

# Automated Segmentation of Cervical Cell Images Using IMBMDCR-Net

Yanjing Ding, Weiwei Yue, and Qinghua Li\*

**Abstract**—Early screening of cervical lesions is of great significance in pathological diagnosis. Owing to the complexity of cell morphological changes and the limitations of medical images, accurate segmentation of cervical cells is still a challenging task. In this paper, an isomorphic multi-branch modulation deformable convolution residual model is proposed to extract features for enhancing the segmentation of small cells and overlapping cytoplasmic boundaries. Then the regional feature extraction, boundary box recognition, and adding a single pixel-level mask at the last level are integrated and optimized based on the cascade regional convolution neural network (Cascade R-CNN) to complete the segmentation of cervical cells for getting better accuracy. The proposed framework was evaluated on the ISBI2014 cervical cell segmentation competition public dataset. Experimental results show that the average accuracy of the network model in cervical cell segmentation is 81.1%, and the accuracy of small targets is 77%. To some extent, it can assist pathologists in screening cervical cancer in the early phase.

**Index Terms**—Cervical cell, instance segmentation, cascade RCNN, modulation deformable convolution, residual network, deep learning

## I. INTRODUCTION

Cervical cancer is the world's fourth most prevalent malignancy among women [1]. However, it is worth noting that cervical cancer is the only tumor that can be found and cured early in gynecologic cancer [2]. In traditional cervical cytology screening, the Pap smear is observed by pathologists using a manual screening method, which is extremely tedious and time-consuming and is easily misdiagnosed and missed due to the subjective limitations of pathologists. With the continuous development of digital pathology, the automation of the Pap smear diagnosis process brings great benefits to the prevention and treatment of cervical cancer. The automation system replaces the manual screening process of Pap smear with a computer system that can simulate the behavior and knowledge of experts, this intelligent system replaces the analysis and decision-making of experts with feature extraction and classification [3]. The main feature used by pathologists to analyze cell classes is the nuclear-cytoplasmic ratio(N/C). However, the complex situations of overlapping cells, blurred edges, uneven staining, and poor contrast in cervical images lead to false diagnoses. So an effective segmentation algorithm becomes an essential first step to detect the contours of the nucleus and cytoplasm.

Manuscript received September 19, 2022; revised December 26, 2022; accepted January 10, 2023.

Yanjing Ding and Weiwei Yue are with School of Physics and Electronics, Shandong Normal University, Jinan, China.

Qinghua Li is with College of Artificial Intelligence and Big Data for Medical Sciences SDFMU, Jinan, China.

\*Correspondence: liqinghua1977@163.com (Q.L.)

The lack of current pathologists and the insensitivity of manual interpretation lead to widespread misdiagnosis and missed diagnoses. Combined with the advantages of end-to-end instance segmentation, this paper proposed an isomorphic multi-branch modulation deformable convolution residual network for automatic segmentation of cervical cells. To improve the utilization of spatial information, we developed an isomorphic multi-branch modulation deformable convolution residual model to refine feature representations to realize the segmentation of cervical cells. In addition, instance segmentation is realized based on candidate regions, for obtaining a more accurate prediction box, we adopt integration and optimization region feature extraction, boundary regression, classification, and adding a single segment mask prediction at the last stage based on cascade region convolution neural network (Cascade R-CNN) [4], completed the functional test, and finally obtained the mask effect maps of cervical cells divided into single nucleus and cytoplasm.

The remainder of this paper is organized as follows: Section II introduces the previous work. Section III analyzes the methods. The experimental results are given in section IV. Finally, Section V provides conclusions of the work.

## II. THE RELATED WORK

Segmentation is a complex and critical step in medical image processing, which provides a reliable basis for clinical diagnosis and pathological research. There have been many research methods and achievements in medical image segmentation technology. For example, natural image segmentation based on effective histogram threshold T technology [5], fuzzy color image segmentation based on watershed transformation [6], image segmentation using improved FCM watershed algorithm and DBMF [7], density-based clustering for interactive liver segmentation [8], and medical image edge detection technology based on ant colony algorithm[9]. Many methods have also been published for cervical cell segmentation thus far. Some traditional methods include thresholding [10, 11], edge detection [12], watershed [13, 14], and superpixel [15] methods. Some others combine the above algorithms. For instance, Tareef *et al.* [16] proposed using a gradient threshold to extract cell boundaries and then combining morphological technology to infer cytoplasmic segmentation. Ushizima *et al.* [17] introduced a technical method combining superpixel and Voronoi diagrams to detect cells, the algorithm combines low average pixels with adaptive histogram equalization to improve the problem of poor contrast and uses the Voronoi diagram to realize cell segmentation. Wang *et al.* [18] reported a method of combining the Gaussian mixture model (GMM) with the

regularized level set model to extract overlapping cervical cells. Most of these algorithms extract cell images based on the cell contours, colors, and textures. Although the methods are very effective, it is still a challenging task to realize fast, accurate, and automatic segmentation.

In recent years, deep learning techniques have made great achievements in the field of automatic medical image segmentation [19]. A specific segmentation algorithm developed based on medical images, U-net [20], becomes a more classical semantic segmentation network. Many improved U-net models are proposed for the automatic segmentation of the liver [21], spine [22], Cine-MR images [23], infant ventricles [24], retinal images [25], and blood vessels in fundus images [26]. Huang *et al.* [27] Proposed combining the convolutional neural network U-Net with the improved level set method to realize the segmentation of overlapping cervical cells. Although the classical U-Net semantic segmentation has attained good result, it does not focus on regions of interest (ROI), such as over-processing irrelevant regions, which leads to a waste of computing resources. Later, the instance segmentation model based on R-CNN [28] was applied to cervical cell image segmentation. For instance, Kurnianingsih *et al.* [29] proposed a method to segment the whole cervical cells using a mask region convolution neural network. Long *et al.* [30] introduced the multi-scale feature fusion method based on the Mask R-CNN for cervical image segmentation. Both methods have made significant progress in the field of cytology segmentation, but the instance segmentation based on R-CNN still has some limitations. such as R-CNN is a target detection method, and its CNN network is used to complete the classification task, which is not suitable for the accurate localization of cervical cells. Although a down-sample in feature extraction enriches high-level semantic features, it loses a lot of information that cannot be ignored in the algorithm. since detection is coarser than segmentation, it is no problem for simple classification. However, this spatial information is very crucial for accurate segmentation. To segment the overlapping cell region and split it into a single cell, an isomorphic multi-branch modulation deformable convolution residual network was proposed to locate more accurate boundary information.

#### A. Dataset Preparation and Preprocessing

In the images of cervical cells viewed under a microscope, where cells are isolation, adhered, or overlapped, we cannot make any assumptions about the number of cells, nor can we expect the cells in the images to be segregated from each other. Therefore, in this study, two sets of ISBI 2014 test images [31] were selected for performance evaluation, which consists of 900 images. The details of the data sets used in this experiment are shown in Table I. All these composite images are  $512 \times 512$  grayscale images, generated by overlapping a set of isolated cells extracted from real Pap smears, reconstructed by applying random linear luminance transforms and random rigid transforms, and finally positioned on the synthetic images. The number of overlapping cells in these images varied between 2 and 10, with varying degrees of overlap, contrast, and texture. Image samples are shown in Fig. 1.

In order to facilitate training, this paper creates a dataset with the same format as COCO. The model needs to segment

the nucleus and cytoplasm of cervical cells. Therefore, under the guidance of the pathologists, 900 cervical images are labeled and stored in JSON format. Data augmentation is applied as a preprocessing step to improve the performance and prevent overfitting in our limited images, including translation, flipping operations, and image contrast enhancement methods are used to reasonably and effectively create slightly variable “copies” to make up for the imbalance of sample data. These transformations occur in real-time scenarios during image capturing.

TABLE I: INTRODUCTION TO USING DATASETS IN OUR EXPERIMENTS

Dataset	Image type	Image size	cells	Image number
ISBI-14 test-90	Synthetic	$512 \times 512$	2-10	90
ISBI-14 test-810				810

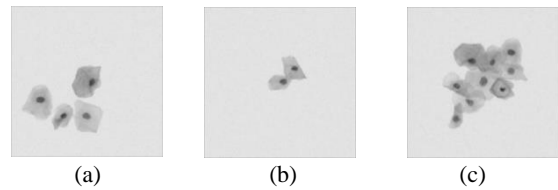


Fig. 1. Data sample: (a) isolate cell (b) adhering cell (c) overlapping cell.

#### B. Cervical Cell Segmentation Algorithm Flow

The accurate segmentation of cervical images provides important support for the subsequent quantitative analysis of pathological cells and classification. Extracting image features by training neural networks has obvious advantages, which can help experts reduce the burden and the errors caused by subjective reasons. Fig. 2 shows the overall flow chart of cervical cell segmentation. The preprocessed images are trained through the segmentation model and parameter optimization. The trained model realizes the visualization of cervical image instance segmentation and the prediction of nuclear and cytoplasm masks.

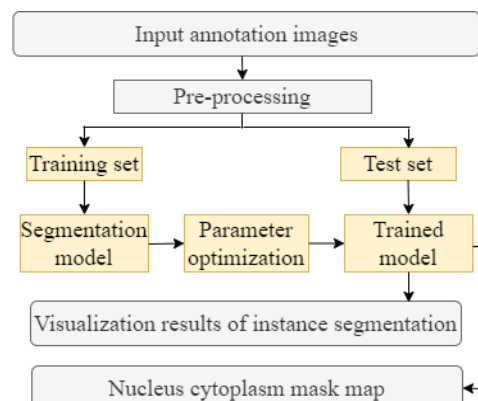


Fig. 2. Overall flowchart of cervical cell segmentation.

### III. METHODS

For the methods used in this study, this section divides the proposed cervical cell instance segmentation model into two parts: one is to extract the cervical cell image feature map based on an isomorphic multi-branch modulation deformable convolution residual network; the other is to integrate and optimize by the regional feature extraction, boundary box recognition and adding a mask segmentation branch [32] based on Cascade R-CNN for obtaining the mask prediction

results. The general outline of our proposed methodology is shown in Fig. 3.

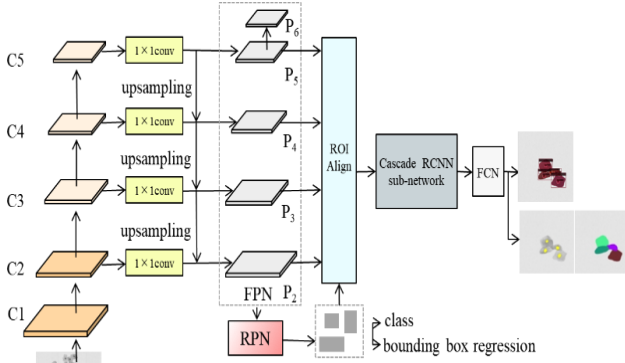


Fig. 3. Cervical cell segmentation model based on cascade R-CNN.

### A. Isomorphic Multi-Branch Modulation Deformable Convolution Residual Network

Regional convolutional neural network (R-CNN) and its extensions (Fast R-CNN, Faster R-CNN, etc.) have been proven to be successful in image segmentation and target detection. Mask R-CNN outputs an object mask in parallel based on Faster R-CNN [33] (generate target classification label + window) and has successfully become one of the most extensive instance segmentation networks. Liu *et al.* [34] proposed to use the ResNet [35] as the backbone of a Mask R-CNN and modify it according to the image of the cervical nucleus. However, the network pays too much attention to the optimization of deep features, so the low-level features containing more detailed information are not fully utilized, which can not be ignored for identifying small objects. Too deep a network will reduce the detection generalization ability of cervical cells, resulting in missed diagnosis, and requiring a more refined spatial layout of extraction objects.

Inspired by the idea of Inception [36] and deformable convolution [37], this paper improves the residual network by combining grouped convolution and modulated deformable convolution, so that the process of cervical image feature extraction can focus more on the effective information area to obtain better detection performance. The details are shown in Fig. 4. In order to prevent network distortion and ensure the accuracy of the segmentation effect, the original images are zero-padding, and then convoluted (the number of channels is 64, the size of convolution kernel is  $7 \times 7$ , the step size is 2), the length and width are compressed to the 1/2 of the original image. The feature layer generated by the convolution filter is input to Batch Normalization and Relu to prevent vanishing/exploding gradients. The feature layer of Relu is then input into the max-pooling layer. The pooling result C1 passes through different residual modules to output feature maps C2, C3, C4 (Res\_block and 22 Conv\_blocks), and C5 of various scales with various levels. Conv\_Block and Res\_block adopt the bottleneck structure. The only difference between them is Res\_Block adds a convolution to the skip connection, which can not only deepen the depth of the network but also change the dimension of the network. The first convolution layer size of  $1 \times 1$  is responsible for dimension reduction, the size of the second convolution layer is  $3 \times 3$ , and the third convolution layer size  $1 \times 1$  is responsible for dimension increase, so the structure gives the

second convolution layer to learn fewer weights. The improved model is called an isomorphic multi-branch modulation deformable convolution residual network in Fig. 5.

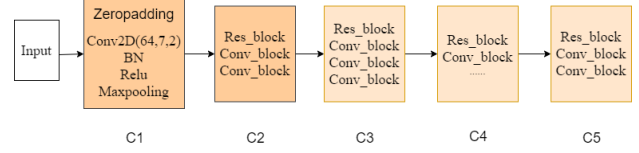


Fig. 4. The backbone architecture used in this study.

Cervical images are simpler than natural scene images, but the continuous deepening of the network leads to excessive hyperparameters and easier overfitting. Therefore, this paper puts all  $3 \times 3$  Conv layers replaced by group convolution with 64 groups in the residual network, and the number of channels of each group is 4. Group convolution can not only flexibly change the number of the group to obtain more effective accuracy, At the same time, the parameters are also reduced as a whole (the input and output are originally 128 channel convolutions. Because they are divided into 64 different groups of convolution kernel, Although the number of output channels remains the same, the input changes from 128 to 2, reducing the parameters). Besides, the convolution kernel of traditional feature extraction is generally fixed size, so it is difficult to make the corresponding adjustment according to the actual situation of cervical images, and the generalization ability is weak. In order to better extract the input features, all  $3 \times 3$  Conv layers add an additional modulation offset (light white arrow in Fig. 5) to each pixel of the standard convolution to ensure the accurate extraction of effective information in the C3-C5 stage. Among them, offset is to find the location of the region of interest, and end-to-end learning can be carried out through gradient backpropagation. the modulation mechanism is simply weighted. The target region is modified by assigning different weights to offset, these two aspects realize the accurate feature extraction of effective information. The modulation deformable convolution output is shown as:

$$y(p) = \sum_{s=1}^S w_s \cdot x(p + p_s + \Delta p_s) \cdot \Delta m_s \quad (1)$$

where  $\Delta p_s$  and  $\Delta m_s$  are the learnable offset and modulation scalar of  $s$  position respectively, both of which are obtained by applying a separate convolution layer on the same input characteristic map,  $w_s$  is the initialization weight,  $p$  is the original position of the feature map, and  $p_s$  is the position listed by the sampling points.  $\Delta p_s$  is a real number with unrestricted range,  $\Delta m_s$  is in  $[0,1]$ , and the offset value  $x(p + p_s + \Delta p_s)$  maybe a floating-point number, so bilinear interpolation is used to calculate the eigenvalue of the offset position.

### B. Cervical Cell Segmentation Model Based on Cascade R-CNN

It has been proved that using pyramid representation to process multi-scale image tasks can obtain better performance [38]. Therefore, the obtained feature maps C2-C5 are used to construct the feature pyramid for more detailed detection, so as to realize the multi-scale fusion of low features with high resolution and deep features with rich

semantic information. As shown in Fig. 3. The output of the last residual block of the feature layers C2-C5 is extracted go through a  $1 \times 1$  convolution layer to maintain channel unity, then up-sampling is carried out, and the up-sampling features fuse with the down-top features of the same dimension to form the final feature mapping set P2-P5, P6 is simply a max-pooling of P5 and only for covering a larger cell region

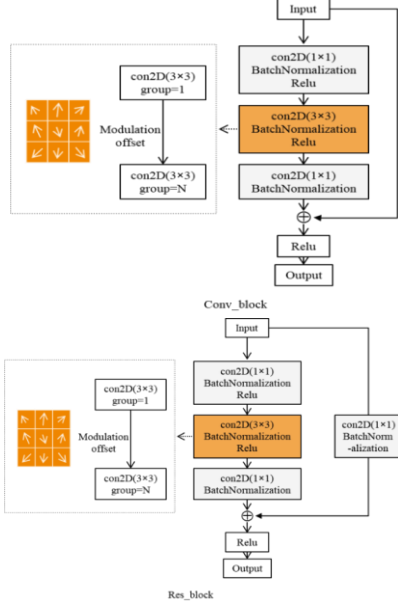


Fig. 5. Isomorphic multi-branch modulation deformable convolution residual network.

Thus, the proposed feature extraction network generates region proposals through RPN at five different scales of P2-P6. By using the anchors to predict the region proposals, in order to guarantee that the anchor boxes can cover the authentic cervical cells, we define the basic scale of the anchor as 8 and the aspect ratios of the anchor as  $[1/2, 1, 2]$ . Then, after filtering and cutting, the category of cervical cells is judged through the softmax function. At the same time, the anchor boxes are modified by the other branch bounding box regression to form a rough suggestion box, Finally, the obtaining a series of region proposals by the above operations. In order to avoid the impact of the twice quantization error of RoI Pooling on target segmentation, those region proposals and corresponding effective feature maps P2-P5 are converted into ROIAlign to obtain a fixed size feature map, the size of the feature map with segmentation accuracy in this paper is  $7 \times 7$  for cell recognition,  $14 \times 14$  for cell segmentation.

After ROIAlign, as shown in Fig. 6. since the additional mask output of Mask R-CNN network depends on the output of the target boxes, for obtaining a more accurate prediction box, this paper based on Cascade R-CNN sub-network continuously revises new region proposals successively through detection network with threshold values of  $[0.5, 0.6, 0.7]$ , which makes proposals resampled in the previous stage can adapt to the detection network with a higher threshold, similar to RPN operation. By cascading several classifiers and regressors, the target categories and locations obtained in the last stage are taken as the output results of the whole network. In order to make the network have the best balance between cost and AP performance, this paper selects to add a mask head in the last stage. The classification and regression

network in the last stage is used to judge whether the detected content contains the target, and the proposal is adjusted to obtain the prediction box. The prediction box is used to intercept the effective feature layer again, the intercepted results are transmitted to ROIAlign to obtain the local feature layer of the same size and then transmitted to the full convolution neural network (FCN) to obtain the cervical cell segmentation results.

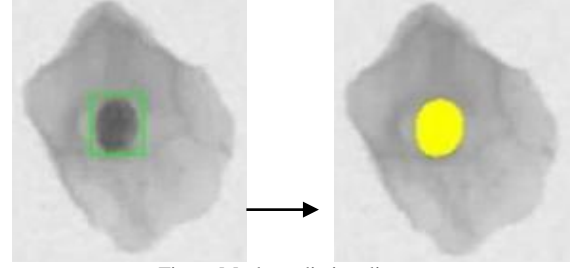


Fig. 6. Mask prediction diagram.

## IV. EXPERIMENTS AND RESULTS

### A. Experimental Environment and Parameter Setting

In this section, we conducted all of the experiments using NVIDIA GeForce RTX 3080 with 16GB memory, and Windows 10 installed on an Intel Core i7 CPU with 64 GB RAM. We implemented the model using the PyTorch framework. The experiment is based on MMDetection [39] and detectron2 [40]. The super parameters, configuration, and label allocation procedures follow the settings in [41, 42]. The training model for 24 epochs using the Stochastic Gradient Descent algorithm with a momentum of 0.9, the batch size is set to 2 with SGD optimizer, the initial learning rate is 0.001, and reduce it for the 8th and 11th epochs, the weight decay by 0.0001 to avoid overfitting. To evaluate the validity of our model proposed, we randomly divided the limited cervical cell dataset into train-validation-test at a ratio of 8:1:1 to ensure that images of the same case can only appear in the training set, validation set, or test set.

### B. Evaluation Metrics

Standard COCO indicators were used in this study for understanding the segmentation effect of the segmentation network on different scale targets. Including AP (averaged over IoU thresholds),  $AP_{50}$ ,  $AP_{75}$  (AP with IoU of 0.5 and 0.75 respectively, as (2)), and  $AP_S$ ,  $AP_M$ ,  $AP_L$  (AP with different scales). To prove the availability of our proposed method, we take AP as an important index to measure the segmentation accuracy, The mean of accuracy is calculated as (3). Besides, the cell belongs to small target objects in the studied cervical image, so  $AP_S$  (the number of pixels in the segmentation mask  $< 32^2$ ) is taken as another index to measure the segmentation accuracy. Unless otherwise stated, the AP uses the evaluation conducted by mask IOU.

$$IoU(i) = n_{ii} / (\sum_j (n_{ij} + n_{ji}) - n_{ii}) \quad (2)$$

$$mIoU = \frac{1}{N} \sum_i IoU(i) \quad (3)$$

In the formula:  $n_{ii}$  represents the number of pixels with category  $i$  predicted as category  $i$ ,  $n_{ij}$  represents category  $i$  the number of pixels predicted as category  $j$ ,  $n_{ji}$  represents the number of pixels predicted as category  $i$  for category  $j$ ,

and  $N$  represents the number of categories.

### C. Comparative Analysis and Results

For the quantitative evaluation of the segmentation model. We compared the proposed IMBMDCR-net with the baseline model ResNet50, ResNet101, Res2Net101, ResNext101(32×4d), and ResNext101(64×4d) in the segmentation process. The quantitative results of the sample validation images are shown in Table II. We found that the proposed CNN architecture achieved superior performance compared to the other five backbone networks, the average accuracy and small target accuracy of segmentation are improved, which shows the effectiveness of the cervical cell segmentation method based on isomorphic multi-branch modulation deformable convolutional residual (IMBMDCR) network. since the proposed residual model can enhance feature information more effectively, which makes it detect various sizes of cervical cells more correctly.

We also compared the segmentation performance of the proposed method with the original Mask R-CNN over six network architectures. The relevant results are presented in Table II. We found that the proposed model is superior to the original Mask RCNN in average precision (AP) and small target precision (APs) over all backbone networks. Particularly, compared with Mask R-CNN, the improved

model increases the AP by an average of 0.38% and the APs by an average of 0.5%. For instance, the AP and APs of our proposed method using IMBMDCR-net as the backbone network are 81.1% and 77% respectively, which is better than the Mask R-CNN method (80.5% and 76.5%). Fig. 7 presents the precision value on the validation dataset for 24 epochs, we use the last epoch for segmentation.

For a more intuitive explanation, a cervical image with 10 highly overlapping cells is selected for visualization, Fig. 8 provides the qualitative results of Cascade Mask RCNN compared with the Mask R-CNN method using IMBMDCR-net, ResNet50, ResNet101, Res2Net101, ResNext101(32×4d), and ResNext101(64×4d) as the backbone network. As from Fig. 8, even though the results do not differ much from each other, when we look into details, our model is more conducive to the segmentation of the nucleus and overlapping cells in the cervical image. Moreover, the qualitative results of more sample test images of the proposed model are shown in Fig. 9. As can be seen from the diagram, the proposed method can not only precisely divide the nucleus, but also clearly extract the boundaries of isolate, adhere and overlap the cytoplasm for good segmentation.

TABLE II: COMPARISON OF THE PROPOSED METHOD SEGMENTATION RESULTS (%) WITH RELATED METHODS ON VALIDATION DATASET

Method	Backbone	AP	AP <sub>50</sub>	AP <sub>75</sub>	AP <sub>S</sub>	AP <sub>M</sub>	AP <sub>L</sub>
Cascade Mask R-CNN	ResNet-50-FPN	80.4	99.4	97	76.5	82.4	88.6
	ResNet-101-FPN	80.3	99.4	96.5	76.4	81.8	88.8
	Res2Net-101-FPN	80.7	99.4	97	76	83.5	89
	ResNext101-32×4d-FPN	80.3	99.4	96.3	76.1	82.4	88.5
	ResNext101-64×4d-FPN	80.6	99.4	97	76.7	82.1	88.6
	<b>IMBMDCR-Net(ours)</b>	<b>81.1</b>	99.4	97	<b>77</b>	83.5	88.9
Mask R-CNN	ResNet-50-FPN	80	99.4	96.9	75.7	82.3	88.9
	ResNet-101-FPN	80.1	99.4	97	76.1	82.3	88
	Res2Net-101-FPN	80.3	99.4	97	75.6	83.9	88.7
	ResNext101-32×4d-FPN	80.2	99.4	96.2	75.9	83.1	87.9
	ResNext101-64×4d-FPN	80	99.4	96.8	75.9	82.3	88.6
	<b>IMBMDCR-Net(ours)</b>	<b>80.5</b>	99.4	97.1	<b>76.5</b>	83.4	87.7

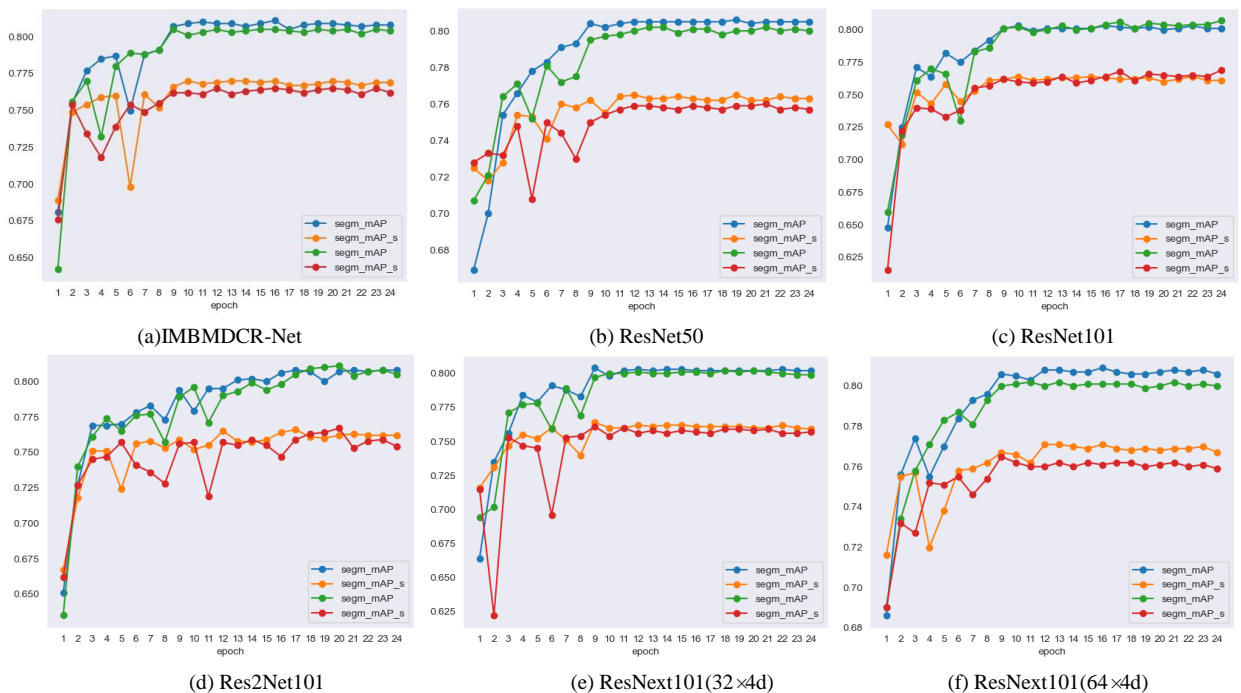
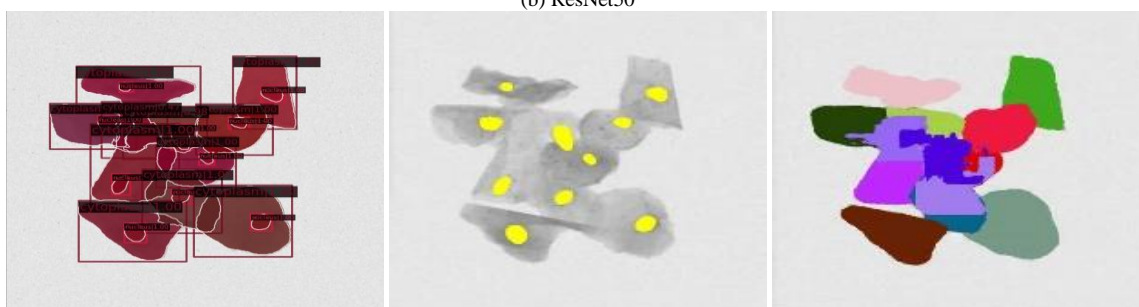
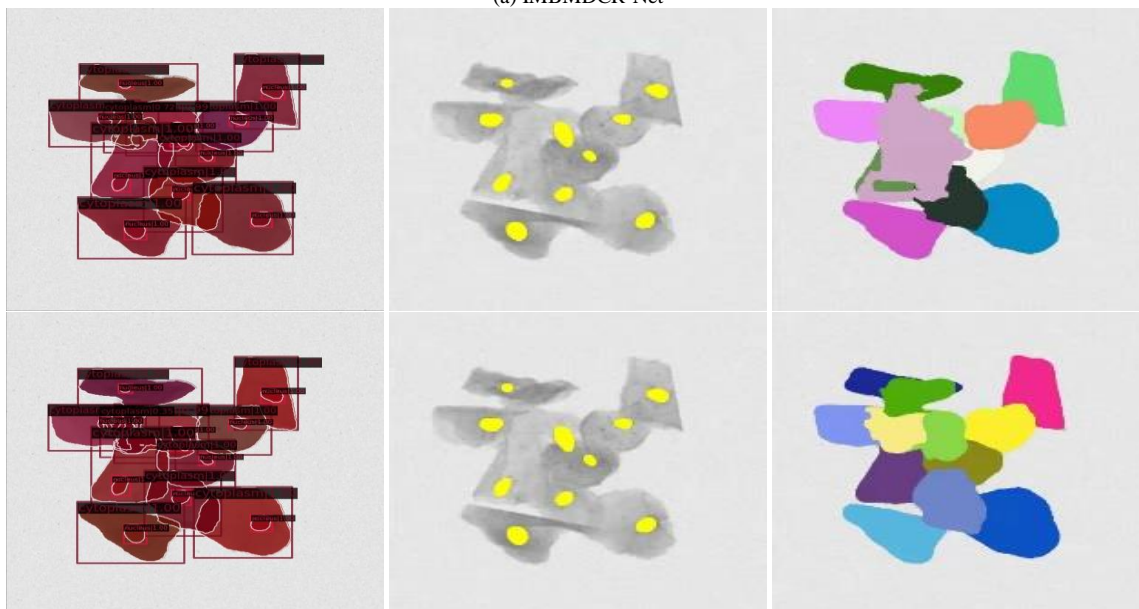
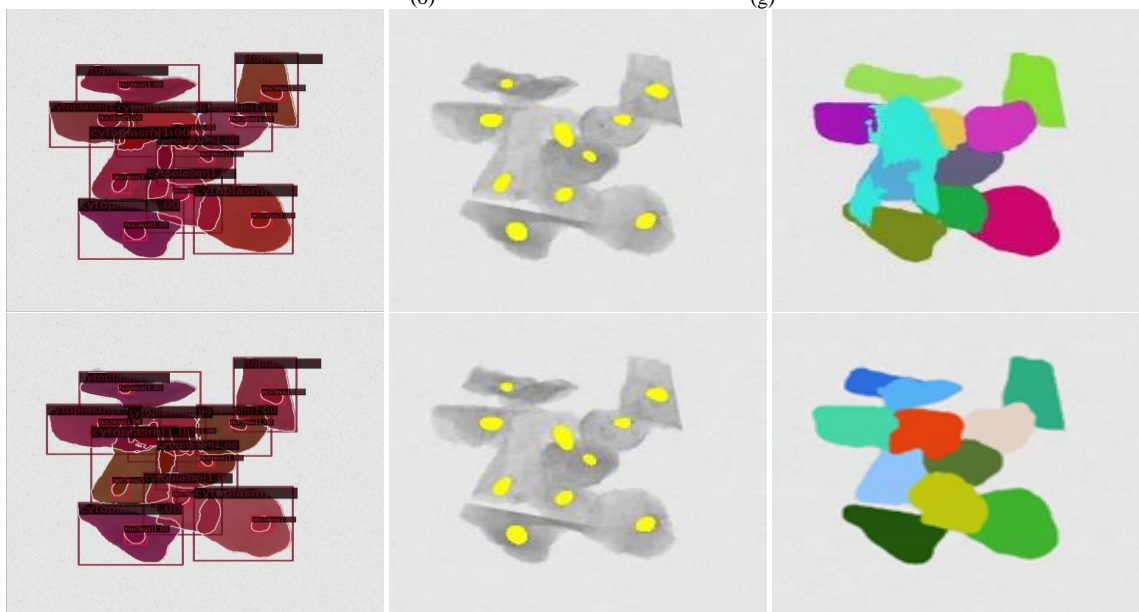
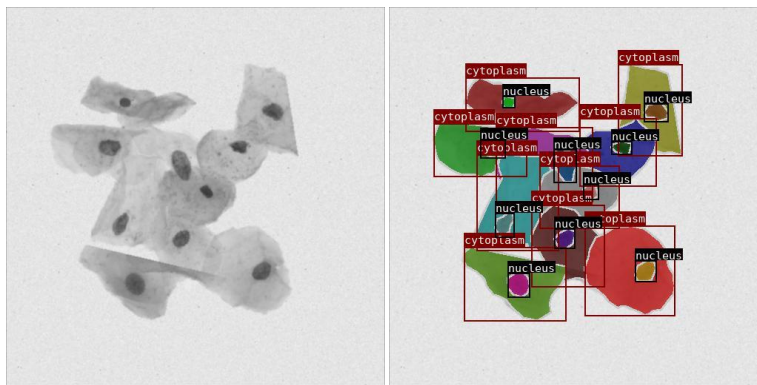
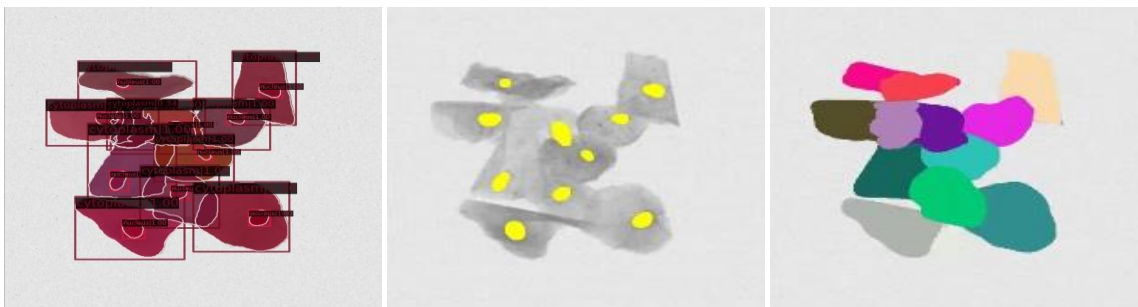
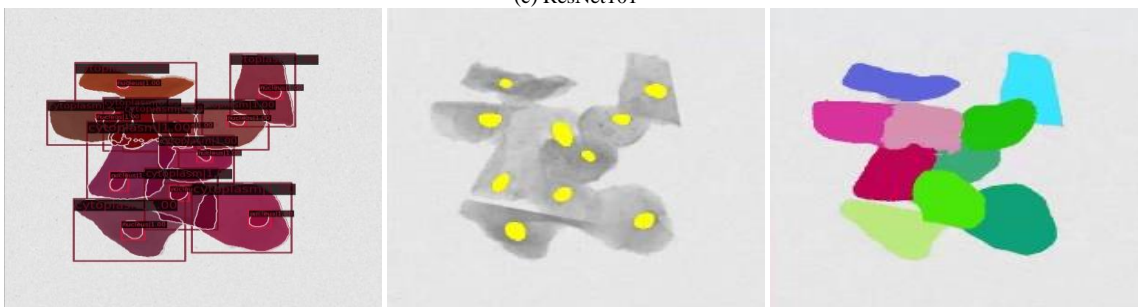


Fig. 7. “Green “ and “Red” represent the average AP value and small target APs value of Mask RCNN; “Blue” and “Orange “ represent the average AP value and small target APs value of Cascade Mask RCNN. IMBMDCR-net, ResNet, Res2Net, and ResNext are used as backbone networks, respectively.





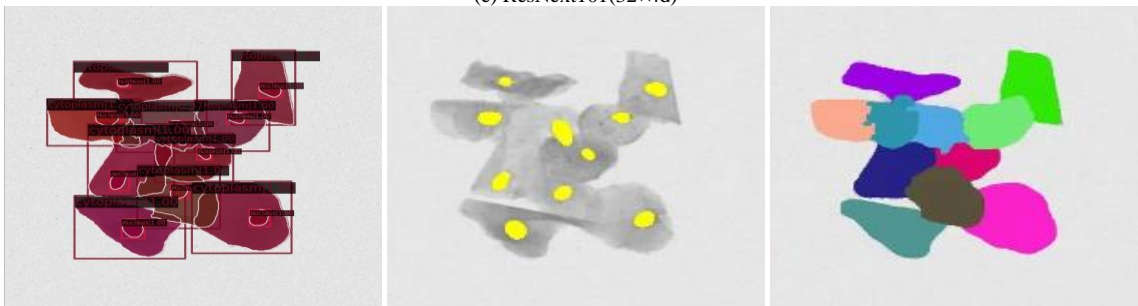
(c) ResNet101

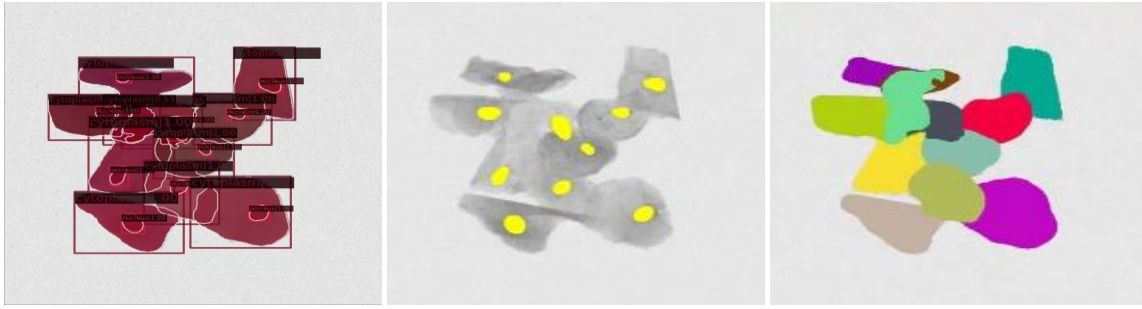


(d) Res2Net101



(e) ResNext101(32x4d)





(f) ResNext101(64×4d)

Fig. 8. Visualization of the predicted segmentation results. (o) The original picture. (g) Visualization of the ground-truth annotations. The segmentation mask results and the instance segmentation effect pictures are predicted by Mask R-CNN and Cascade Mask RCNN. IMBMDCR-Net, ResNet, Res2Net, and ResNext are used as backbone networks, respectively. As shown in (a)-(f).

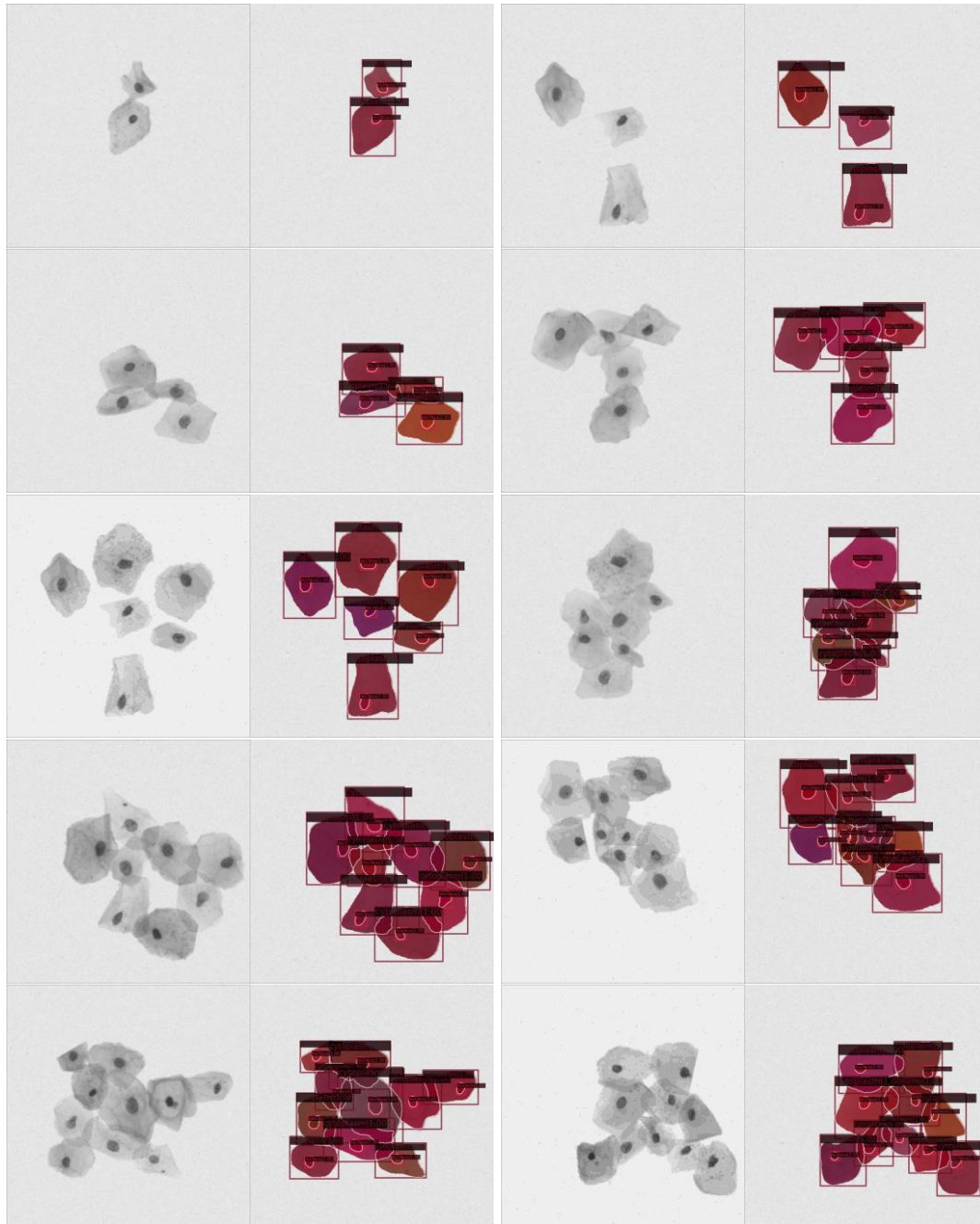


Fig. 9. More examples of visualization results for instance segmentation.

## V. DISCUSSION

In this study, we utilized an isomorphic multi-branch modulation deformable convolutional residual network named IMBMDCR-Net, which can extract features from

images in a training model. since the propagation efficiency of low-level feature information has a great impact on enhancing the global feature level. however, too much attention is paid to the extraction of deep features in the residual network, resulting in the insufficient utilization of



the location information of low-level features, and a large number of computing tasks will be generated in the deep network. Hence, we chose IMBMDCR-Net not only to reduce the amount of calculation but also to adapt the deformable convolution kernel to adjust the receptive domain, which reduces the loss of extracting high-level feature image information and refines cell boundaries. Our results imply that the IMBMDCR-Net method improves the accuracy of different morphological cells.

So far, the Mask R-CNN method is used to perform segmentation based on the pixel-level prior information. However, the instance segmentation is based on candidate regions, which are easily prone to noise interference. Cascade R-CNN cascades several detection networks with different IoU thresholds, which can improve the accuracy of target detection. therefore, we chose the integrated optimization algorithm based on Cascaded RCNN to realize a more accurate segmentation of cervical cells. The proposed method obtains better performance than the traditional Mask RCNN methods on the ISBI2014 dataset, and we can successfully divide the Overlapping cells into multiple single cells, according to the experimental results. We believe the results are helpful for the subsequent cervical cancer automatic cytological screening analysis.

## VI. CONCLUSION

It is different from the previous algorithm that the algorithm proposes a feature extraction network IMBMDCR-Net based on Cascade R-CNN to achieve the extraction of cervical cell boundaries. Six backbone networks conducted the effectiveness experiments on 900 ISBI images, which can be used as an aid for pathological analysis of precancerous lesions. The ubiquitous overlapping cells, which affect the accuracy of quantitative estimation. The main purpose of this work is to divide the overlapping area into multiple single cells. Although the algorithm has achieved some results, it is still far from the complicated real cervical liquid-based cytology samples and will be further studied from the perspective of the adaptability of the algorithm in the future.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

For this work, Qinghua Li carried out algorithm design, data analysis, and overall planning. Weiwei Yue directed the writing of the paper and the implementation of the algorithm. Yanjing Ding analyzed the data, completed the labeling of the samples and the writing of the paper.

## REFERENCES

- [1] J. Ferlay, I. Soerjotmataram, M. Ervik, R. Dikshit, S. Eser, and C. Mathers, "Latest world cancer statistics: Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed," 2013.
- [2] T. A. Potapova, J. Zhu, and R. Li, "Aneuploidy and chromosomal instability: a vicious cycle driving cellular evolution and cancer genome chaos," *Cancer and Metastasis Reviews*, vol. 32, no. 3-4, pp. 377–389, 2013.
- [3] Sarwar, Abid, A. A. Sheikh, and N. Koul, "Segmentation of cervical cells for automated screening of cervical cancer: A review," 2017.
- [4] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," 2017.
- [5] K. K. Thanammal, J. S. Jayasudha, R. R. Vijayalakshmi, and S. Arumugaperumal, "Effective histogram thresholding techniques for natural images using segmentation," *Journal of Image and Graphics*, vol. 2, no. 2, pp. 113–116, December 2014. doi: 10.12720/joig.2.2.113-116
- [6] N. Richard, C. Fernandez-Maloigne, C. Bonanomi, and A. Rizzi, "Fuzzy color image segmentation using watershed transform," *Journal of Image and Graphics*, vol. 1, no. 3, pp. 157–160, September 2013. doi: 10.12720/joig.1.3.157-160
- [7] R. Kaur and G. Malik, "An image segmentation using improved FCM watershed algorithm and DBMF," *Journal of Image and Graphics*, vol. 2, no. 2, pp. 106–112, December 2014. doi: 10.12720/joig.2.2.106-112
- [8] N. Borzooie, H. Danyali, and M. S. Helfroush, "Modified density-based data clustering for interactive liver segmentation," *Journal of Image and Graphics*, vol. 6, no. 1, pp. 84–87, June 2018. doi: 10.18178/joig.6.1.84-87
- [9] M. Rezaee and M. B. T. Avakoli, "Cancerous masses segmentation by using heuristic ant colony algorithms in medical images," *Journal of Image and Graphics*, vol. 2, no. 2, pp. 128–134, December 2014. doi: 10.12720/joig.2.2.128-134
- [10] A. Hap, "A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images," *Computerized Medical Imaging and Graphics*, vol. 59, pp. 38–49, 2017.
- [11] T. Chankong, N. Theera-Umporn, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in Pap smears," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 539–556, 2014.
- [12] N. Ashidi, and M. Isa, "Automated edge detection technique for pap smear images using moving k-means clustering and modified seed based region growing algorithm," 2013.
- [13] A. Tareef, "Multi-pass fast watershed for accurate segmentation of overlapping cervical cells," *IEEE Transactions on Medical Imaging*, 1-1, 2018.
- [14] J. Zhang, "Segmentation of overlapping cells in cervical smears based on spatial relationship and overlapping translucency light transmission model," *Pattern Recognition*, vol. 60, pp. 286–295, 2016.
- [15] H. Lee and J. Kim, "Segmentation of overlapping cervical cells in microscopic images with superpixel partitioning and cell-wise contour refinement," *Computer Vision & Pattern Recognition Workshops IEEE*, pp. 1367–1373, 2016.
- [16] A. Tareef, S. Yang, W. Cai, D. D. Feng, and C. Mei, "Automated three-stage nucleus and cytoplasm segmentation of overlapping cells," presented at International Conference on Control Automation Robotics & Vision IEEE, 2015.
- [17] D. M. Ushizima, A. C. Bianchi, and C. Carneiro, "Segmentation of subcellular compartments combining superpixel representation with Voronoi diagrams," presented at International Symposium on Biomedical Imaging, 2014.
- [18] Y. Wang, X. Lu, and M. Zhu, "A fast active contour tracking method based on Gaussian mixture model," Springer International Publishing, 2014.
- [19] Long, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017.
- [20] R. Olaf, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Springer International Publishing, 2015.
- [21] Y. Enokiya, Y. Iwamoto, Y.-W. Chen, and X.-H. Han, "Automatic liver segmentation using U-Net with Wasserstein GANs," *Journal of Image and Graphics*, vol. 7, no. 3, pp. 94–101, September 2019. doi: 10.18178/joig.7.3.94-101
- [22] N. Shigeta, M. Kamata, and M. Kikuchi, "Effectiveness of Pseudo 3D feature learning for spinal segmentation by CNN with U-Net architecture," *Journal of Image and Graphics*, vol. 7, no. 3, pp. 107–111, September 2019. doi: 10.18178/joig.7.3.107-111
- [23] K. Otsuki, Y. Iwamoto, Y.-W. Chen, A. Furukawa, and S. Kanasaki, "Cine-MR image segmentation for assessment of small bowel motility function using 3D U-Net," *Journal of Image and Graphics*, vol. 7, no. 4, pp. 134–139, December 2019. doi: 10.18178/joig.7.4.134-139
- [24] K. Ono, Y. Iwamoto, Y.-W. Chen, and M. Nonaka, "Automatic segmentation of infant brain ventricles with hydrocephalus in MRI based on 2.5D U-Net and transfer learning," *Journal of Image and Graphics*, vol. 8, no. 2, pp. 42–46, June 2020. doi: 10.18178/joig.8.2.42-46
- [25] A. N. Almustofa, A. Handayani, and T. L. R. Mengko, "Optic disc and optic cup segmentation on retinal image based on multimap

- localization and U-Net convolutional neural network,” *Journal of Image and Graphics*, vol. 10, no. 3, pp. 109–115, September 2022.
- [26] A. O. Joshua, F. V. Nelwamondo, and G. Mabuza-Hocquet, “Blood vessel segmentation from fundus images using modified U-Net convolutional neural network,” *Journal of Image and Graphics*, vol. 8, no. 1, pp. 21–25, March 2020. doi: 10.18178/joig.8.1.21-25
- [27] Y. Huang, “Segmentation of overlapping cervical smear cells based on U-Net and improved level set,” presented at 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) IEEE, 2019.
- [28] R. Girshick, “Rich feature hierarchies for accurate object detection and semantic segmentation tech report (v5),” 2017.
- [29] Kurnianingsih, “Segmentation and classification of cervical cells using deep learning,” *IEEE Access*, 2019.
- [30] M. Long, “Cervical cell TCT image detection and segmentation based on multi-scale feature fusion,” presented at 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) IEEE, 2021.
- [31] A. P. Bradley, G. Carneiro, and Z. Lu, *Overlapping Cervical Cytology Image Segmentation Challenge-ISBI 2014*.
- [32] H. Kaiming, “Mask r-cnn,” *ICCV*, 2017.
- [33] S. Ren, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [34] Y. Liu, “Automatic segmentation of cervical nuclei based on deep learning and a conditional random field,” *IEEE Access*, vol. 6, pp. 53709–53721, 2018.
- [35] H. Kaiming, “Deep residual learning for image recognition,” *IEEE*, 2016.
- [36] C. Szegedy, “Going deeper with convolutions,” *IEEE Computer Society*, 2014.
- [37] X. Zhu, “Deformable ConvNets V2: More deformable, better results,” presented at 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2019.
- [38] T. Y. Lin, “Feature pyramid networks for object detection,” presented at 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE Computer Society, 2017.
- [39] K. Chen, “MMDetection: Open MMLab Detection Toolbox and Benchmark,” 2019.
- [40] W. Yuxin, A. Kirillov, and F. Massa. (2019). Detectron2. [Online]. Available: <https://github.com/facebookresearch/detectron2>,
- [41] N. Carion, *End-to-End Object Detection with Transformers*, 2020.
- [42] X. Zhu, *Deformable DETR: Deformable Transformers for End-to-End Object Detection*, 2020.

Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).