Journal of Artificial Intelligence and Control Systems

CoSci Press

Research Article

UCRD-Net: End-to-End Chromosome Segmentation Based on Uncertainty Correction and Region Decomposition

Xinyu Fan, Hui Liu, Lian Xiong, Hongdang Zheng, Lin Zhang*

School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China.

Article History:

Received: 24 April 2025

Revised: 14 May 2025

Accepted: 25 May 2025

Published: 30 May 2025

Abstract: Chromosome segmentation is the basis of karyotype analysis, and the performance of segmentation directly affects the subsequent classification and identification. However, chromosomes are free in cells, presenting as non-rigid and highly susceptible to overlapping adhesions, which leads to the problems of inaccurate instance detection and incomplete segmented masks in chromosome instance segmentation. In this context, we develop an end-to-end chromosome instance segmentation network, named UCRD-Net, using uncertainty correction and region decomposition to detect and segment more accurately. In this model, the output uncertainty of the uncertainty branch is designed to correct the original classification confidence by Gaussian modeling the detection box of the chromosome instance to improve the positioning accuracy of the detection box. A strategy of decomposing Region of Interest (RoI) is proposed to weaken the semantic ambiguity of overlapping chromosomes in RoI and strengthen the characteristic response of chromosomes to be segmented. Based on the above ideas, this study has completed the end-to-end architecture design. Experiments on clinical datasets show that the average segmentation accuracy of UCRD-Net is 4.1% higher than that of the baseline model Mask R-CNN, AP50 is 0.4% higher, and AP75 is 1.0% higher. The improved accuracy and completeness of segmentation provided by UCRD-Net can offer reliable support for downstream karyotype analysis and facilitate the early detection of chromosomal abnormalities in clinical diagnostic workflows.

Keywords: chromosome segmentation; instance segmentation; uncertainty; region decomposition

1.Introduction

Normal human cells contain 22 pairs of autosomes and one pair of sex chromosomes (XY for males and XX for females) [1]. Chromosome number abnormalities or structural aberrations can lead to genetic diseases such as mental retardation and congenital malformations [2,3], making prevention crucial through prenatal screening and prenatal diagnosis. At present, prevention is mainly carried out by prenatal screening and prenatal diagnosis [4]. The current gold standard for prenatal diagnosis is karyotyping [5], which involves imaging chromosomal ribbons during mid-division in amniotic fluid cell samples, followed by segmenting and sorting to identify abnormalities [6]. However, the free and flexible nature of chromosomes within cells [7], as well as the occurrence of multiple overlapping crosses and varying morphologies, makes karyotype analysis heavily reliant on professionals. It is expensive, inefficient, and highly dependent on experience. Consequently, there is an urgent need for efficient and accurate intelligent aids [8]. Chromosome segmentation, as the primary task in karyotype analysis, serves as the foundation for subsequent classification and identification, with a particular focus on addressing the

---

* Corresponding author: Lin Zhang, School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China, lin.zhang@cumt.edu.cn

segmentation of overlapping chromosomes [9].

Traditional overlapping chromosome segmentation is mainly based on geometric morphology. For example, Grisan et al. first segmented overlapping chromosome clusters using a local adaptive method, then combined it with geometric feature untangling to segment overlapping chromosomes [10]. Minaee et al. achieved chromosome segmentation by identifying chromosome boundary crossing points to extract cut lines [11]. Karvelis et al. adopted a recursive approach based on the watershed transform to decompose chromosome regions and then segmented single chromosomes by merging adjacent regions through a multi-channel gradient path [12]. These traditional methods are difficult to cope with the complex and variable chromosome overlap cases, have insufficient generalization ability and low accuracy, and are difficult to implement automatic segmentation. These conventional techniques encounter challenges in handling complex and variable chromosome overlap scenarios, exhibiting limited generalization capabilities and accuracy, which makes it difficult to implement automation in the segmentation process. These traditional geometry-based methods lack robustness when handling complex overlap patterns and often fail in ambiguous regions due to rigid assumptions.

With the rapid development of deep learning in recent years, automatic chromosome segmentation models based on convolutional neural networks have emerged [13-17]. These models have demonstrated significant improvements when compared to traditional algorithms. To employ deep learning algorithms for training chromosome images, an initial exploration of chromosome labeling data is essential. For example, Hu et al. used individually segmented chromosomes from metaphase images to synthesize overlapping chromosome images by rotating and superimposing them two-by-two. They constructed a neural network based on UNet to segment two overlapping chromosomes, which notably improved accuracy and speed in contrast to traditional segmentation methods [18]. Wang et al. developed ARMS Net to address the characteristics of chromosome overlapping regions with varying sizes, and adaptively extracted multi-scale features to enhance the detailed segmentation of chromosome overlapping regions [19]. However, both methods were designed for synthetic datasets containing two-by-two overlapping chromosomes, and their labels are not tailored to real chromosome data, leading to difficulties in applying their algorithms.

To improve segmentation accuracy, some researchers have proposed the correction of detection boxes during the detection stage before segmentation. For example, Wang et al. introduced rotated bounding boxes in the detection phase to accommodate the flexible bending characteristics of chromosomes. They developed an enhanced rotated Mask R-CNN chromosome segmentation model to improve the accuracy of object detection. However, labeling for rotated boxes is intricate and challenging to implement [20]. Liu et al. introduced regression confidence and mask confidence to concurrently correct the classification confidence, addressing the problem of low correlation between classification confidence and object localization accuracy in instance segmentation. Nevertheless, this confidence metric does not incorporate detection box coordinate information and thus cannot reflect the accuracy of localization [21]. To address inaccurate segmentation resulting from overlapping chromosomes, Xu et al. developed a two-stage model called ChromSeg, specifically designed for segmenting individual chromosomes within chromosome clusters. In the first stage, they expanded UNet++ into a dual-output framework, distinguishing between foreground-background and overlapping regions. In the second stage, the integration of foreground and overlapping regions was achieved through joint search, resulting in high computational complexity [22]. On the other hand, Lin et al. constructed AS-PANet according to the characteristics of overlapping chromosomes. They introduced a chromosome counting branch to improve the accuracy of instance segmentation, but their approach is primarily suited for chromosome clusters and is insufficient for the analysis of metaphase chromosome images [23]. However, these deep learning models struggle with the precise localization of overlapping regions and mask integrity, as they typically rely only on classification confidence and lack explicit modeling of uncertainty or semantic interference.

In summary, the overlapping chromosome segmentation network has the following challenges: (1) The privacy concerns and intricacies associated with labeling chromosome images result in a scarcity of labeled data. This scarcity significantly increases the difficulty of applying deep learning methods to this task. (2) The classification confidence does not always reflect the accuracy of localized position prediction. Overlapping chromosomes result in the acquisition of imprecise detection boxes, subsequently leading to missing segmentation masks. (3) The similar appearance characteristics of different chromosomes contribute to semantic ambiguity, making it challenging to correctly segment overlapping chromosomes, as shown in Figure 1.

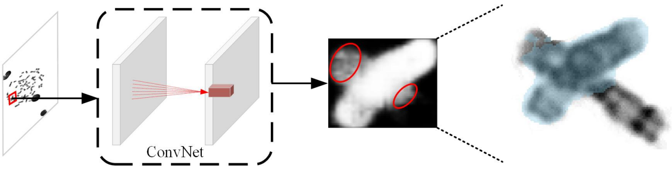


Figure 1. Chromosome overlap causes incorrect segmentation masks.

In order to address the aforementioned problems, we propose an end-to-end chromosome segmentation network called UCRD-Net. Our approach involves several key steps: Firstly, we enhance the effectiveness of capturing the structural characteristics of chromosomes by exploring suitable feature extraction strategies. Then, we model the uncertainty estimate for the detection box as a Gaussian process in our algorithm to correct its position. Finally, we design a region decomposition strategy in each Region of Interest (RoI) generated by detectors to mitigate semantic ambiguity caused by overlapping chromosomes.

The main contributions of this paper are as follows:

(1) To overcome the challenge of limited chromosome data, we have devised tailored feature extraction strategies that leverage the unique characteristics of chromosomes, which are different from natural images. These strategies encompass pre-training, fine-tuning, and the selection of feature maps and anchors of appropriate sizes to enhance the quality of features.

(2) To correct detection boxes and subsequently improve segmentation performance, we innovatively introduce an uncertainty branch. This branch is designed to incorporate reliability information from the detection box into the classification confidence, thereby rectifying the position of the chromosome detection box and generating a more comprehensive mask.

(3) To effectively address the problem of semantic ambiguity caused by overlapping chromosomes, we propose a region decomposition strategy. This strategy dynamically combines the features from different region locations after decomposition to enhance the semantic response of the chromosome targeted for segmentation.

This paper is organized as follows. Section 2 introduces the proposed methods, including the outline of UCRD-Net and details of each module. Section 3 evaluates UCRD-Net comprehensively and compares it with other methods. Section 4 presents our conclusion.

2. Methods

UCRD-Net is implemented through three modules: chromosome feature extraction, uncertainty-based correction of chromosome detection boxes, and chromosome segmentation based on region decomposition, as shown in Figure 2.

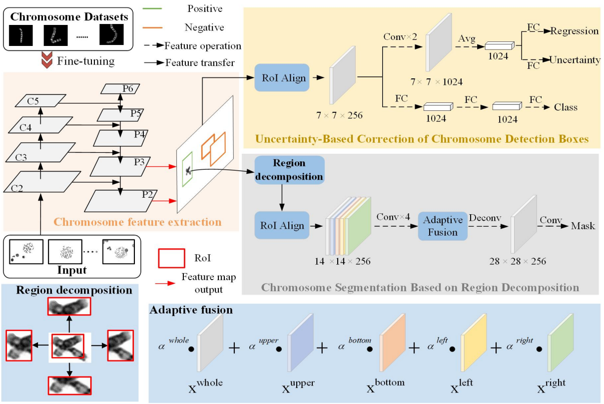


Figure 2. Overview of the proposed UCRD-Net framework. The figure highlights the internal structure of the main processing pipeline and key modules, such as the region decomposition and adaptive fusion.

The chromosome feature extraction module initially undergoes fine-tuning using the public G-band chromosome classification dataset [24] to complete the pre-training of the ResNet101 [25] backbone network. The module responsible for correcting chromosome detection boxes based on uncertainty incorporates dual branch heads. Within this module, the dual branch head enhances localization awareness, the uncertainty branch outputs uncertainty values, which are utilized to correct detection boxes with notable positioning errors. This correction process helps reduce instances of missing segmentation masks resulting from inaccuracies in the detection boxes. The region decomposition module initiates by breaking down the RoI into four directional location regions. Subsequently, features are extracted from these four location directions using RoI Align [26]. These directional features are then adaptively fused with the full RoI features to amplify the semantic response of different chromosomes. This approach effectively mitigates semantic ambiguity within the RoI caused by the presence of other chromosomes and enhances reasoning through multiple directions.

2.1 Chromosomal Feature Extraction

Chromosome banding images offer limited data, while deep learning models typically demand a substantial amount of training data. To maximize feature extraction potential from the available data, the ResNet101 backbone network, commonly used in conjunction with Mask R-CNN [26], is employed for extracting low-level structural features from images. This is accomplished by loading a pre-trained model from ImageNet [27]. Previous methods have traditionally undergone pre-training on ImageNet, but a significant semantic gap exists between natural images and biomedical images, such as chromosomes, which possess distinct semantic characteristics. Consequently, our approach involves fine-tuning the ResNet101 backbone network using the publicly available G-band chromosome classification dataset. Subsequently, we merge these features using Feature Pyramid Networks (FPN) [28], as illustrated in Figure 2.

The proper configuration of anchors in the object detection task plays a crucial role in accurately locating instances. An anchor that is too small may result in a large regression offset for the detection box, potentially leading to overfitting. Conversely, an anchor that is too large may yield a low Intersection over Union (IoU) ratio with the ground-truth, increasing the probability of missed detections. To address this, UCRD-Net improves the extraction of chromosome features by quantifying the pixel occupancy of each chromosome, aiding in the setup of anchors. We calculate the pixel points associated with chromosomes and create a distribution map of their counts, as shown in Figure 3.

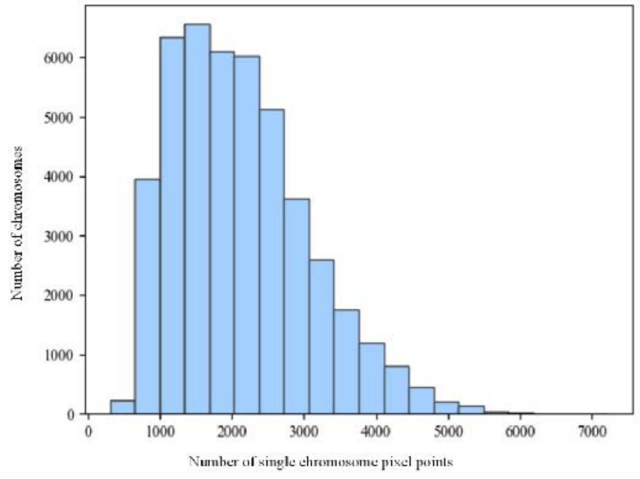


Figure 3. Histogram of chromosome area statistics

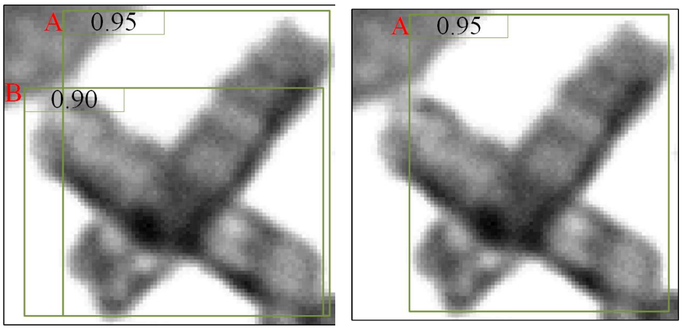
We can observe that the size of an individual chromosome typically falls within the range of [0, 5000], which is consistent with the fact that chromosomes are considered small targets in images of dimensions like 1600×1200 pixels. The default configuration of Mask R-CNN is set to output all feature layers of the FPN, including P2-6, which may not be ideal for small target chromosomes. Small targets do not necessitate the generation of overly large anchors on the high-level feature map. The anchors generated by the low-level feature tensor map sufficiently cover all chromosome targets. Therefore, we choose to utilize the P2 and P3 layers as the output layers. These outputs are fed into the feature extraction network, which incorporates a larger number of anchors in the lower layers. This approach effectively combines semantic and detailed information, thereby improving object detection. Further calculation details on this configuration are provided in Section 3.3.

2.2 Uncertainty-Based Correction of Chromosome Detection Boxes

In this part, we improve the fully-connected structure of the detected branch to enhance the regression localization ability of the network. Gaussian modeling is employed to derive uncertainty, which is then used to correct the position of the detection boxes, thereby enhancing the final segmentation performance.

Mask R-CNN locates the detection box by a regression branch and identifies the category of the detection object by a classification branch, both of which usually use a fully-connected structure and share parameters. In contrast, convolutional layers are better suited for regression location.

However, there is a lack of coordination between the classification and localization tasks. Paper [29] has been observed that a fully connected structure is more adept at distinguishing complete and partial objects, as it employs different parameters for various parts of a detection frame, making it spatially sensitive and better suited for distinguishing between complete and partial targets. Conversely, the convolutional layer, with its weight-sharing property, excels in regressing the entire object. Experiments have shown that achieving better performance can be realized by using the fully connected header for classification and the convolutional head for bounding box regression, as opposed to solely relying on the fully connected header. Therefore, we have configured two convolutional layers in the regression branch and two fully-connected layers in the classification branch, as shown in Figure 2. Besides, the detection boxes of overlapping chromosomes usually overlap, and if the regression and classification branches are completely independent, the detection boxes with inaccurate localization in the non-maximum suppression (NMS) stage may be retained incorrectly. As shown in Figure 4, detection box A and detection box B are associated with different chromosome individuals, the confidence level of box A surpasses that of box B. Simultaneously, the IoU values of both boxes exceed the NMS threshold, which triggers the suppression of box B by the NMS algorithm, consequently resulting in the loss of chromosome instances.



(a) (b)

Figure 4: NMS for overlapping and intersecting chromosomes. (a) The localization status before NMS; (b) The localization status after NMS.

Thus, the overlapping prediction boxes are very likely to be suppressed as redundant boxes, resulting in missed detection. Considering that the predicted values of the same group of candidate boxes approximately obey Gaussian distribution [30], the uncertainty can be estimated by the standard deviation of the corresponding distribution. Therefore, we establish a Gaussian distribution model in the regression branch to predict the uncertainty of each detection box and then correct the localization of overlapping chromosome detection boxes. Specifically, suppose there are *K* ground-truth boxes, the *k-th* ground-truth box corresponds to M_k proposals, and the coordinates of the *k-th* ground-truth box are denoted as (x_k, y_k, w_k, h_k), where the coordinates of the m_k*-th* proposal box are denoted as $(x_{m_k}, y_{m_k}, w_{m_k}, h_{m_k})$. As is known, the regression branch does not directly output the coordinate values, but it outputs the regression loss parameters $d^*_{m_k} = [dx^*_{m_k}, dy^*_{m_k}, dw^*_{m_k}, dh^*_{m_k}]$ of the prediction box for the m_k*-th* proposal box. The coordinates of the regressed prediction box are denoted as $(x^*_{m_k}, y^*_{m_k}, w^*_{m_k}, h^*_{m_k})$. The regression parameters of the m_k*-th* proposal box corresponding to the ground-truth values are denoted as $d^G_{m_k} = [dx^G_{m_k}, dy^G_{m_k}, dw^G_{m_k}, dh^G_{m_k}]$. This parameter transformation relationship is shown as follows (1)、（2）:

$$dx^*_{m_k} = \frac{x^*_{m_k} - x_{m_k}}{w_{m_k}}, \quad dy_{m_k} = \frac{y^*_{m_k} - y_{m_k}}{h_{m_k}}, \quad dw_{m_k} = \log(\frac{w^*_{m_k}}{w_{m_k}}), \quad dh_{m_k} = \log(\frac{h^*_{m_k}}{h_{m_k}}) \tag{1}$$

$$dx^G_{m_k} = \frac{x_k - x_{m_k}}{w_{m_k}}, \quad dy^G = \frac{y_k - y_{m_k}}{h_{m_k}}, \quad dw^G = \log(\frac{w_k}{w_{m_k}}), \quad dh^G_{m_k} = \log(\frac{h_k}{h_{m_k}}) \tag{2}$$

Accordingly, a Gaussian distribution model is developed for each regression loss parameter $d^*_{m_k} = [dx^*_{m_k}, dy^*_{m_k}, dw^*_{m_k}, dh^*_{m_k}]$, which is expressed as follows （3）:

$$p(d^G_{m_k} \mid d^*_{m_k}) = N(d^G_{m_k} \mid \mu(d^*_{m_k}), \sigma(d^*_{m_k})) \tag{3}$$

where $\mu(d^*_{m_k}) = [\mu(dx^*_{m_k}), \mu(dy^*_{m_k}), \mu(dw^*_{m_k}), \mu(dh^*_{m_k})]$ denotes the mean values of the regression loss parameters for each coordinate of the m_k proposal box, respectively, as shown as follows（4）:

$$\mu(dx^*_{m_k}) = \sum_{m_k=1}^{M_k} dx^*_{m_k}, \quad \mu(dy^*_{m_k}) = \sum_{m_k=1}^{M_k} dy^*_{m_k}, \quad \mu(dw^*_{m_k}) = \sum_{m_k=1}^{M_k} dw^*_{m_k}, \quad \mu(dh^*_{m_k}) = \sum_{m_k=1}^{M_k} dh^*_{m_k} \tag{4}$$

where $\sigma(d_{m_k}) = [\sigma(dx_{m_k}), \sigma(dy_{m_k}), \sigma(dw_{m_k}), \sigma(dh_{m_k})]$ is the output of UCRD-Net's uncertainty prediction for the mk-th proposal box after logistic transformation. The loss of uncertainty branch loss L_{un} is computed as follows, $\varepsilon = 10^{-9}$ denoting the error（5）：

$$L_{\text{un}} = -\sum_k \sum_m \log(N(d^G_{m_k} \mid \mu(d^*_{m_k}), \sigma(d^*_{m_k})) + \varepsilon) \qquad (5)$$

The loss of the regression branch L_{reg} is smoothLoss, which is computed as follows（6）、（7）：

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \qquad (6)$$

$$L_{\text{reg}} = \sum_k \sum_m \text{smooth}_{L_1}(d^G_{m_k} - d^*_{m_k}) \qquad (7)$$

The loss of the classification branch L_{class} is Binary Cross Entropy (BCE), which is denoted as follows（8）、（9）：

$$p^G_{m_k} = \begin{cases} 0 & \text{positive} \\ 1 & \text{negative} \end{cases} \qquad (8)$$

$$L_{\text{class}} = -\frac{1}{N_{\text{class}}} \sum_k \sum_m [p^G_{m_k} \log(p_{m_k}) + (1 - p^G_{m_k}) \log(1 - p_{m_k} \qquad (9)$$

The total loss function of the detection correction module consists of the L_{class}, L_{reg} and L_{un}, which denote classification branch loss, the regression branch loss, and the uncertainty branch loss, respectively.

In the prediction stage, for the mk-th proposal box, the average of the standard deviation of its four coordinates is used as its uncertainty η, which is denoted as follows（10）：

$$\eta = \frac{\sigma(dx^*_{m_k}) + \sigma(dy^*_{m_k}) + \sigma(dw^*_{m_k}) + \sigma(dh^*_{m_k})}{4} \qquad (10)$$

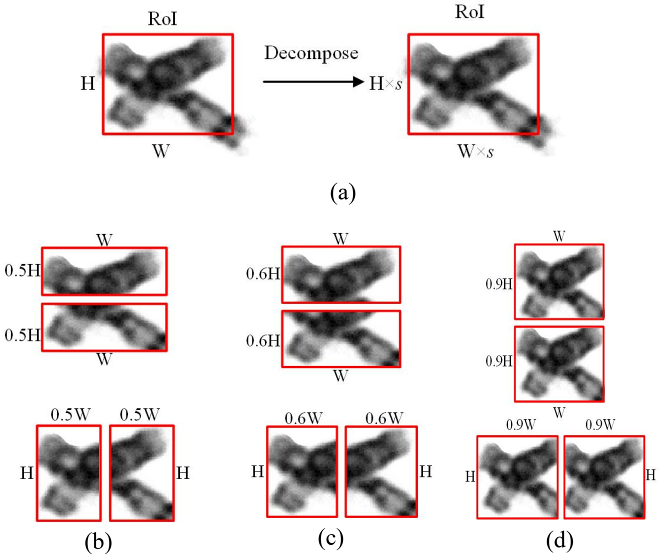
Finally, the cls_score required for the NMS sorting is corrected by uncertainty, and the corrected score is given by（11）：

$$\text{score} = \lambda \times (cls\_score) + (1 - \lambda) \times (1 - \eta) \qquad (11)$$

where $\lambda = 0.3$ is the weight parameter set by grid search in this paper

2.3 Chromosome Segmentation Based on Region Decomposition

The uncertainty-based correction of chromosome detection boxes in 2.2 addresses the issue of partially missing chromosome masks by correcting the detection box. However, overlapping chromosomes can still introduce semantic ambiguity during segmentation by influencing each other, resulting in inaccurate mask segmentation. In cases where chromosomes do not overlap, segmenting chromosomes multiple times within the RoI leverages multi-directional information to better define the target and background, thereby refining the segmentation. When chromosomes overlap, this multi-directional approach enables the network to consider the overlapping chromosomes multiple times, providing inference clues, amplifying semantic responses, and reducing segmentation mask errors. To achieve this, we introduce a decomposition factor *s*, which proportionally decomposes the RoI into upper, lower, left, and right sub-regions. The decomposition factor *s* is defined as the ratio of the length/width of the decomposed region to the length/width of the original figure, as schematically shown in Figure 5(a). For instance, when *s* equals 0.5, there is no overlap between the left and right or the upper and lower decomposed subregions, as depicted in Figure 5(b). With *s* set to 0.6, the decomposition partially overlaps the two sub-regions, preserving structural features between them, as demonstrated in Figure 5(c). However, when *s* reaches 0.9, excessive overlap reinforces the global features of overlapping chromosomes, hindering rather than facilitating their segmentation, as shown in Figure 5(d).



(a)

(b) (c) (d)

Figure 5. Original image and decomposition effect under different decomposition factors. (a) Illustration of the original decomposition; (b) s=0.5; (c)s=0.6; (d) s=0.9.

RoI Align extracts each sub-region after decomposition for 14 × 14 feature maps, respectively, which are denoted as x_{ij}^{upper} , x_{ij}^{bottom} , x_{ij}^{left} , x_{ij}^{right} and the complete features x_{ij}^{whole} of RoI are extracted at the same time. Obviously, a reasonable region decomposition helps to retain the overlapping information, but there is no guidance between regions and it cannot cover the semantic information of the chromosome subject. In order to extract features that are biased towards the overlapping regions and perceive the relationship between the local and the whole chromosome at the same time, UCRD-Net weights the features at each pixel point (i, j) on the feature maps of different sub-regions to adaptively fuse the features. The equation is as follows（12）:

$$\begin{aligned} y_{ij} = {} & \alpha_{ij}^{left} \cdot x_{ij}^{left} + \alpha_{ij}^{right} \cdot x_{ij}^{right} + \alpha_{ij}^{upper} \cdot x_{ij}^{upper} \\ & + \alpha_{ij}^{bottom} \cdot x_{ij}^{bottom} + \alpha_{ij}^{whole} \cdot x_{ij}^{whole} \end{aligned} \qquad (12)$$

where α_{ij}^{upper}, α_{ij}^{bottom}, α_{ij}^{left}, α_{ij}^{right}, α_{ij}^{whole} are the sub-region weights, respectively, which are mapped from the feature map to the weights by 1×1 convolution and then by *softmax*. The equation is as follows（13）:

$$\alpha_{ij}^{*} = \frac{e^{\lambda_{\alpha_{ij}}^{*}}}{e^{\lambda_{\alpha_{ij}}^{left}} + e^{\lambda_{\alpha_{ij}}^{right}} + e^{\lambda_{\alpha_{ij}}^{upper}} + e^{\lambda_{\alpha_{ij}}^{bottom}} + e^{\lambda_{\alpha_{ij}}^{whole}}} \qquad (13)$$

The total loss of the region decomposition module is given by（14）:

$$L_{\text{mask}}(y,p) = -\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}[y_{ij}^{mask}\log(p_{ij}) + (1-y_{ij}^{mask})\log(1-p_{ij})] \qquad (14)$$

where N and M denote the length and width of the image, and denotes p_{ij} the true class of the output mask pixel points, y^{mask} is the prediction mask feature map output by the deconvolution and convolution operations of the feature map obtained from (12).

3. Experiments

3.1 Datasets

To evaluate the performance of the proposed UCRD-Net, we utilized a clinically collected chromosome image dataset provided by a precision medical technology company. The dataset comprises 985 Giemsa-stained microscopic metaphase images, each with a resolution of 1600 × 1200 pixels. All chromosome instances within the images are meticulously annotated by experienced karyotyping experts, including both bounding boxes and pixel-level masks, ensuring high-quality and reliable ground truth for instance segmentation tasks.

To enable model training and performance evaluation, the dataset is randomly split into three subsets following a 6:2:2 ratio: 591 images are used for training, 197 for validation, and 197 for testing. The final evaluation results are reported on the testing set to ensure a fair and unbiased comparison across different methods.

3.2 Evaluations Metrics

In this paper, the effectiveness of the method is evaluated by using three different evaluation metrics of the average accuracy (mAP), AP at the IoU threshold of 0.5 (AP50) and 0.75 (AP75) [31]. The IoU metric is defined as（15）:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \qquad (15)$$

A prediction is considered to be True Positive if the IoU value is higher than the threshold (e.g., IoU > 0.5), and False Positive otherwise. Then, the area under the curve can be calculated as（16）:

$$\text{AP} = \int_0^1 p(r)dr \qquad (16)$$

where $p(r)$ is the precision-recall curve. The precision measures the accuracy of the prediction, and the recall measures whether the positive results are good or not. The precision and recall are computed as（17）、（18）:

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (17)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (18)$$

where *TP* = true positive, *FP* = false positive, and *FN* = false negative.

In practice, in a majority of current papers about instance segmentation, the interpolated curve is calculated in ranges from IoU threshold of 0.5 to 0.95 with a step size of 0.05. This method was also used in this paper. The mAP metric calculates the mean of AP across all of the IoU thresholds. We also used AP at fixed IoUs: IoU = 0.5 (AP50) and IoU = 0.75 (AP75). This means that a prediction is considered as True Positive if its IoU is >0.5 and >0.75 respectively. Thus, the higher the threshold, the greater the requirement for the accuracy of matching the prediction to the ground-truth. The standard metrics used in this study allow direct comparison of future methods with the approach used in this paper.

3.3 Implementation Details

Experiments were conducted using an Intel(R) Xeon(R) W-2175 CPU @2.50GHz processor, 64GB RAM, NVIDIA GeForce RTX 2080Ti GPU, and all networks were built based on the deep learning framework PyTorch.

For the selection of the P2 and P3 layers as the output of the FPN, we conducted the following specific analysis: suppose that the number of pixel points for each chromosome in the dataset is $\{N_1, N_2, ..., N_t\}$ and t is the number of all chromosomes in the dataset. The G-band chromosome dataset used in this paper contains 985 images with a total of 40575 labelled chromosomes. The area histogram of each single chromosome by pixel numbers is shown in Figure. 3. We can see that the area of a single chromosome is mostly in [0, 5000], and accordingly, the image area proportion of a single chromosome is [0-0.0026]. Reference to the anchor setting of Mask R-CNN, the area share of anchors on each feature layer of P2-6 is calculated as [0.0011, 0.0047, 0.0188, 0.0752, 0.3011], respectively. So only the output P2, and P3 feature layers can cover all chromosome instances. The anchors with area of $\{32^2, 64^2, 96^2, 128^2\}$ are set according to aspect ratio {1:5,1:4,1:3,1:2,1:1,2:1,3:1,4:1,5:1}, respectively. Due to the hardware limitations and image size, the batch size is set to 1. The initial learning rate is 0.02, the momentum is 0.9, and the weight decay is 0.0001, with the random gradient descent method used for network optimization.

3.4 Comparisons with Other Methods

We compare UCRD-Net with the generic instance segmentation networks, including Mask R-CNN [26], PANet [32], Cascade R-CNN [33], HTC [34], Mask Scoring R-CNN [35], as well as current instance segmentation networks designed for chromosome datasets such as AS-PANet[23], ESNet [17] and RCNet [36]. The results of the comparison experiments are presented in Table 1.

Test Accuracy. Mask R-CNN achieved a mAP of 79.4%, AP50 of 97.4%, and AP75 of 93.6%. PANet enhances multi-scale feature representation by introducing feature pyramid augmentation and a top-down path aggregation strategy. However, its performance on the chromosome segmentation task is inferior to that of Mask R-CNN. Other general-purpose instance segmentation methods, such as Cascade R-CNN, adopt a cascaded architecture that progressively refines multi-stage regression and classification. HTC further builds upon Cascade R-CNN by incorporating a parallel mask branch and introducing inter-task information flow. Compared to Mask R-CNN, these methods achieve certain performance improvements, which can be mainly attributed to their enhancements in joint optimization of detection and segmentation, multi-stage refinement mechanisms, and mask quality estimation. When Mask Scoring R-CNN introduced the Mask IoU head to predict mask quality scores, it improved the segmentation performance slightly, with the mAP increasing by 0.4%, AP50 by 0.1%, and AP75 by 0.2%. AS-PANet, a method specifically designed for chromosome datasets, demonstrated an improvement of 0.7% in mAP and 0.2% in AP50 compared to Mask R-CNN. ESNet incorporates chromosome edge information, resulting in a 2.3% improvement in mAP compared to Mask R-CNN. This performance gain may be attributed to the use of edge cues, which help the model produce more precise segmentation results. RCNet demonstrated a 3.7% increase in mAP, a 0.3% increase in AP50, and a 0.5% increase in AP75 compared to Mask R-CNN, with significant improvements in all evaluation metrics. These positive results stem from the proper adaptation of the network architecture to the chromosome dataset. In comparison, UCRD-Net, designed with the chromosome dataset in mind, showed the most significant improvement. It achieved a remarkable 4.1% increase in mAP, 0.4% in AP50, and 1% in AP75, surpassing all other approaches in all performance metrics.

Table 1 Comparison with existing methods

Methods	**mAP**	**AP50**	**AP75**
Mask R-CNN	79.4	97.4	93.6
PANet	76.2	97.1	92.1
Cascade R-CNN	80.3	97.6	93.6
HTC	80.4	97.4	93.8
Mask Scoring R-CNN	79.9	97.6	93.6
AS-PANet	80.1	97.6	93.8
ESNet	80.7	98.5	94.1
RC Net	83.1	97.7	94.1
UCRD-Net	83.5	97.8	94.6

Analysis of results: Mask Scoring R-CNN shows marginal performance gains compared to Mask R-CNN, which can be attributed to inherent challenges posed by the chromosome dataset, characterized by similar chromosome shapes and small objects. The distinctions in Mask IoU for this dataset are relatively minor, resulting in limited corrections and thus the observed marginal improvement. AS-PANet involves the addition of a counting branch to estimate the number of chromosomes in the image. However, the improvement in accuracy was not substantial, mainly due to differences between the dataset used in AS-PANet and the dataset in this paper. The dataset of AS-PANet is mainly characterized by images of chromosome clusters, which contain a small number of chromosomes (typically 2-5). In contrast, the dataset in this paper consists of images directly extracted from human cells, typically containing 46 chromosomes (including images of chromosomes with abnormal data). In this case, counting branches produce large errors and losses, making network training challenging and limiting the effectiveness of the method. RCNet, designed specifically for the chromosome dataset, achieved notable improvements. UCRD-Net outperformed other methods by optimizing feature extraction, addressing the challenges of detection and segmentation caused by overlapping instances, and introducing specific branches and strategies. This comprehensive enhancement effectively improved chromosome segmentation accuracy.

3.5 Ablation Studies

We designed ablation experiments to evaluate the effectiveness of each module in UCRD-Net for feature extraction, and the results are shown in Table 2.

Table 2. Ablation study

Methods	mAP	AP50	AP75
Mask R-CNN	79.4	97.4	93.6
Mask R-CNN + FE	79.9	97.6	93.6
Mask R-CNN + FE + UC	82.1	97.7	94.4
Mask R-CNN + FE + RD	81.9	97.7	93.9
Mask R-CNN + FE + UC+ RD	83.5	97.8	94.6

* FE: chromosomal feature extraction; UC: uncertainty-based correction of chromosome detection boxes; RD: chromosome segmentation based on region decomposition.

Feature extraction module enhances downstream segmentation: Compared with the training weights obtained on natural images only, the feature extraction module with the addition of chromosome image fine-tuning can extract more features about chromosomes, and its feature selection is more specific to chromosome shape features, so compared to the baseline (Mask R-CNN), except for the more demanding AP75 item, the performance on mAP improves the performance by 0.5% and AP50 by 0.2%.

UC improves detection performance: As illustrated in Figure 6(a), the segmentation result of Mask R-CNN showcases missed and inaccurately positioned detection boxes, leading to mask omissions. In contrast, Figure 6(b) shows the outcome after adding uncertainty correction, effectively addressing these issues. In Table 1, the addition of uncertainty correction results in a 2.7% improvement in AP relative to the baseline, a 0.3% improvement in AP50 relative to the baseline, and a 0.8% improvement in AP75 relative to the baseline, with improvements observed across all metrics. The baseline method relies on a classification score representing the probability of the presence of chromosomes in the detection box to rank the boxes. However, it lacks a mechanism to assess the reliability of the detection boxes and determine their accuracy. The introduction of uncertainty metrics reflects the reliability of the detection boxes. Sorting scores post-uncertainty correction alleviates over-suppression due to overlapping instances, resulting in more accurately localized boxes. Consequently, the masks obtained in the subsequent segmentation task are more precise. Moreover, the UC module demonstrates substantial improvements in the challenging AP75 metric compared to other modules, demonstrating that the use of uncertainty enables UCRD-Net to obtain more accurate masks.

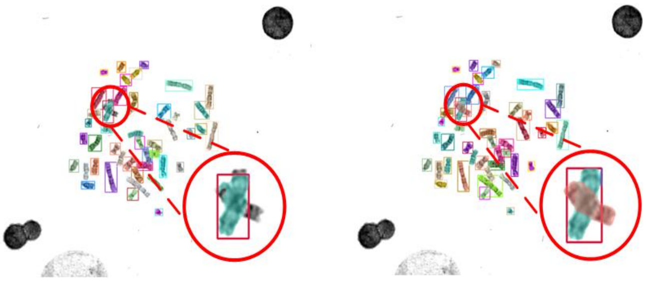
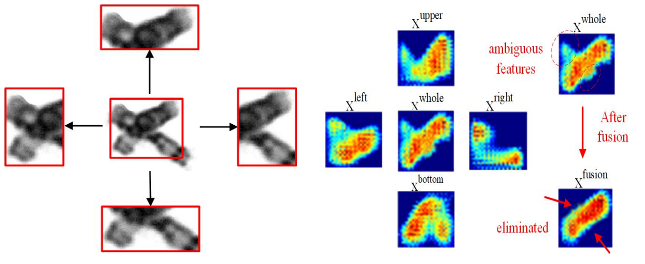

(a) (b)

Figure 6. Visualization results of the UC module. (a) Mask R-CNN segmentation results; (b) segmentation results after adding uncertainty correction.

RD mitigates the semantic ambiguity: We employ a decomposition factor (s=0.6) to break down the RoI region, the multi-region reminder network diminishes the influence of features from other chromosomes within the RoI, strengthening the features of the target chromosome for segmentation. In Table 1, it is evident that reasonable decomposition and fusion can alleviate semantic ambiguity, yielding a higher IoU of the segmentation mask and the ground-truth mask. Consequently, more accurate chromosome masks are achieved, with mAP improving by 2.5% relative to the baseline, AP50 by 0.5%, and AP75 by 0.3% relative to the baseline. RD module assists the network in distinguishing the segment corresponding to the object chromosome and alleviates the semantic ambiguity associated with overlapping chromosomes.

Figure 7 shows the result of feature map visualization, where x^{whole} is the feature before the fusion operation and x^{fusion} is the feature after the fusion operation. All other regions correspond to the parts of the original region decomposition, as shown in Figure 7(a). The semantic ambiguities circled in red above in x^{whole} reflected in x^{left} and x^{upper}, while the semantic ambiguities circled in red below reflected in x^{right} and x^{bottom}. We note that when no measures are taken, the ambiguous features caused by other chromosomes are not eliminated, resulting in redundant masks for subsequent mask segmentation. When each region is extracted and adaptively fused with the strategy, the four regions together strengthen the semantic response of the chromosome to be segmented, while the semantic response of the weaker part is weakened, making the final mask obtained more complete, as shown in Figure 7(b).



(a) (b)

Figure 7. Visualization results of the RD module. (a) schematic diagram of the original image region decomposition; (b) feature map of each region and comparison of features before and after fusion.

4. Conclusion

Chromosome segmentation serves as the cornerstone of automated karyotype analysis, and accurately segmenting overlapping or cross-over chromosomes remains a critical challenge. To address this, we propose UCRD-Net, an end-to-end chromosome instance segmentation framework designed for datasets with limited annotations and high degrees of chromosome overlap. UCRD-Net integrates two key innovations: uncertainty correction, which refines bounding box localization through confidence modeling, and region decomposition, which alleviates semantic ambiguity in overlapping regions by disentangling and adaptively fusing feature representations. To validate the efficacy of our designed end-to-end chromosome instance segmentation method, which leverages uncertainty correction and region decomposition, we conducted experiments using a clinical chromosome dataset. Comparative analysis against other standard instance segmentation methods confirms that our proposed approach surpasses them in terms of average segmentation performance. It also demonstrates superior segmentation performance at threshold values of 0.5 and 0.75. Remarkably, our algorithm accomplishes end-to-end segmentation of individual chromosomes within the original image, eliminating the need for preprocessing into chromosome cluster images.

Although the proposed model demonstrates superior performance compared to previous methods on the clinical chromosome dataset, it still has certain limitations. In particular, the current region decomposition strategy, while effective in reducing semantic ambiguity in moderately overlapping cases, may struggle to precisely segment highly entangled or densely overlapped chromosomes. In these challenging scenarios, the decomposed directional features may not fully capture the intricate spatial relationships, leading to incomplete boundaries or inaccurate mask assignments. In the future, we will explore more advanced mechanisms, such as integrating contour-based prior reasoning and designing adaptive or learnable decomposition strategies, to further improve segmentation accuracy under complex overlapping conditions.

Funding Statement: This work was supported in part by the National Natural Science Foundation of China under Grant 61971422, Xuzhou Science and Technology Innovation Plan - Key Special Project for Social Development under Grant KC22112.

Author Contribution: **Xinyu Fan:** Validation; formal analysis; visualization; writing – original draft. **Hui Liu:** Conceptualization; supervision; writing – review and editing. **Lian Xiong:** Investigation; formal analysis. **Hongdang**

Zheng: Conceptualization; supervision; writing – review and editing. **Lin Zhang:** Conceptualization; supervision; writing – review and editing.

Statements of Ethical Approval: Not applicable.

Declaration of Competing Interest: The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] Gartler, S.M. The chromosome number in humans: a brief history. *Nature Reviews Genetics* **2006**, 7, 655-660. https://doi.org/10.1038/nrg1917

[2] Quach, T.T., Stratton, H.J., Khanna, R., Kolattukudy, P.E., Honnorat, J., Meyer, K., & Duchemin, A. M. Intellectual disability: dendritic anomalies and emerging genetic perspectives. *Acta Neuropathologica* **2021**, 141, 139-158. https://doi.org/10.1007/s00401-020-02244-5

[3] Travessa, A.M., Dias, P., Santos, A., Custódio, S., Sousa, A., & Sousa, A.B. Upper limb phocomelia: A prenatal case of thrombocytopenia-absent radius (TAR) syndrome illustrating the importance of chromosomal microarray in limb reduction defects. *Taiwanese Journal of Obstetrics and Gynecology* **2022**, 59, 318-322. https://doi.org/10.1016/j.tjog.2020.01.024

[4] Liu, X., Zhang, W., Zhang, L., Ma, Y., Gao, Z., You, Y., & Wang, C. The Relationship between Prenatal Diagnosis Indications and Abnormal Chromosomal Karyotypes: A Retrospective Cohort of 4646 Cases in Beijing from 2012-2019. *Current Topics in Medicinal Chemistry* **2021**, 21, 1301-1306. https://doi.org/10.2174/1568026621666210714160232

[5] Chen, L., Du, J., Wang, J., Chen, S., Wang, W., Yang, W. Study on the application value of BACs-on-Beads technology combined with chromosome karyotype analysis in prenatal diagnosis. *Translational Pediatrics* **2022**, 11. 212. https://doi.org/10.21037/tp-22-16

[6] Bhalegaonkar, S.A., Munot, M.V., Anuse, A.D., & Kute, R.S. Automated Metaphase Chromosome Image Selection Techniques for Karyotyping: Current Status and Future Prospects. *Turkish Journal of Computer and Mathematics Education* **2021**, 12, 3258-3266. https://doi.org/10.1109/ICACC63692.2024.10845351

[7] Arora, T., Dhir, R., & Mahajan, M. An algorithm to straighten the bent human chromosomes. 2017 fourth international conference on image information proce ssing (iciip), Shimla, India, 21-23 December 2017;pp. 1-6. https://doi.org/10.1109/ICIIP.2017.8313772

[8] Kang, S., Han, J., Chu, Y., Lee, I., Joo, H., & Yang, S. Automated chromosomes counting systems using deep neural network. 2022 International Conference on Electronics, Information, and Communication (ICEIC), Jeju, Korea, 6-9 February 2022;pp. 1-3. https://doi.org/10.1109/ICEIC54506.2022.9748307

[9] Wang, W., Yousaf, M., Liu, D., & Sohail, A. A comparative study of the genetic deep learning image segmentation algorithms. *Symmetry* **2022**, 14, 1977. https://doi.org/10.3390/sym14101977

[10] Grisan, E., Poletti, E., & Ruggeri, A. Automatic segmentation and disentangling of chromosomes in Q-band prometaphase images. *IEEE Transactions on Information Technology in Biomedicine* **2009**, 13, 575-581. https://doi.org/10.1109/TITB.2009.2014464

[11] Minaee, S., Fotouhi, M., & Khalaj, B.H. A geometric approach to fully automatic chromosome segmentation. 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB),USA, 13 December 2014;pp. 1-6. https://doi.org/10.1109/SPMB.2014.7163174

[12] Karvelis, P., Likas, A., & Fotiadis, D.I. Identifying touching and overlapping chromosomes using the watershed transform and gradient paths. *Pattern Recognition Letters* **2010**, 31, 2474-2488. https://doi.org/10.1016/j.patrec.2010.08.002

[13] Mei, L., Yu, Y., Shen, H., Weng, Y., Liu, Y., Wang, D. Adversarial multiscale feature learning framework for overlapping chromosome segmentation. *Entropy* **2022**, 24, 522. https://doi.org/10.3390/e24040522

[14] Pijackova, K., Gotthans, T., & Gotthans, J. Deep learning pipeline for chromosome segmentation. In 2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA), Kosice, Slovakia, 21-22 April 2022;pp. 01-05. https://doi.org/10.1109/RADIOELEKTRONIKA54537.2022.9764950

[15] Xiang, X., Meng, F., Lv, N., & Yin, H. Engineering vehicles detection for warehouse surveillance system based on modified yolov4-tiny. *Neural Processing*

Letters **2023**, 55, 2743-2759. https://doi.org/10.1007/s11063-022-10982-8

[16] Nikolaou, A., & Papakostas, G.A. Exploiting deep learning for overlapping chromosome segmentation. Computer Vision and Robotics: Proceedings of CVR, Lucknow, India, May 2022; pp. 309-329. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-7892-0_24

[17] Liu, H., Li, X., Fan, X., Lu, Z., & Zhang, L. ESNet: End-to-end Chromosome Instance Segmentation Method Based on Edge Supervised Network. *IEEE Transactions on Computational Biology and Bioinformatics* **2025** 22, 1,368-374.https://doi.org/10.1109/TCBBIO.2024.3525409.

[18] Hu, R. L., Karnowski, J., Fadely, R., & Pommier, J. P. Image segmentation to distinguish between overlapping human chromosomes. *arXiv preprint* https://arxiv.org/abs/1712.07639

[19] Wang, G., Liu, H., Yi, X., Zhou, J., & Zhang, L. ARMS Net: Overlapping chromosome segmentation based on Adaptive Receptive field Multi-Scale network. *Biomedical Signal Processing and Control* **2021**, 68, 102811. https://doi.org/10.1016/j.bspc.2021.102811

[20] Wang, P., Hu, W., Zhang, J., Wen, Y., Xu, C., & Qian, D. Enhanced rotated mask r-cnn for chromosome segmentation. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),online meeting,30 October -5 November, 2021;pp. 2769-2772. https://doi.org/10.1109/EMBC46164.2021.9630695

[21] Liu, H., Wang, G., Song, S., Huang, D., & Zhang, L. RC-Net: Regression correction for end-to-end chromosome instance segmentation. *Frontiers in Genetics* **2022**, 13, 895099. https://doi.org/10.3389/fgene.2022.895099

[22] Cao, X., Lan, F., Liu, C. M., Lam, T.W., & Luo, R. ChromSeg: two-stage framework for overlapping chromosome segmentation and reconstruction. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), online meeting,16-19 December, 2020;pp. 2335-2342. https://doi.org/10.1109/BIBM49941.2020.9313458

[23] Lin, C., Zhao, G., Yin, A., Ding, B., Guo, L., & Chen, H. AS-PANet: A chromosome instance segmentation method based on improved path aggregation network architecture. *Journal. of Image Graphics* **2020**, 25, 2271-2280. https://doi.org/10.11834/jig.200236

[24] Lin, C., Zhao, G., Yang, Z., Yin, A., Wang, X., Guo, L. Cir-net: Automatic classification of human chromosome based on inception-resnet architecture. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2020**, 19, 1285-1293. https://doi.org/10.1109/TCBB.2020.3003445

[25] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition NV, USA, 27 June-30 June 2016;pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[26] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask r-cnn. Proceedings of the IEEE international conference on computer vision, Venice, Italy,22 Octpber-29 Octpber 2017;pp. 2961-2969. https://doi.org/10.1109/ICCV.2017.322

[27] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, Miami, USA, 22 June-24 June 2009;pp. 248-255. https://doi.org/10.1109/CVPR.2009.5206848

[28] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition,Venice, Italy,22-29 Octpber 2017;pp. 2117-2125. https://doi.org/10.1109/CVPR.2017.106

[29] Wu, Y., Chen, Y., Yuan, L., Liu, Z., Wang, L., Li, H., & Fu, Y. Rethinking classification and localization for object detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,Seattle, USA,13 June-16 June 2020; pp. 10186-10195.https://doi.org/10.1109/CVPR42600.2020.01020.

[30] Levi, D., Gispan, L., Giladi, N., & Fetaya, E. Evaluating and calibrating uncertainty prediction in regression tasks. *Sensors* **2022**, 22, 5540. https://doi.org/10.3390/s22155540

[31] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. Microsoft coco: Common objects in

context. Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, 6 September-12 September 2014; pp. 740-755. Springer International Publishing.https://doi.org/10.1007/978-3-319-10602-1_48

[32] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. Path aggregation network for instance segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, USA, 18 June-23 June 2018;pp. 8759-8768. https://doi.org/10.1109/CVPR.2018.00913

[33] Cai, Z., & Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. Proceedings of the IEEE conference on computer vision and pattern recognition,Salt Lake City, USA, 18 June-23 June 2018;pp. 6154-6162. https://doi.org/10.1109/CVPR.2018.00644

[34] Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., & Lin, D. Hybrid task cascade for instance segmentation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,Long Beach, USA, 15 June-20 June 2019;pp. 4974-4983. https://doi.org/10.1109/CVPR.2019.00511

[35] Huang, Z., Huang, L., Gong, Y., Huang, C., & Wang, X. Mask scoring r-cnn. IProceedings of the IEEE/CVF conference on computer vision and pattern recognition,Long Beach, USA, 15 June-20 June 2019;pp. 6409-6418. https://doi.org/10.1109/CVPR.2019.00657

[36] Liu, H., Wang, G., Song, S., Huang, D., & Zhang, L. RC-Net: Regression correction for end-to-end chromosome instance segmentation. *Frontiers in Genetics* **2019**, 13, 895099. https://doi.org/10.3389/fgene.2022.