KEMR-Net: A Knowledge-Enhanced Mask Refinement Network for Chromosome Instance Segmentation

Renhao Zhou [a], Linfeng Yu [a], Ding Chen [a], Haoxi Zhang [a b], Edward Szczerbicki [c]

[a] School of Cybersecurity, Chengdu University of Information Technology, Chengdu, China; [b] Advanced Cryptography and System Security Key Laboratory of Sichuan Province, Chengdu, China; [c] Faculty of Management and Economics, Gdansk University of Technology, Gdansk, Poland

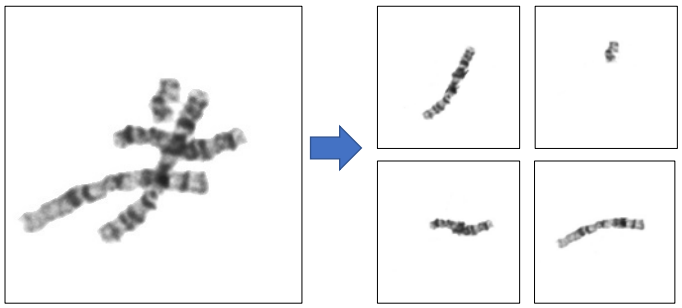
Abstract. This paper proposes a mask refinement method for chromosome instance segmentation. The proposed method exploits the knowledge representation capability of Neural Knowledge DNA (NK-DNA) to capture the semantics of the chromosome's shape, texture, and key points, and then it uses the captured knowledge to improve the accuracy and smoothness of the masks. We validate the method's effectiveness on our latest high-resolution chromosome image dataset. The experimental results show that our proposed method's mask average precision (MaskAP) is 3.66% higher than Mask R-CNN and outperforms advanced Cascade Mask R-CNN by 1.35%.

Address correspondence to Ding Chen, School of Cybersecurity, Chengdu University of Information Technology, No. 24 Block 1, Xuefu Road, Chengdu, China, 610225. E-mail: chending@cuit.edu.cn

Keywords: Chromosome instance segmentation, knowledge-enhanced, mask refinement, Neural Knowledge DNA.

INTRODUCTION

Chromosome segmentation is a fundamental task in human chromosome karyotype analysis. Chromosomes are randomly located and may overlap with others in the metaphase images. The manual segmentation of chromosomes is time-consuming and labour-intensive due to the complexity of the metaphase image (see Figure 1). Therefore, automatic chromosome segmentation can greatly facilitate karyotype analysis.



(a) **(b)**

Figure 1. The instance segmentation of chromosome. The goal of the chromosome instance segmentation is to separate chromosomes into instances (b) from the given image (a).

Traditional chromosome segmentation methods heavily relied on computer image processing techniques, such as threshold-based, region-based, and edge-based segmentation methods (Minaee et al. 2014; Saiyod et al. 2014; Poletti Pallavoor et al. 2012). These methods have achieved some success in pre-processing chromosome images and segmenting non-overlapping chromosomes. However, traditional

segmentation methods based on fixed rules are not robust and fail in complex situations like chromosome overlaps or adhesions.

The feasibility of neural networks in chromosome segmentation was first introduced in Hu's pioneering work (Hu et al. 2017) using a U-Net (Ronneberger et al. 2015). They focus on segmenting overlapping and non-overlapping regions of chromosomes. Their method constrains that overlaps only happen between two chromosomes, and their dataset is constructed under this constraint via synthesis images. Subsequent improvements and attempts based on U-Net networks have been made by other researchers (Mei et al. 2022; Bai et al. 2022). These approaches are highly accurate in segmenting overlapping and non-overlapping regions of chromosomes. However, they require complex pre-processing steps of dividing chromosomes into chromosome clusters, and they can only perform semantic segmentation that cannot produce complete chromosome instances. Moreover, combining the overlapping and non-overlapping regions into complete chromosomes will dramatically increase the method's complexity.

To directly generate the complete mask of each chromosome, instance segmentation methods were introduced by researchers. Several studies (Pijackova et al. 2022; Wang et al. 2022) apply the Mask R-CNN model (He et al. 2017) to chromosome instance segmentation. Lin proposed an improved AS-PathNet based on the path enhancement network (Lin et al. 2020). Feng et al. (2020) proposed a Mask Oriented R-CNN chromosome instance segmentation framework based on components such as a directed enclosing frame, AwIoU metric, and directed convolutional pathway structure.

The existing chromosome instance segmentation networks use the fully convolutional network (FCN) as the segmentation branch. However, the deconvolution operation used in FCN inevitably loses low-dimensional semantic information, resulting in poor performance predicting instance masks. We propose a knowledge-enhanced method for refined masks to improve the accuracy and robustness of chromosome instance segmentation. Our approach uses a cascaded network for initial instance segmentation. Then, it utilizes probability graphs to describe the relationship between the pixel values and the initial masks of chromosome instances. This relationship effectively expresses the knowledge of the chromosome's shape, texture, and key points and is captured by the NK-DNA. The introduction of knowledge into chromosome instance segmentation notably increases the accuracy and robustness of the model. Moreover, our method avoids complex pre-processing and post-matching procedures, making it more practical for real-world applications.

We create a high-resolution chromosome image dataset with 901 metaphase cases to evaluate our proposed method. Our method achieves a 3.66% improvement to the baseline method Mask R-CNN and is 1.35% more accurate than Cascade Mask R-CNN on the Mask AP.

NEURAL KNOWLEDGE DNA

Neural knowledge DNA (NK-DNA) is proposed to store and represent knowledge captured in machine learning systems that use artificial neural networks as the central driving force of their intelligence. The construction of NK-DNA is similar to the formation of DNA (Travers 2015): it is composed of four key elements. When DNA produces a phenotype, neural knowledge DNA carries information and knowledge through its four key components: States, Actions, Experiences, and Networks (see Figure 2).

Figure 2. The conceptual structure of the NK-DNA.

Generally, knowledge is acquired as models after training in machine learning systems. The model usually stores information about weights and biases of the connections between neurons of the neural network and the hierarchy of the neural network in detail. In the NK-DNA, the Networks is used to carry such models, and in this study, we utilize the fully connected conditional random fields (FC-CRFs) (Lafferty et al. 2002) as the Networks to interpret the relationship among all pixels in the masks.

CASCADE MASK R-CNN

Cascade Mask R-CNN is a multi-stage extension of the Faster R-CNN architecture. It aims to get high-quality target detection. It is achieved by combining cascade bounding box regressions. As shown in figure 3: the Cascade Mask R-CNN has three heads linked together in a series. It means that the former head's output is the latter's input. This form of linked structure is called: cascading.

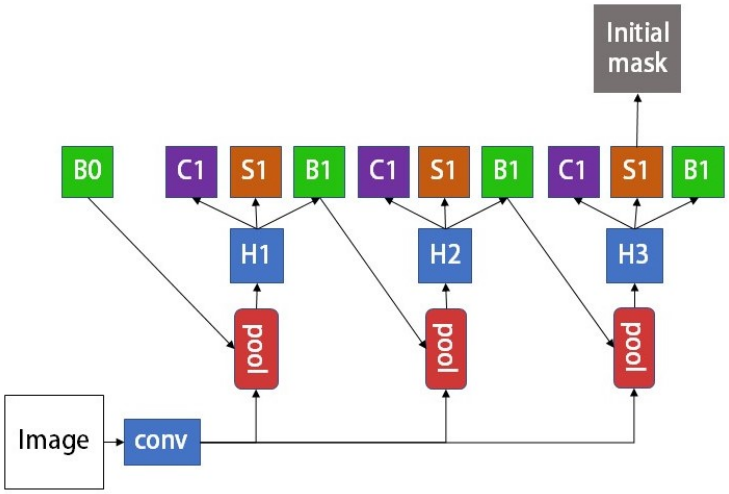


Figure 3. The architectures of Cascade Mask R-CNN

The cascade detector has three increasing IoU thresholds (*i.e.*, 0.5, 0.6, 0.7) for its three heads, respectively. It solves the problem of the degrading performance with increased IoU thresholds and improves the hypotheses quality, ensuring a positive training set for three detectors and minimizing overfitting.

The identical three cascade detectors in the segmentation branches output the initial mask in a sequential manner. The last head produces the initial masks of our approach from an ensemble of three segmentation branches.

KNOWLEDGE-ENHANCED MASK REFINEMENT NETWORK

A. Overview

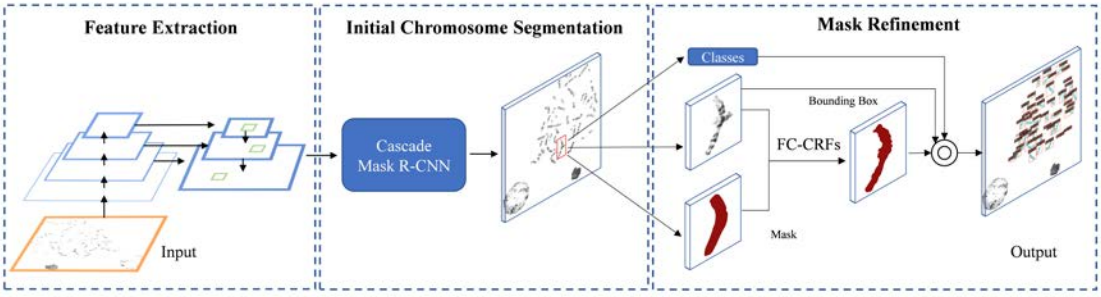


Figure 4. The architecture of our Knowledge-Enhanced Mask Refinement Network.

The proposed mask refinement method for chromosome instance segmentation mainly consists of *Feature Extraction*, *Initial Chromosome Segmentation*, and *Mask Refinement*. As shown in Figure 4: Firstly, a backbone network is used to extract image features, and then, the extracted features are fed into the Cascade Mask R-CNN for predicting classes, bounding boxes, and generating initial masks. Finally, these initial masks and bounding boxes are sent to the FC-CRFs stored as NK-DNA knowledge for refinement. After mask refinement, our approach produces the final masks the same size as the input metaphase images.

B. KNOWLEDGE BUILDING AND REASONING

Humans can naturally utilize the sense of objects' boundaries and contents for segmentation tasks. Inspired by this, we introduce the semantics of boundaries and contents as a kind of knowledge into chromosome instance segmentation in this work. In order to capture the semantics among all pixels in the initial masks, the FC-CRFs (fully connected conditional random fields) are used. FC-CRFs have good Markovianity, and

the maximal cliques of FC-CRFs are all mutually independent sets of random variables, which enables FC-CRFs the good divisibility and effectiveness for refining the chromosome instance masks.

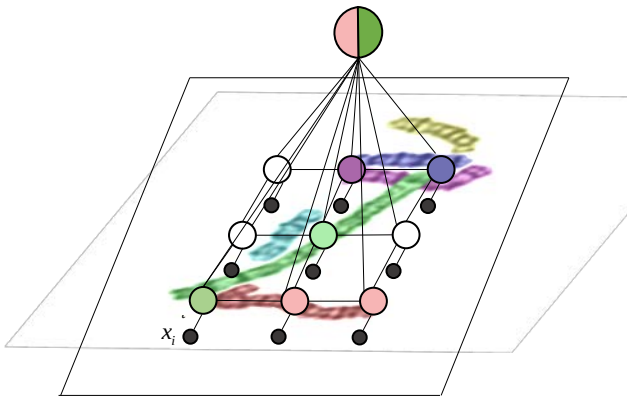


Figure 5. The conditional random field of the chromosome image

First, we build a random field $X = \{X_1, X_2, \cdots, X_N\}$ for the N pixels of a chromosome mask. X_i denotes that the i^{th} pixel is assigned to a category label by the cascade segmentation network. Next, we define another random field $I = \{I_1, I_2, \cdots, I_N\}$, consisting of RGB values of pixels (see Figure 5), where I_i represents the RGB value of the i^{th} pixel, and the distribution of the conditional random field (I, X) conforms to the Gibbs distribution. Therefore, the task of instance mask refinement can be expressed as minimizing the conditional probability:

$$P(X|I) = \frac{1}{Z(I)} e^{-E(X|I)}, \tag{1}$$

where Z is a matrix that has the same size as the instance mask. It is the normalization factor of the Gibbs distribution. In FC-CRFs, the corresponding energy $E(X|I)$ is defined as:

$$E(X) = \sum_i \psi(x_i) + \sum_{i<j} \psi_p(x_i \cdot x_j) \ , \quad (2)$$

where $E(X|I)$ takes values in the range (0, 1], with a maximum value of 1 and a minimum value infinitely converging to 0. The unary potential function $\psi(x_i)$ is used to measure the class probability of a pixel. It represents the probability that the i^{th} pixel predicted to be x_i. And the unary potential comes from the output sampled on the segmentation network. The pairwise potential $\psi_p(x_i \cdot x_j)$ describes the relationship between pixels. We want pixels with the same relationship to be given the same category label, while pixels with dissimilar relationships are determined to be in different categories. The $\psi_p(x_i \cdot x_j)$ is calculated as:

$$\psi_p(x_i \cdot x_j) = \mathrm{u}(x_i \cdot x_j) \sum w\, K(f_i, f_j) \ , \quad (3)$$

which considers the pixels' RGB values and the distance between pixels: Where $u(x_i, x_j)$ represents label compatibility, constraining that energy can be transferred between pixels only if they have the same label. The w is the weight parameter, and $K(f_i, f_j)$ is the feature function representing the pixels' RGB values and distance as:

$$K(f_i, f_j) = w_1 \, exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\beta^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2} \right) + w_2 \, exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2} \right) \ . \quad (4)$$

Where it has two Gaussian kernels: The first Gaussian kernel takes both the pixel's position (denoted as p) and RGB value (denoted as I), while the second only calculates the pixels' distance. The hyperparameters σ_m , σ_β and σ_γ control the "scale" of the Gaussian kernel.

$E(X|I)$ equals 0 when each pixel is correctly classified. Consequently, the optimization process is to minimize this energy by finding the correct parameters w_1 and

w_2 in this conditional distribution. The smaller the energy $E(X|I)$, the more accurate the predicted category label X. As a result, FC-CRFs enable the masks to be segmented along the border as precisely as possible. Figure 6 demonstrates the effectiveness of our method on chromosome instance segmentation.

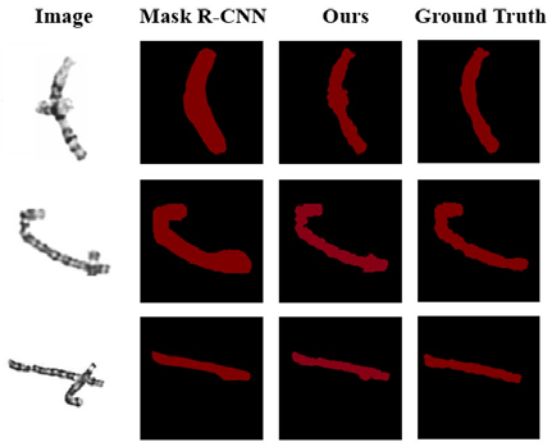


Figure 6. Mask qualitative comparison of Mask R-CNN and the proposed approach.

THE EXPERIMENT

A. DATASET

The chromosome images used by cytogeneticists are not standardized. The resolution of chromosome images undergoes a series of changes (Figure 7). Rather than tackle the pre-process-to-non-overlapping chromosome images shown in Figure 7 (d), we collect and create our high-resolution chromosome dataset from authentic day-to-day clinic case images like Figure 7(e), which have overlaps and adhesions among chromosomes. These images are collected at the Department of Medical Genetics/Prenatal Diagnostic Center, West China Second Hospital. We label images with the annotation tool Labelme and

resize them to 1280 × 1024. The dataset consists of 901 metaphase case images, including 720 case images in the training set and 181 in the test set.

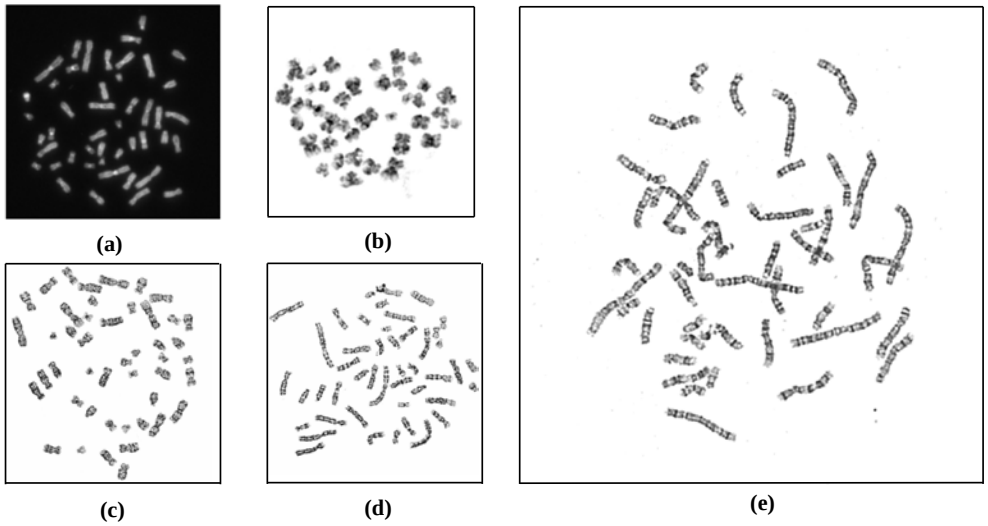


Figure 7. Datasets used for chromosome segmentations. The early chromosome images only have the outlines of chromosomes (a). In recent years, chromosome image resolution has been increasing: (b) and (c) show G200 and G300 resolution chromosome images, and (d) and (e) illustrate resolution G550 resolution chromosome images, respectively. In this work, we create our own dataset using images as (e).

B. EVALUATION METRICS

Bounding box average precision(BoxAP), MaskAP, and Instance Accuracy(IA) metrics are used to evaluate the results. MaskAP is directly related to segmentation accuracy, which is the mean average precision (mAP) under different IoU thresholds. BoxAP measures bounding box average precision. BoxAP and MaskAP are calculated using the method defined in the COCO dataset (Dataset of Common Objects in Context Visual Recognition Challenge (Lin et al. 2014). AP50 signifies the IoU threshold of 0.5 is used to determine whether the predicted mask is positive in the assessment, and the rest of the IoU thresholds were expressed similarly. AP without threshold refers to the average result when the IoU is thresholded at 0.05 intervals between 0.5 to 0.95. Instance

Accuracy refers to the number of instances that are accurately segmented as a percentage of instances in the ground truth, abbreviated as IA:

$$IA = \frac{N'}{N} \quad , \tag{5}$$

where N denotes the actual number of all chromosome instances in the dataset, and N' denotes the number of chromosome instances segmented by the model. The higher metric of IA, the more accurate the model is.

C. EXPERIMENT SETTINGS

The method proposed in this paper is implemented based on Pytorch 1.8.0. The experimental device is a computer with 16GB of RAM and an Intel(R) Core (TM) i7-8700 CPU with ubantu18.04, carrying an NVIDIA TITAN XP with 12GB memory. Each model was set to train for 100 epochs, and all models were trained using ResNet-101 as the backbone. We optimize our model using Adam optimizer, with an initial learning rate of 10^{-3}, and weight decay 10^{-4}. We set the interval to [0, 160,000, 200,000]. Each batch has two images. We adjust the learning rate at 0, 18000, and 240000 iterations and adjust the size of the gamma of the current learning rate.

D. RESULTS

Table 1: Comparison of models for chromosome instance segmentation.

Model	$\boldsymbol{BoxAP}$	$\boldsymbol{MaskAP}$	$\boldsymbol{AP_{50}^{mask}}$	$\boldsymbol{AP_{75}^{mask}}$	$\boldsymbol{IA}$
Mask R-CNN	88.91	75.12	95.35	87.15	90.21
Blend Mask	88.97	76.09	95.86	87.67	90.15
Cascade Mask R-CNN	$\mathbf{91.42}$	77.43	97.47	90.83	95.19
Ours	91.38	$\mathbf{78.78}$	$\mathbf{97.76}$	$\mathbf{91.95}$	$\mathbf{95.24}$

Segmentation accuracy is crucial for chromosome karyotype analysis. We compare the performances of our proposed methods with Mask R-CNN, Blendmask, and Cascade Mask R-CNN. The experiment result is summarised in Table 1. Our method outperforms Mask R-CNN by 3.66% and 2.47% in terms of mask average precision (MaskAP) and bounding box average precision (BoxAP), respectively.

Specifically, in the maskAP50 metric, our method has a 2.41% improvement. In the MaskAP 75, our model obtains a 4.8% improvement. Meanwhile, Cascade Mask R-CNN and our model significantly outperformed Mask R-CNN and Blendmask, which are not cascaded on both BoxAP and IA. Figure 8 shows the prediction comparison between our method and Mask R-CNN. In bounding box and mask accuracy, our method shows better results.

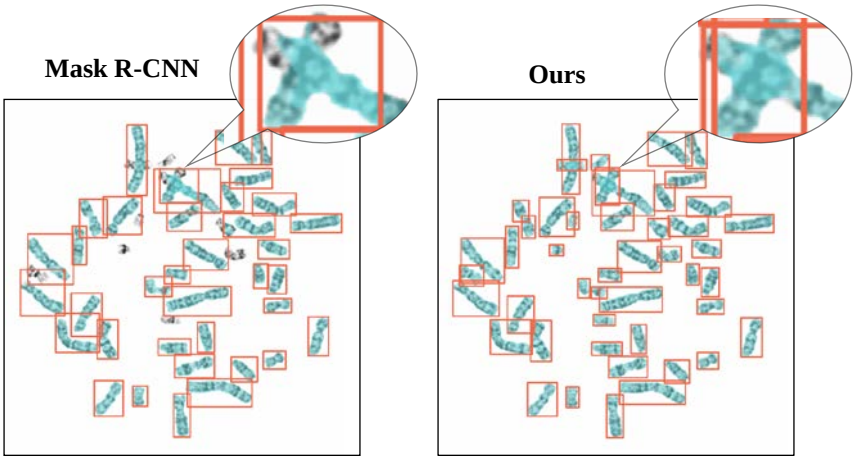


Figure 8. Segmentation result comparison on high-resolution image.

CONCLUSION AND FUTURE WORK

In this paper, we propose KEMR-Net, a knowledge-enhanced mask refinement network for chromosome instance segmentation, and validate it on our latest high-resolution chromosome dataset. Our method uses a cascaded network to produce high-quality chromosome detections and initial instance masks. Then it optimizes the initial masks with the knowledge contained by the NK-DNA. The experiment results show that our method obtains a 3.66% improvement compared to the baseline method (Mask R-CNN) and outperforms Cascade Mask R-CNN by 1.35% on mask AP. There is still room to improve the segmentation accuracy in regions with multiple chromosomes overlapping. We plan to work on analyzing the semantic knowledge of overlapping regions to achieve more accurate segmentation results.

ACKNOWLEDGEMENT

The authors would like to thank the editors and anonymous reviewers for their valuable comments and suggestions on this paper. This work was supported by the Kemoshen Science Research Program.

REFERENCES

Bai, H., Zhang, T., Lu, C., Chen, W., Xu, F., & Han, Z. B. (2020). Chromosome extraction based on U-Net and YOLOv3. IEEE Access, 8, 178563-178569.

Balagalla, U. B., Samarabandu, J., & Subasinghe, A. (2022). Automated human chromosome segmentation and feature extraction: Current trends and prospects. F1000Research, 11(301), 301.

Cai, Z., & Vasconcelos, N. (2019). Cascade R-CNN: high quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence, 43(5), 1483-1498.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

Hu, R. L., Karnowski, J., Fadely, R., & Pommier, J. P. (2017). Image segmentation to distinguish between overlapping human chromosomes. arXiv preprint arXiv:1712.07639.

Krähenbühl, P., & Koltun, V. (2011). Efficient inference in fully connected crfs with gaussian edge potentials. Advances in neural information processing systems, 24.

Lafferty, J. , Mccallum, A. , & Pereira, F. . (2002). Conditional random fields: probabilistic models for segmenting and labeling sequence data. proceedings of icml.

Mei, L., Yu, Y., Shen, H., Weng, Y., Liu, Y., Wang, D., ... & Lei, C. (2022). Adversarial Multiscale Feature Learning Framework for Overlapping Chromosome Segmentation. Entropy, 24(4), 522.

Minaee, S., Fotouhi, M., & Khalaj, B. H. (2014). A geometric approach to fully automatic chromosome segmentation. In 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB) (pp. 1-6). IEEE.

Pijackova, K., Gotthans, T., & Gotthans, J. (2022). Deep Learning Pipeline for Chromosome Segmentation. In 2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA) (pp. 01-05). IEEE.

Poletti, E., Zappelli, F., Ruggeri, A., & Grisan, E. (2012). A review of thresholding strategies applied to human chromosome segmentation. Computer methods and programs in biomedicine, 108(2), 679-688.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Saiyod, S., & Wayalun, P. (2014). A new technique for edge detection of chromosome g-band images for segmentation. In Advanced Approaches to Intelligent Information and Database Systems (pp. 315-323). Springer, Cham.

Sharma, M., Saha, O., Sriraman, A., Hebbalaguppe, R., Vig, L., & Karande, S. (2017). Crowdsourcing for chromosome segmentation and deep classification. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 34-41).

Travers, A., & Muskhelishvili, G. (2015). DNA structure and function. The FEBS journal, 282(12), 2279-2295.

Wang, P., Hu, W., Zhang, J., Wen, Y., Xu, C., & Qian, D. (2021). Enhanced Rotated Mask R-CNN for Chromosome Segmentation. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 2769-2772). IEEE.

Zhang, H., Sanin, C., Szczerbicki, E., & Zhu, M. (2017). Towards neural knowledge DNA. Journal of Intelligent & Fuzzy Systems, 32(2), 1575-1584.