images," in *Intelligent Robotics and Applications*, Cham, 2022, pp. 556–567.

[14] H. Atici, H. E. Kocer, A. SiVriKaya, and M. Dagli, "Analysis of Urine Sediment Images for Detection and Classification of Cells," *Sakarya University Journal of Computer and Information Sciences*, vol. 6, no. 1, pp. 37–47, Apr. 2023,

[15] S. Dong, S. Zhang, L. Jiao, and Q. Wang, "Automatic Urinary Sediments Visible Component Detection Based on Improved YOLO Algorithm," in *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, Jul. 2020, pp. 485–490.

[16] M. Yan, Q. Liu, Z. Yin, D. Wang, and Y. Liang, “A Bidirectional Context Propagation Network for Urine Sediment Particle Detection in Microscopic Images,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 981–985.

[17] T. Gonçalves, I. Rio-Torto, L. F. Teixeira, and J. S. Cardoso, “A Survey on Attention Mechanisms for Medical Applications: are we Moving Toward Better Algorithms?” *IEEE Access*, vol. 10, pp. 98909–98935, 2022,

[18] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, “GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond,” in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Oct. 2019, pp. 1971–1980.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in *Computer Vision – ECCV 2018*, Cham, 2018, pp. 3–19.

[20] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 7132–7141.

[21] S. Dodge and L. Karam, “Understanding how image quality affects deep neural networks,” in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, Jun. 2016, pp. 1–6.

[22] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, “BAM: Bottleneck Attention Module,” arXiv, Jul. 18, 2018.

[23] Y. Lee and J. Park, “CenterMask: Real-Time Anchor-Free Instance Segmentation,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 13903–13912.

[24] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 658–666.

[25] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, “Focal and efficient IOU loss for accurate bounding box regression,” *Neurocomputing*, vol. 506, pp. 146–157, Sep. 2022,