

Enhanced Rotated Mask R-CNN for Chromosome Segmentation

Penglei Wang, Wenjing Hu, Jiping Zhang, Yaofeng Wen, Chenming Xu*, Dahong Qian*

Abstract— Karyotyping is an important process for finding chromosome abnormalities that could cause genetic disorders. This process first requires cytogeneticists to arrange each chromosome from the metaphase image to generate the karyogram. In this process, chromosome segmentation plays an important role and it is directly related to whether the karyotyping can be achieved. The key to achieving accurate chromosome segmentation is to effectively segment the multiple touching and overlapping chromosomes at the same time identify the isolated chromosomes. This paper proposes a method named Enhanced Rotated Mask R-CNN for automatic chromosome segmentation and classification. The Enhanced Rotated Mask R-CNN method can not only accurately segment and classify the isolated chromosomes in metaphase images but also effectively alleviate the problem of inaccurate segmentation for touching and overlapping chromosomes. Experiments show that the proposed approach achieves competitive performances with 49.52 AP on multi-class evaluation and 69.96 AP on binary-class evaluation for chromosome segmentation.

I. INTRODUCTION

Karyotyping plays an important role in the screening and diagnosis of birth defects. The normal human karyotype includes 22 pairs of autosomal chromosomes and one pair of sex chromosomes. To achieve karyotyping, the cytogeneticists first need to stain cells and record the metaphase images under a microscope. Then the metaphase images will be preprocessed and individual chromosomes will be segmented from the preprocessed metaphase images. After that, the single chromosome will be classified to generate a karyogram. Cytogeneticists will analyze each chromosome in a karyogram to find chromosome abnormalities that are related to genetic disorders. Among the process of karyotyping, chromosome segmentation is a vital step because the quality of segmentation will directly affect the results of karyotyping. Unfortunately, the chromosomes in metaphase images are complex and the clusters of touching and overlapping chromosomes exist in almost all metaphase images. In traditional methods, cytogeneticists need to spend a lot of time and energy to manually segment touching and overlapping chromosomes.

With the wide application of karyotyping in genetic diagnosis, there are many methods to solve the problem of chromosome segmentation. Based on traditional image processing algorithms, the global threshold is used by the OTSU method [1] or rethresholding scheme [2] to identify every single chromosome. K-means clustering [3] and watershed segmentation [4] are also applied to metaphase

images for chromosome segmentation. From the perspective of chromosome geometry, Popescu et al. [5] proposed the Cross Section Sequence Graph (CSSG) method for shape analyzing and Agam et al. [6] applied Discrete Curvature Function (DCF) methods for segmenting touching and overlapping chromosomes. With the advent of deep learning that tends to employ convolutional neural networks (CNNs) for segmentation tasks, Esteban et al [7] used Fully Convolutional Networks (FCN) for semantic segmentation in the M-FISH of chromosome images. Hu et al. [8] and Saleh et al. [9] applied U-Net to segment partially overlapping and non-overlapping chromosomes. However, it is still difficult to apply these methods to segment chromosomes in practice for karyotyping, because they are not directly deal with the actual complicated metaphase images and only consider typical touching and overlapping problems, but the actual situation in metaphase images is more complicated.

This paper proposes a new method for automatic chromosome segmentation in metaphase images. This method views chromosome segmentation as an instance segmentation task that requires not only classifying the chromosomes but also separating them from the neighboring instances. Some existed box-free methods handle instance segmentation problems directly without relying on additional detection boxes. For example, Deep Snake [10] is a contour-based approach that iteratively deforms an initial contour to match the object boundary and can reduce the redundant detection through regressing the contour of instances. In contrast, box-based methods perform instance segmentation on the pixel level within a region proposal. Mask R-CNN [11] is a typical instantiation that first detects the bounding boxes of objects by a region proposal network (RPN), then uses a branch to segment instances within the detected boxes. At the same time, Mask R-CNN is a multi-task model that can achieve detection, classification and segmentation through different branches, so it is effective for automatic chromosome recognition and segmentation. This paper achieves the chromosome instance segmentation based on the Mask R-CNN framework. Meanwhile, considering the slender shape of chromosomes, we believe that detecting rotated bounding boxes in Mask R-CNN, which is the minimum circumscribed rectangle of the objects, can reduce redundant detection and improve the quality of segmentation for touching and overlapping chromosomes. So this paper attempted to use Rotated Mask R-CNN [12] for chromosome segmentation. To enhance the performance of chromosome segmentation, we further improve the segmentation accuracy through Enhanced Rotated Mask R-CNN, which combines axis aligned and rotated

Penglei Wang and Jiping Zhang are with the School of Biomedical Engineering, Shanghai Jiao Tong University, China.

Wenjing Hu is with The International Peace Maternity and Child Health Hospital of China, China.

Yaofeng Wen is with Institute of Medical Robotics, Shanghai Jiao Tong University, China

Chenming Xu is with The International Peace Maternity and Child Health Hospital of China, China (*corresponding author to email chenming_xu2006@163.com).

Dahong Qian is with Institute of Medical Robotics, Shanghai Jiao Tong University, China (*corresponding author to email dahong.qian@sjtu.edu.cn).

bounding box detection and extracts additional position information by endpoint detection branch.

II. METHOD

A. Dataset

The metaphase images and the corresponding karyograms are collected from The International Peace Maternity and Child Health Hospital of China (IPMCH). All data used in this paper is approved by the Hospital's Ethical Review Board. The cells are stained with the G-staining method and recorded by the MetaSystems' Metafer 4. To obtain the available dataset for chromosome instance segmentation, we first crop the karyogram to obtain each chromosome, then use the SIFT [13] and DBSCAN [14] algorithm to match the corresponding chromosome in metaphase image. After obtaining the matching results, we further screened and manually labeled the data to obtain a high-quality chromosome instance segmentation dataset. And we obtain the approximate endpoints of each chromosome by using the skeleton extraction algorithm to process the existed chromosome mask. Finally, we get the high-quality benchmark dataset which includes the mask label, category label and endpoints for each chromosome in the metaphase image. The benchmark dataset includes 1378 metaphase images with annotated chromosomes, we randomly select 1103 images for training, 138 images for validation and 137 images for testing.

B. Network Architecture

Our work is inspired by the Mask R-CNN for object instance segmentation. In order to segment the individual chromosome in metaphase images, we view each chromosome as an instance object, which means that we can use existed instance segmentation methods to solve this problem. Firstly, we use Rotated Mask R-CNN to replace the axis aligned bounding boxes by rotated bounding boxes in detection for the special geometry features of chromosomes. Secondly, we have modified the architecture of Rotated Mask R-CNN to obtain the Enhanced Rotated Mask R-CNN which enables the network to detect axis aligned and rotated bounding boxes simultaneously and adds an endpoint detection branch in the network. Fig. 1 shows the overall framework of Enhanced Rotated Mask R-CNN.

In contrast to Mask R-CNN, the Rotated Mask R-CNN uses rotated bounding boxes in detection. For the touching and overlapping chromosomes, axis aligned bounding box likely contains more than one chromosome and results in

missing detection and segmentation. To avoid this problem, we try to use the rotated bounding box to detect the slender chromosome. Usually, the representation of the axis aligned box is $(x_{min}, y_{min}, x_{max}, y_{max})$. It respectively represents the coordinates of the top left and bottom right corners of the bounding box. And we use (x, y, w, h, θ) to uniquely represent a rotated bounding box, which represents the coordinate of the center point and side length of the bounding box. The parameter $\theta \in [-90, 0)$ is the angle at which the horizontal axis rotates counterclockwise to the longer edge of the bounding box. The predicted offsets for rotated bounding box regression can be described as:

$$\begin{aligned} d_x &= \frac{\hat{p}_x - p_x}{w}, d_y = \frac{\hat{p}_y - p_y}{h}, \\ d_w &= \ln \frac{\hat{p}_w}{p_w}, d_h = \ln \frac{\hat{p}_h}{p_h}, d_\theta = \frac{\pi}{180} (\hat{p}_\theta - p_\theta) \end{aligned} \quad (1)$$

where \hat{p}_* is the predicted parameters, p_* is the ground truth ($* \in \{x, y, w, h, \theta\}$). Finally, the Rotated Mask R-CNN replaces the axis aligned bounding boxes with rotated bounding boxes to achieve better detection accuracy for touching and overlapping chromosomes. The accurate detection means that effective segmentation can be performed in a smaller area, which will significantly reduce the difficulty of segmentation tasks. We adopt an end-to-end approach to train the Rotated Mask R-CNN in this paper, and the loss function L_R can be represented as:

$$L_R = L_{cls} + L_{box} + L_{mask} + L_{RPN} \quad (2)$$

where L_{cls} is the cross-entropy loss for classification, L_{box} is the smooth l_1 loss expressed in (3) for rotated bounding box regression, where the x is the error between predicted parameters and labels, L_{mask} is the average binary cross-entropy for the binary mask of resolution $m \times m$ and L_{RPN} is the loss for RPN, which includes a binary classification loss and the region proposal regression loss.

$$smooth_l_1(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \quad (3)$$

In order to enhance the quality of chromosome segmentation, we have further improved the Rotated Mask R-CNN. Firstly, we make the network detect both the axis aligned and rotated bounding boxes, and then synthesize the predicted results to combine the advantages of rotated bounding boxes for touching and overlapping chromosomes

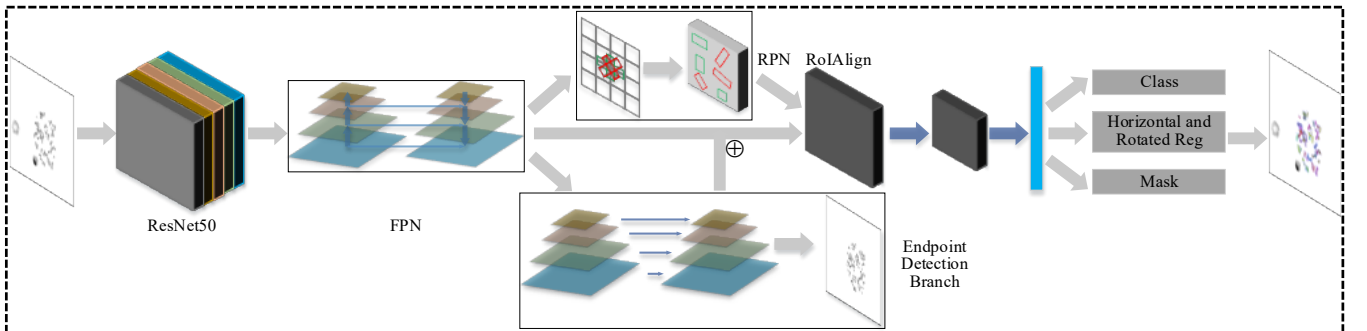


Fig. 1. The overall framework of Enhanced Rotated Mask R-CNN. The input image is fed into ResNet50 and FPN backbones and the endpoint detection branch provides additional position information. Then RoIs is generated via RPN and RoIAlign. The network predicts class, bounding box and mask respectively through different branches.

and axis aligned bounding boxes for better detection in isolated chromosomes. Secondly, we add an endpoint detection branch in the network to provide additional position information of chromosomes. In general, the slender chromosome can be roughly located by its two endpoints, so we consider that add an endpoint detection branch as an additional task can provide more abundant information for chromosome location. The branch aims to predict a heatmap that generates from all endpoints in a metaphase image. The five feature maps from FPN are fed into the branch and it consists of 6 convolution layers and interpolate operation to obtain a heatmap of the same size as the input image. The features of the fifth convolution layer are directly added to the original feature maps from FPN to introduce additional position information for the next tasks. And the loss function of network can be updated like this:

$$L = L_R + \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (x_{ij} - \hat{x}_{ij})^2 \quad (4)$$

where N and M represent the size of the input image is $N \times M$.

III. RESULTS

The network is trained with 100 epochs while the input images are resized to 640×512 , and the training process will early stop when the mask loss doesn't decrease in 10 epochs later. The network parameters are optimized with Stochastic Gradient Descent (SGD). The momentum and learning rate is set as 0.9 and 0.005 respectively. The network is implemented by maskrcnn-benchmark and trained on RTX 2080Ti. After obtaining masks based on the top 100 detection boxes, we use non-maximum suppression (NMS) operation on the results at mask level to eliminate the redundant results. Finally, we use the average precision (AP) metric to evaluate the chromosome instance segmentation performance of the network.

A. Results of Rotated Mask R-CNN

We first evaluate the performance of Rotated Mask R-CNN for chromosome segmentation. We compare the instance segmentation results for chromosomes with Mask R-CNN (baseline) and Deep Snake. And we calculate the AP metric of the networks under the multi-class and binary-class conditions. The AP under multi-class condition reflects the average segmentation performance in each category, and the AP under binary-class condition reflects the performance without considering chromosome classification results. The quantitative and qualitative results are reported in Tab I and Fig. 2. We can observe that Rotated Mask R-CNN performs better than other methods for chromosome segmentation in Tab I. The AP for Rotated Mask R-CNN is 47.87 which is 12.82% higher than Mask R-CNN. The results show that the rotated bounding box provides more accurate detection results and is beneficial to improve the chromosome segmentation performance. As shown in Fig. 2, since the axis aligned bounding box doesn't fit the shape of chromosome, the axis aligned bounding box lacks a precise location description for the touching and overlapping chromosomes and the overlapping bounding boxes will also be suppressed in NMS operation. It leads to Mask R-CNN can't effectively differentiate the touching and overlapping chromosomes.

However, Rotated Mask R-CNN performs well in separating the touching and overlapping chromosomes because the rotated bounding boxes can significantly reduce redundant detection for the slender chromosomes.

TABLE I. CHROMOSOME INSTANCE SEGMENTATION EVALUATION RESULTS WITH DIFFERENT NETWORKS UNDER THE CONFIDENCE SCORE THRESHOLD OF 0.5. MR, RMR AND ERMR REPRESENT MASK R-CNN, ROTATED MASK R-CNN AND ENHANCED ROTATED MASK R-CNN RESPECTIVELY.

Network	AP	AP@0.5	AP@0.75	Mode
Deep Snake	47.01	58.26	56.95	
MR	42.43	64.16	55.99	Multi-class
RMR	47.87	64.27	61.15	
ERMUR	49.52	65.89	63.11	
Deep Snake	64.16	80.87	77.83	
MR	58.23	90.72	75.25	Binary-class
RMR	66.04	90.81	84.39	
ERMUR	69.96	95.83	89.71	

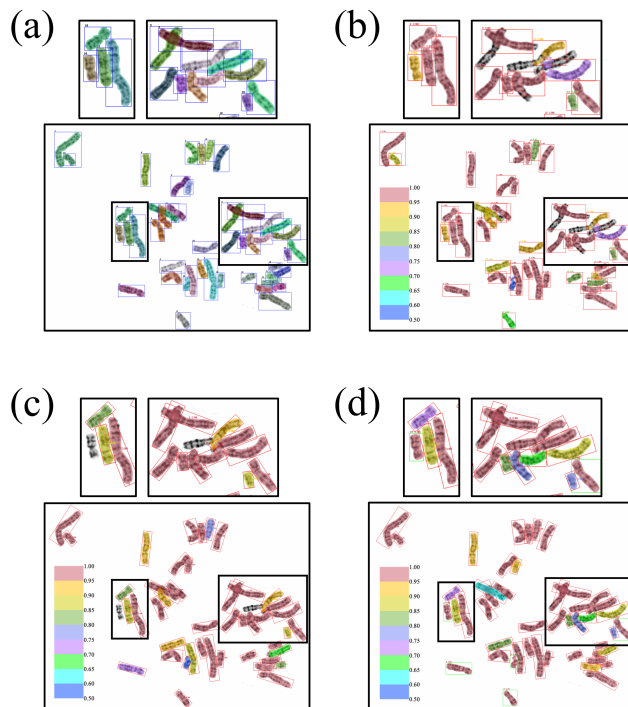


Fig. 2. Qualitative chromosome instance segmentation results in metaphase images. The color bar from bottom to top in the figure represents the confidence score from 0.5~1.0. (a) Ground truth. (b) The segmentation results of Mask R-CNN. (c) The segmentation results of Rotated Mask R-CNN. (d) The segmentation results of Enhanced Rotated Mask R-CNN.

B. Results of Enhanced Rotated Mask R-CNN

As shown in the Tab I, the Enhanced Rotated Mask R-CNN obtain the best performance in all methods. In order to more accurately evaluate the performance of Enhanced Rotated Mask R-CNN, the performance of ablation experiments under the multi-class condition are reported in Tab II. ERMR-1 and ERMR-2 are designed to evaluate the effect of the endpoint detection branch. There is slight performance improvement when we add the endpoint

detection branch in Rotated Mask R-CNN. We consider that the endpoint detection branch prediction the endpoints of chromosome which can provide better position information for the segmentation task. However, we still need more research to design the architecture of endpoint detection branch for more sophisticated endpoint detection, which means that it can provide richer location information for chromosome detection and segmentation. ERM-1 and ERM-3 show that detecting the axis aligned and rotated bounding boxes at the same time also has a certain positive effect, and leads to 1.09 AP performance improvement. For the isolated chromosomes in the metaphase images, using the axis aligned and rotated bounding box will not influence the results at the segmentation step, but the rotated bounding box means more complicated detection process and may lead to missing detection for isolated chromosomes at the detection step. Therefore, detecting the axis aligned and rotated bounding box at the same time can make the model accurately detect the touching and overlapping chromosomes by rotated bounding box detection while avoiding the missed detection for isolated chromosomes by axis aligned bounding box detection. Finally, we obtain the best AP with 0.4952 by combing the axis aligned and rotated bounding box detection and endpoint detection branch in ERM-4, which is 16.71% higher than the Mask R-CNN baseline and 3.45% higher than the Rotated Mask R-CNN.

TABLE II. COMPARISON OF PERFORMANCE FOR EACH SEGMENTATION METHOD UNDER THE CONFIDENCE SCORE THRESHOLD OF 0.5 ON MULTI-CLASS. H&R, ED REPRESENT HORIZONTAL AND ROTATED BOUNDING BOXES DETECTION, ENDPOINT DETECTION RESPECTIVELY. THE ERM-1 EQUALS TO RMR IN TABLE I, THE ERM-4 EQUALS TO ERM IN TABLE I.

Network	H&R	ED	AP	AP@0.5	AP@0.75
ERM-1	×	×	47.87	64.27	61.15
ERM-2	×	✓	47.98	64.26	61.28
ERM-3	✓	×	48.96	65.96	63.37
ERM-4	✓	✓	49.52	65.89	63.11

IV. CONCLUSION

In this paper, we proposed to use a general instance segmentation network for chromosome segmentation in metaphase images, which can be directly applied to the actual process of karyotyping. Considering that the special morphological features of chromosomes in metaphase images, we first applied the Rotated Mask R-CNN to the chromosome segmentation task and obtained a significant improvement in the performance of chromosome segmentation.

In order to enhance the performance of the Rotated Mask R-CNN for chromosome segmentation, we further proposed Enhanced Rotated Mask R-CNN to detect axis aligned and rotated bounding boxes at the same time, which enable the network to combine the advantages of axis aligned and rotated bounding boxes for chromosome detection in chromosome segmentation. Meanwhile, we have added an endpoint detection branch in the network to obtain additional position information for chromosomes. The experimental results demonstrate the advantages of our method for chromosome instance segmentation and indicate our method can effectively promote the development of the automatic chromosome karyotyping process.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China under Grant 81974276.

REFERENCES

- [1] L. Ji, "Intelligent splitting in the chromosome domain," *Pattern Recognition*, vol. 22, no. 5, pp. 519-532, 1989/01/01/ 1989.
- [2] L. Ji, "Fully automatic chromosome segmentation," (in eng), *Cytometry*, vol. 17, no. 3, pp. 196-208, Nov 1 1994.
- [3] B. Lerner, "Toward a completely automatic neural-network-based human chromosome analysis," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 28, no. 4, pp. 544-552, 1998.
- [4] P. Karvelis, D. Fotiadis, M. Syrou, I. Georgiou, and Greece, "Segmentation of chromosome images based on a recursive watershed transform," vol. 11, pp. 1727-1983, 01/01 2005.
- [5] M. Popescu, P. Gader, J. Keller, C. Klein, J. Stanley, and C. Caldwell, "Automatic karyotyping of metaphase cells with overlapping chromosomes," *Computers in Biology and Medicine*, vol. 29, no. 1, pp. 61-82, 1999/01/01/ 1999.
- [6] G. Agam and I. Dinstein, "Geometric separation of partially overlapping nonrigid objects applied to automatic chromosome classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1212-1222, 1997.
- [7] E. Pardo, J. M. Morgado, and N. Malpica, "Semantic segmentation of mFISH images using convolutional networks," *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, vol. 93 6, pp. 620-627, 2018.
- [8] R. L. Hu, J. Karnowski, R. Fadely, and J.-P. Pommier, "Image Segmentation to Distinguish Between Overlapping Human Chromosomes," *ArXiv*, vol. abs/1712.07639, 2017.
- [9] H. M. Saleh, N. H. Saad, and N. A. M. Isa, "Overlapping Chromosome Segmentation using U-Net: Convolutional Networks with Test Time Augmentation," *Procedia Computer Science*, vol. 159, pp. 524-533, 2019/01/01/ 2019.
- [10] S. Peng, W. Jiang, H. Pi, H. Bao, and X. Zhou, "Deep Snake for Real-Time Instance Segmentation," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8530-8539, 2020.
- [11] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980-2988, 2017.
- [12] S. Looi. Rotated Mask R-CNN: From Bounding Boxes To Rotated Bounding Boxes [Online] Available: https://github.com/mrlooi/rotated_maskrcnn
- [13] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004/11/01 2004.
- [14] S. Louhichi, M. Gzara, and H. B. Abdallah, "A density based algorithm for discovering clusters with varied density," in 2014 World Congress on Computer Applications and Information Systems (WCCAIS), 17-19 Jan. 2014 2014, pp. 1-6.