2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10341092

A Key-Points Based Anchor-Free Cervical Cell Detector

Tong Shu[1], Jun Shi[2*], Yushan Zheng[3,5*], Zhiguo Jiang[4,5], Lanlan Yu[6]

***Abstract*— Cervical cell detection is crucial to cervical cytology screening at early stage. Currently most cervical cell detection methods use anchor-based pipeline to achieve the localization and classification of cells, e.g. faster R-CNN and YOLOv3. However, the anchors generally need to be pre-defined before training and the detection performance is inevitably sensitive to these pre-defined hyperparameters (e.g. number of anchors, anchor size and aspect ratios). More importantly, these preset anchors fail to conform to the cells with different morphology at inference phase. In this paper, we present a key-points based anchor-free cervical cell detector based on YOLOv3. Compared with the conventional YOLOv3, the proposed method applies a key-points based anchor-free strategy to represent the cells in the initial prediction phase instead of the preset anchors. Therefore, it can generate more desirable cell localization effect through refinement. Furthermore, PAFPN is applied to enhance the feature hierarchy. GIoU loss is also introduced to optimize the small cell localization in addition to focal loss and smooth L1 loss. Experimental results on cervical cytology ROI datasets demonstrate the effectiveness of our method for cervical cell detection and the robustness to different liquid-based preparation styles (i.e. drop-slide, membrane-based and sedimentation).**

I. INTRODUCTION

Cervical cytology has been the primary screening to detect and prevent cervical cancer. In practical application, pathologists generally need to detect the potential cervical abnormal cells from the slide and then make the diagnostic report according to the Bethesda System (TBS) [1]. The whole process is labor-intensive, time-consuming and relatively subjective. Therefore, the efficiency and quality of screening is heavily influenced.

To alleviate this problem, automatic cervical cell detection methods have been proposed to aid manual screening. In past few years, the object detection methods [2-5] based on convolution neural network (CNN) have been paid more attention and applied in cervical cytology. Du et al. [6] and Zhang et al. [7] introduce the two-stage object detector, Faster R-CNN [2], to achieve cervical cell detection. Xiang et al. [8] use one-stage object detector, YOLOv3 [9], to detect and classify the cervical cells. Liang et al. [10] propose a global context-aware framework based on YOLOv3 for cell-wise detection. Furthermore, Chai et al. [11] present a deep semi-supervised metric learning framework for cervical cell detection, which leverages labeled and unlabeled samples and takes Faster R-CNN as the base. Sompawong et al. [12] apply Mask R-CNN[5] to segment and identify the normal and abnormal cells. These CNN-based cervical cell detection methods effectively improve the efficiency and quality of manual screening. Essentially, they usually apply the anchor-based pipeline for the localization and classification of cells. The anchor boxes with different sizes and aspect ratios are taken as the detection candidates and then trained to determine whether they are the objects and their corresponding positions. However, these anchor boxes need to be defined in advance before training and thus the setting of anchor hyperparameters, e.g. number of anchors, anchor size and aspect ratios, is an open problem. Inevitably the detection effect is sensitive to these hyperparameters. More importantly, there exists the cells with different geometries and small cells which are difficult to detect. Therefore, the pre-defined anchors cannot effectively deal with this problem at inference phase.

Recently, the anchor-free detectors [13-15] have been proposed, which aims to directly regress the coordinates of objects through the neural network instead of using pre-defined anchor boxes. It can be roughly divided into two categories: center-based and keypoint-based. The center-based methods (e.g. YOLO [4] and FCOS [13]) usually use the center of object to define the positive sample and predict the distances between positive samples to the object bounding box for localization. On the other hand, the keypoint-based methods (e.g. CornerNet [14] and RepPoints [15]) generally represent the object by the specific points and convert them to the object bounding box. Notably, RepPoints uses a set of representative points to describe the object, which can capture the shape, pose and potential semantic position of objects. Consequently, it can generate finer localization effect and explore semantically significant local area. Overall, the anchor-free detectors reduce the computational cost caused by the hyperparameters of anchors and obtain better detection performance in terms of accuracy and speed. However, few work has been done to apply the anchor-free pipeline for cervical cell detection.

In this paper, we propose a key-points based anchor-free cervical cell detector (KPAFD) which uses YOLOv3 as the baseline and switched it to anchor-free mode. Different from the traditional YOLOv3 which uses the feature pyramid network (FPN) [3] to combine the multi-scale features, path aggregation feature pyramid network (PAFPN) [16] is introduced to strengthen the information propagation through bottom-up path augmentation and thus boost the hierarchical

*Corresponding author: juns@hfut.edu.cn; yszheng@buaa.edu.cn
[1]School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China
[2]School of Software, Hefei University of Technology, Hefei, China
[3]School of Engineering Medicine, Beihang University, Beijing, China
[4]Image Processing Center, School of Astronautics, Beihang University, Beijing, China
[5]Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing, China
[6]Motic (Xiamen) Medical Diagnostic Systems Co. Ltd., Xiamen, China

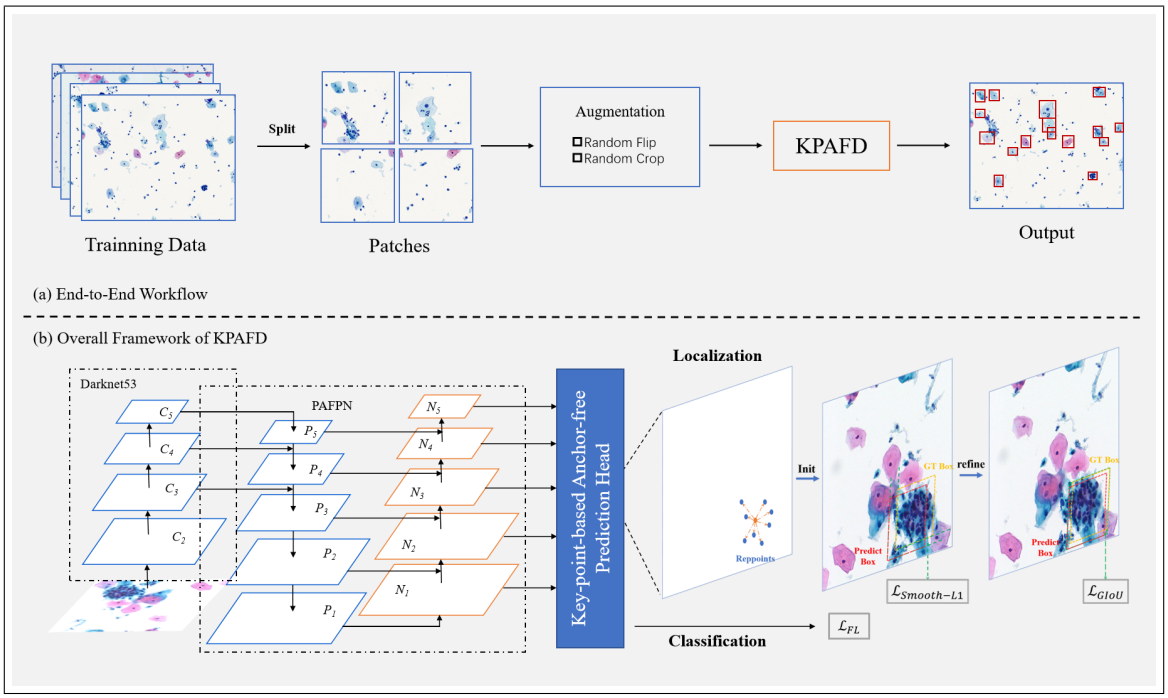


Fig. 1. Pipeline of our method for cervical cell detection

feature representation ability. Considering the cervical cells have different morphological structures, rectangular anchor boxes fail to conform to the cell shapes. Therefore, RepPoints [15] (i.e. point set representation) are used to represent the cells instead of anchors in the prediction head. It helps focus on the specific morphology of different cells and local semantics. Particularly, GIoU[17] loss is also applied to the localization optimization of small cells besides smooth L_1 loss and focal loss. Experiments on cervical cytology ROI datasets verify the effectiveness of our method for cell detection.

The contribution and novelty of this paper is two-folds:

- We present a key-points based anchor-free cervical cell detector. To the best of our knowledge, this is the first to use the anchor-free method for cervical cell detection. It applies PAFPN to strengthen the feature hierarchy and then uses the RepPoints to achieve the cell localization instead of anchor boxes. Finally, GIoU loss is incorporated into the learning procedure to further detect the small cells. Therefore, it yields finer cell localization and classification effect.
- We conduct experiments to evaluate the proposed method on cervical cytology ROIs. Particularly, the ROIs with different liquid-based preparation styles (i.e. drop-slide, membrane-based and sedimentation) are collected to verify the robustness of our method. The results indicate our method is more accurate and robust than the anchor-based methods which are widely used in cervical cell detection and the representative anchor-free methods (e.g Reppoints and YOLOX).

II. METHODOLOGY

A. Overview

The pipeline of our method is present in Fig. 1. It includes a backbone for feature extraction, a neck for feature enhancement, and a head for coordinate regression (localization) and label classification. In the entire end-to-end workflow exhibited in Fig. 1(a), the training ROIs are divided into patches with fixed size. After data augmentation through random flip and crop, the training patches are fed into our proposed KPAFD and finally the coordinate and label of the cells or clumps are automatically generated for further cervical cytology analysis. Noted that our proposed KPAFD applies the backbone Darknet53 to maintain the comparability with the baseline YOLOv3.

B. Hierarchical feature enhancement

Different from YOLOv3 which uses FPN to combine the multi-scale features, our method applies the PAFPN [16] to enhance hierarchical features which adds the bottom-up path augmentation, denoted in Fig. 1(b). Blue and orange parallelograms $\{C_i,P_i,N_i\}$ indicate the feature maps from specific layers. Specifically, C_i are the feature maps from specific layers of backbone Darknet53. P_i are the feature levels yielded by FPN. They are generated by the top-down pathway and lateral connection with the original feature maps $\{C_3,C_4,C_5\}$. Noted that P_2 and P_1 are down-sampled from the previous layer with factor 2 through convolution layers.

Additionally, we introduce the augmented bottom-up structure of PAFPN. N_i are the newly generated feature maps corresponding to P_i. Each new feature N_i firstly is performed on a 3×3 convolutional layer with stride 2 to reduce the spatial size. Then P_{i+1} and the down-sampled map are

TABLE I

DETAILED DATA DISTRIBUTION OF EACH CATEGORY IN TRAINING, VALIDATION AND TEST SETS

Category	Superficial	Intermediate	Glandular	ASC	Koilocytotic	High-N/C-Ratio	LGIN
Train	4429	186	128	473	518	252	94
Valid	693	35	14	88	62	53	6
Test	1393	92	42	159	114	166	26
Total	**6515**	**313**	**184**	**720**	**694**	**471**	**126**

added with lateral connection just like FPN. The generated feature map then goes through another 3×3 convolutional layer to produce the following N_{i+1}. N_i are the final feature hierarchy enhanced by PAFPN and then feed into our key-point based anchor-free prediction head in parallel. Note that P_5 is C_5 and N_1 is P_1, without any processing.

C. Key-points based anchor-free prediction head

To avoid the anchor setting in YOLOv3, we use RepPoints [15] to represent object in an anchor-free fashion which does not involve anchor hyperparameters. It aims to conform to the shapes of the cells with different morphology and the local semantically significant area in the detection procedure. Concretely, RepPoints can be characterized by a set of adaptive sample points:

$$\mathcal{R} = \{(x_i, y_i)\}_{i=1}^{n} \tag{1}$$

where x_i and y_i are the sample points and n is the number of points.

The cell localization procedure of key-points based anchor-free prediction head can be described as two-stage RepPoints regression: initialization and refinement. In the initialization phase, RepPoints is taken as the initial cell representation gained from regressing offsets over the center points through deformable convolution, as shown in Eq. (1).

Then the refinement can be described as:

$$\mathcal{R}_r = \{(x_i + \Delta x_i, y_i + \Delta y_i)\}_{i=1}^{n} \tag{2}$$

where $\Delta = \{(x_i, y_i)\}_{i=1}^{n}$ are the regression offsets of the refined sample points. The min-max function [15] is applied to convert RepPoints into a bounding box. Therefore, the annotated training bounding box can be used and the evaluation of detection performance is facilitated.

D. Learning loss

Our proposed KPAFD can be learnt by the cell localization loss and the classification loss. Following RepPoints [15], smooth L_1 loss $L_{smooth-L1}$ is used to measure the difference between the converted pseudo box and ground-truth bounding box in the initial localization phase. It directly takes the smooth L_1 distance between the top-left and bottom-right corners of prediction and ground-truth. Furthermore, we introduce GIoU loss [17] L_{GIoU} for localization refinement which calculates the intersection between this two rectangular boxes as a metric of loss. Thus avoid the model bias towards large cell clumps and neglect small cells. It can be regarded as the complementation for smooth L_1 loss and generates finer localization. Focal loss L_{FL} is served as the classification loss. Therefore, the entire loss used to train the network is given as:

$$\mathcal{L} = L_{FL} + \alpha L_{smooth-L1} + \beta L_{GIoU} \tag{3}$$

where α and β control the weights of the corresponding loss.

TABLE II

COMPARISON FOR ROIS IN MAP AND RECALL (%)

Methods	mAP	Recall
Faster R-CNN	51.6	84.6
YOLOv3	55.1	90.0
RepPoints	55.8	94.3
YOLOX-S	52.5	83.7
YOLOX-L	55.7	81.8
Ours	**59.9**	**97.0**

III. EXPERIMENTS

A. Dataset

To verify the performance of our method, the liquid-based cervical cytology ROI dataset (denoted as Motic-ROI) provided by Motic is proposed which has 424 ROIs and 7 categories: Superficial cells, Intermediate cells, Glandular cells, Atypical Squamous cells (ASC), Koilocytotic cells, cells with high nuclear-cytoplasmic ratio (High-N/C-Ratio) and Low Grade Intraepithelial Neoplasia (LGIN). The ROIs are scanned at 20X and the cells within ROIs are manually annotated by pathologists. The size of ROIs ranges from 1337×1921 to 5890×7446. The ROIs are randomly spilt into training, validation and test sets following the ratio of 7:1:2.

Detailed data distribution of each category in training validation and test sets are listed in Table I.

TABLE III

ABLATION EXPERIMENT ON THREE COMPONENTS

RepPoints	PAFPN	GIoU Loss	mAP	Recall
✘	✘	✘	55.1	90.0
✔	✘	✘	55.8	94.3
✔	✔	✘	57.5	87.9
✔	✘	✔	59.0	95.1
✔	✔	✔	**59.9**	**97.0**

B. Experimental settings

Our method is compared with Faster R-CNN and YOLOv3 which are widely used in cervical cell detection[6-8, 10-12]. It is also compared with anchor-free methods, such as RepPoints [15] and YOLOX [18]. The sliding-window strategy is used to divide each ROI into the patches with the size of 1024×1024 and the stride of 512 pixels. The augmentation methods are MinIoURandomCrop with min_ious=(0.4, 0.5, 0.6, 0.7, 0.8, 0.9) and min_crop_size=0.3 and RandomFlip with flip_ratio=0.5, then each augmented image will be resized to 640×640 before inputted the network. The network is trained for 300 epochs by stochastic gradient descent (SGD). The learning rate is 0.001. Momentum is 0.9 and weight decay is 0.0005. The IoU thresholds for calculating Average Precision(AP) and mean Average Precision(mAP) are respectively 0.45 and 0.5. The number of sample points n in Eq. (1) is set to 9. All the experiments are conducted on a computer with an AMD Ryzen Threadripper 3960X 24-Core Processor and a GPU of NVIDIA GTX 3090.

TABLE IV

ABLATION EXPERIMENT ON LOSS WEIGHTS α AND β

Init Phase, α	**Refine Phase, β**	**mAP**
SmoothL1, 0.5	GIoU, 1.0	56.9
SmoothL1, 0.75	GIoU, 1.0	57.8
SmoothL1, 1.0	**GIoU, 1.0**	**59.9**
SmoothL1, 1.0	GIoU, 0.75	57.0

C. Experimental results and analysis

The cervical cell detection evaluation of different methods for ROIs are given in Table II. Note that the anchor-free methods are basically better than anchor-based methods, especially RepPoints which has relatively higher mAP and recall. It shows that the key-points based anchor-free representation contributes to exploring the different shapes of cells and local semantical area and thus improving the detection performance. More importantly, our method outperforms other methods. It indicates that PAFPN enhances the feature representation ability of our method with bottom-up path augmentation and GIoU loss helps balance large cell clumps and small cells in the learning procedure.

Qualitative detection results are illustrated in Fig. 2. We visualize the predictions of Faster R-CNN, YOLOv3, RepPoints and our method. Obviously, our method generates more precise localization and classification performance, and the missed case is relatively rare. Particularly the detection effect of the small cells is more promising.

D. Ablation study

Effectiveness of the components. Ablation experiments are designed to verify the effectiveness of the three components (RepPoints, PAFPN and GIoU Loss) used in our method. Note that our method without these three components is the conventional YOLOv3.

According to Table III, The involved anchor-free representation RepPoints contributes to the improvement of YOLOv3.

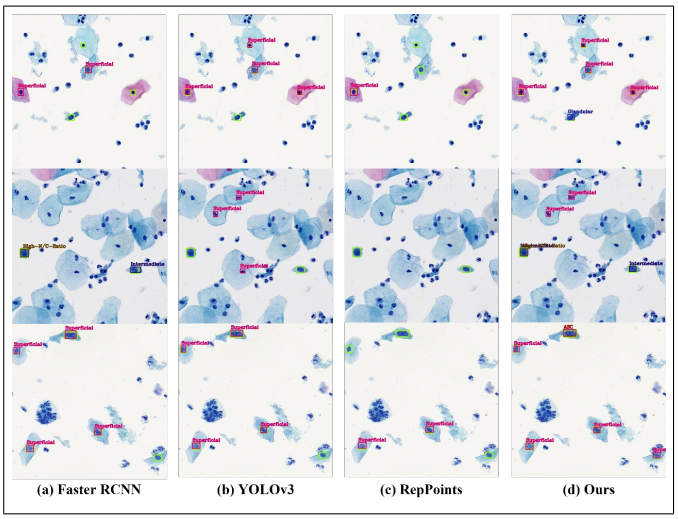


Fig. 2. Qualitative comparisons of cervical cell detection on ROIs. Green rectangles denote annotations by pathological experts. Non-green rectangle and its label denote the detection result of each method

Besides, the use of PAFPN boosts the hierarchical feature representation ability. More importantly, GIoU loss greatly improves the mAP and recall. It benefits from the balance between large cell clumps and small cells. Overall, our method obtains more desirable results incorporating these three components.

Loss weights. Table IV shows the influence of different loss weights (i.e. α and β) on cell detection performance. The KPAFD reaches the relatively optimal result when α and β are taken as 1.

E. Performance on different preparation styles

Considering the liquid-based cervical cytology slide has different preparation styles (i.e. drop-slide, membrane-based and sedimentation), we collect these three types of ROI datasets in clinical to evaluate the robustness of our method, and the number of these three types of ROI is 797, 454 and 1492 respectively. Note that different from the Motic ROI dataset, the membrane-based ROIs are from another laboratory. The staining differences of different preparation style are shown in Fig. 3.

We conduct the experiments on each dataset and simultaneously test on the mixed dataset (Motic, Drop-slide, Membrane-based and Sedimentation). The ratio of training, validation and test sets is 7:1:2. We compare the proposed method with YOLO-L since it is a recently proposed anchor-free detector in YOLO series.

The results are given in Table V and the Mixed style denotes the combination of the four styles mentioned above. It can be found that the detection performance is easily influenced by the preparation styles. Compared with YOLO-L, our method has the robustness for various preparation styles.

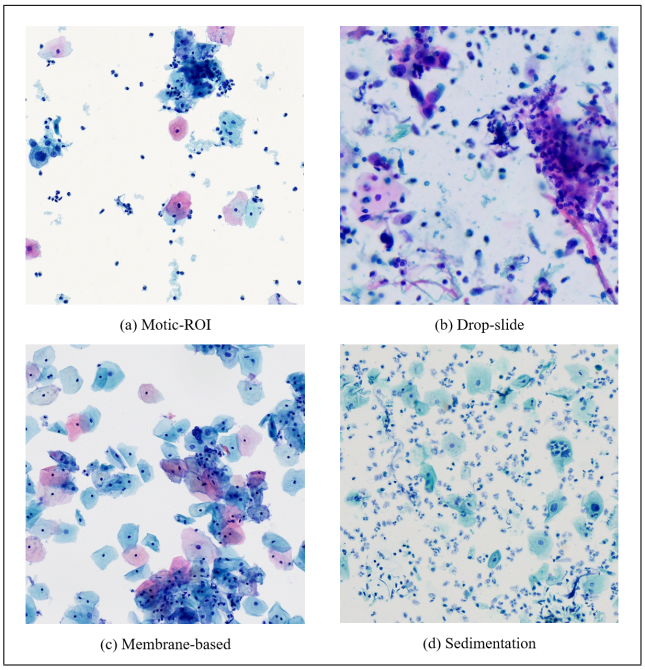
IV. CONCLUSIONS

In this paper, we present a key-points based anchor-free cervical cell detector based on YOLOv3. It applies

TABLE V

COMPARISON UNDER DIFFERENT PREPARATION STYLES

Preparation Styles	Methods	mAP	Recall
Motic	YOLOX-L	55.7	81.8
	Ours	**59.9**	**97.0**
Drop-slide	YOLOX-L	49.0	73.8
	Ours	**49.1**	**90.1**
Membrane-based	YOLOX-L	**42.4**	72.2
	Ours	41.8	**86.7**
Sedimentation	YOLOX-L	44.9	73.5
	Ours	**46.7**	**87.7**
Mixed	YOLOX-L	38.3	71.3
	Ours	**40.0**	**86.8**



(a) Motic-ROI (b) Drop-slide (c) Membrane-based (d) Sedimentation

Fig. 3. The staining differences of different preparation style

PAFPN for the hierarchical feature enhancement through the bottom-up path augmentation and uses the key-points based representation to achieve anchor-free cell detection instead of anchor box. Besides, GIoU loss is introduced into the learning procedure jointly with smooth L_1 loss and focal loss and thus the small cells can be better detected. Experiments on the cervical cytology ROI datasets demonstrate the effectiveness and robustness of our method.

ACKNOWLEDGMENT

This work was partly supported by the National Natural Science Foundation of China (grant no. 61906058, 61901018, 62171007 and 61771031), the Fundamental Research Funds for the Central Universities of China (grant no. JZ2022HGTB0285).

REFERENCES

[1] R. Nayar, D. C. Wilbur., "The Bethesda System for Reporting Cervical Cytology: Definitions, Criteria, and Explanatory Notes", Springer, 2015.

[2] S.Ren, K. He, R. B. Girshick, J Sun., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *in Proceedings of Conference on Neural Information Processing System(NeurIPS)*, Montreal, Quebec, Canada, 2015, pp. 91-99.

[3] T. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, S. J. Belongie., Feature Pyramid Networks for Object Detection", *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*,Honolulu, HI, USA, 2017, pp. 936-944.

[4] J. Redmon, S. K. Divvala, R. B. Girshick, A. Farhadi., "You Only Look Once: Unified, Real-Time Object Detection". *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788.

[5] K. He, G. Gkioxari, P. Dollár, R. B. Girshick., "Mask R-CNN", *in Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2980-2988.

[6] J. Du, X. Li, Q. Li., "Detection and Classification of Cervical Exfoliated Cells Based on Faster R-CNN", *in Proceedings of IEEE 11th international conference on advanced infocomm technology (ICAIT)*, Jinan, China, 2019. pp. 52–57.

[7] C. Zhang, D. Liu, L. Wang, Y. Li, X. Chen, R. Luo, S. Che, H. Liang, Y. Li, S. Liu, et al., "DCCL: A Benchmark for Cervical Cytology Analysis", *in International Workshop on Machine Learning in Medical Imaging*, Springer, 2019, pp. 63–72.

[8] Y. Xiang, W. Sun, C. Pan, M. Yan, Z. Yin, Y. Liang., "A Novel Automation-Assisted Cervical Cancer Reading Method Based on Convolutional Neural Network", *arXiv:1912.06649*, 2019.

[9] J. Redmon, A. Farhadi., "YOLOv3: An Incremental Improvement", *arXiv:1804.02767*, 2018.

[10] Y. Liang, C. Pan, W. Sun, Q. Liu, Y. Du., "Global Context-aware Cervical Cell Detection with Soft Scale Anchor Matching", *Computer Methods and Programs in Biomedicine*, vol 204, pp. 106061, 2021.

[11] Z. Chai, L. Luo, H. Lin, H. Chen, A. Han, P. Heng., "Deep Semi-Supervised Metric Learning with Dual Alignment for Cervical Cancer Cell Detection", *in Proceedings of the International Symposium on Biomedical Imaging (ISBI)*, Kolkata, India, 2022, pp. 1-5.

[12] N. Sompawong, J. Mopan, P. Pooprasert, W. Himakhun, K. Suwannarurk, J. Ngamvirojcharoen, T. Vachiramon, C. Tantibundhit., "Automated Pap Smear Cervical Cancer Screening Using Deep Learning", *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 2019, pp. 7044-7048.

[13] Z. Tian, C. Shen, H. Chen, T. He., "FCOS: Fully Convolutional One-Stage Object Detection", *in Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 9626-9635.

[14] H. Law, J. Deng., "CornerNet: Detecting Objects as Paired Keypoints", *in Proceedings of 15th European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 765-781.

[15] Z. Yang, S. Liu, H. Hu, L. Wang, S. Lin., "RepPoints: Point Set Representation for Object Detection".*in Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 9656-9665.

[16] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia., "Path Aggregation Network for Instance Segmentation", *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 8759-8768.

[17] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. D. Reid, S. Savarese., "Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression",*in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 658-666.

[18] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun., "YOLOX: Exceeding YOLO Series in 2021", *arXiv:2107.08430*, 2021.