Biomedical Signal Processing and Control 68 (2021) 102811

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

# ARMS Net: Overlapping chromosome segmentation based on Adaptive Receptive field Multi-Scale network

Guangjie Wang [b], Hui Liu [a,b], Xianpeng Yi [b], Jinjun Zhou [b], Lin Zhang [a,b,*]

[a] Engineering Research Center of Intelligent Control for Underground Space, Ministry of Education, China University of Mining and Technology, Xuzhou 221116, China
[b] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

ARTICLE INFO

Communicated by S. Sarkar

*Keywords:*
Karyotype analysis
Chromosome segmentation
Adaptive receptive field
Multi-scale

ABSTRACT

Karyotype analysis has become the crux for diagnosis of genetic diseases, which requires automatic precise identification of quantitative and structural abnormalities. Thus, segmentation of chromosomes from microscope captured photos by image processing technique can help promote the automatic identification of chromosomes. However, chromosomes are often curled and overlapped together randomly in different images, which makes overlapping chromosome segmentation the hot topic in karyotype analysis. Herein an Adaptive Receptive field Multi-Scale network (ARMS Net) based on UNet architecture is proposed. The number of pooling operations is optimized to balance the requirements of deep semantic information extraction and high precision segmentation. The adaptive multi-scale feature extraction module is designed to replace the standard convolution at the bottom of UNet, such that the receptive fields can adaptively match the size of feature map. Besides, an adaptive smooth weighted cross entropy loss function is defined to resolve category imbalance issue. Experimental results show that the Intersection of Union (IoU) score of ARMS Net segmented overlapping area is 99.45%, which is 3.2% higher than that achieved by UNet (96.38%), and 10.1% higher than that achieved by CE-Net (90.35%). In a word, ARMS Net is expected to be used as the backbone network for chromosome instance segmentation in its end-to-end identification.

## 1. Introduction

### 1.1. Motivation

There are 23 pairs of chromosomes in healthy human cells, including 22 pairs of autosomes and 1 pair of sex chromosomes [1]. Chromosomes are carriers of genes, and chromosomal abnormalities account for more than 50% of spontaneous abortion, stillbirth and premature death [2]. They are also important causes of many congenital diseases. Chromosomal abnormalities include quantitative and structural abnormalities, which may occur on any chromosome [3]. Thus, identification of chromosomal abnormalities has become the key to the diagnosis, especially early diagnosis of genetic diseases. To be specific, it has become an important auxiliary method for prenatal diagnosis and genetic disease screening [4].

Chromosomes show a clear and stable morphology in the metaphase of mitosis [5], which has become the main research object of chromosome karyotype analysis. Dyed by banding technique chromosome instances can be segmented and identified by karyotyping. Common banding techniques include G-band, Q-band, R-band, C-band techniques and Fluorescence in Situ Hybridization (FISH) [6]. FISH came out in the late 1970s, combing fluorescent probes with DNA in chromosomes to better display the structure of the stained part of the target DNA under a fluorescence microscope [7]. Thus, Chromosomes be analyzed, compared, classified and numbered by extracting information such as length of chromosomes, centromeric position, the ratio of long and short arms, and whether there are any satellites [8] identify chromosomal abnormalities such as deletion, addition and mutation [9]. However, as a flexible substance [10], even chromosomes with the same number show different gestures in the nucleus due to different bending. Besides, clusters occur due to touch and overlap [11], where the analysis needs to be done manually. Thus, most karyotype analyses are achieved by manual segmentation, which is lengthy and repetitive. For example, Monika et al. used crowdsourcing to distribute data sets to various crowdsourcing platforms [12] for manual segmentation and then aggregated for classification and anomaly identification [13]. Therefore,

* Corresponding author at: Engineering Research Center of Intelligent Control for Underground Space, Ministry of Education, China University of Mining and Technology, Xuzhou 221116, China.
*E-mail address:* lin.zhang@cumt.edu.cn (L. Zhang).

https://doi.org/10.1016/j.bspc.2021.102811
Received 19 January 2021; Received in revised form 11 May 2021; Accepted 23 May 2021
Available online 31 May 2021

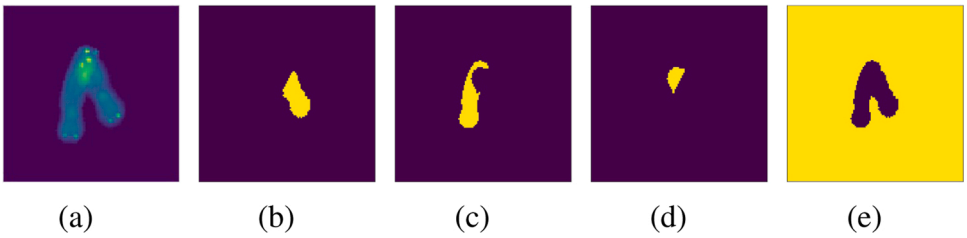


**Fig. 1.** Overlapping chromosomes image and category label. (a) overlapping chromosomes image synthesized by chromosomes $\alpha$ and $\beta$, (b) non-overlapping regions label of chromosome $\alpha$, (c) non-overlapping regions label of chromosome $\beta$, (d) overlapping regions label of chromosome $\alpha$ and $\beta$, (e) background label.

automatic segmentation of chromosomes especially touching and overlapping chromosomes has become an urgent issue in karyotype [11].

### 1.2. Related work and problems

Most traditional automatic segmentation methods are based on geometric morphology. For example, Wacharapong et al. first extracted the contour of overlapping chromosomes [14], then calculated their curvature based on the Otsu threshold processing method [15] to locate the intersection point (concave point) and tangent point, and finally implemented semi-automatic segmentation of overlapping chromosomes through Voronoi diagram [16] and Delaunay triangulation [17–19]. Somasundaram et al. first separated a single chromosome by the Multi-Objective Geodesic Contour method, then, the cut points on the image identified by curvature function to draw hypothesis lines to segment overlapping regions [20]. Yilmaz et al. first proposed a thresholding and watershed segmentation method [21] to separate non-overlapping single chromosomes and chromosome clusters, then calculated the tangency points of the chromosome clusters through the curvature function, and finally divided the overlapping chromosomes through the optimal geodesic path between tangency points [22]. These methods mainly determine the concave points of the overlapping part for segmentation, but the effective concave points may be misjudged. Thus, it remains unsatisfactory.

With the development of deep learning, neural networks are applied to transform image segmentation into pixel-level classification, which effectively improves the segmentation accuracy. For example, UNet adopts skip connections between encoder and decoder, effectively fuses deep feature map with rich semantic information and shallow feature map with rich, detailed information, and less loss of features. It has been widely used in medical image segmentation due to its ideal segmentation performance under a smaller scale of training data [23]. As for the chromosome segmentation tasks, Hu et al. constructed the UNet with two-layer pooling to segment overlapping chromosomes with less computation and storage costs [24]. The segmentation accuracy and IoU of overlapping regions are 99.22 and 94.70, respectively. Thus, the segmentation accuracy is high, but the IoU score still needs to be improved. Saleh et al. believed that the increase of pooling and convolution operation in the network was conducive to the extraction of more input feature information. Thus, they built three-layer pooling in UNet to segment overlapping chromosomes, and the segmentation accuracy and IoU were improved slightly [25]. It can be seen that Hu and Saleh et al. set the number of pooling layers from different perspectives, but neither of them elaborated on the principles. As is known, a deeper network can better extract features, but the resulting feature map may show lower resolution [26]. Thus, feature extraction and feature map resolution should be considered harmoniously. Besides, skip connection is adopted to reduce the loss of features. It may lead to semantic gaps, which degrade the segmentation performance [27]. Thus, it is also necessary to improve the connection path.

Much work shifts attention to the feature map at the bottom of the network and improves network performance by extracting and fusing the multi-scale features [28]. Ruan et al. proposed a Pyramid Pooling Module (PPM), which realized multi-scale fusion of shallow and deep information through pooling operations. They have different pooling sizes and steps in different branches but the same pooling size and steps in the same branch [29]. Gu et al. introduced Dense Atrous Convolution (DAC) module and Residual Multi-kernel Pooling (RMP) at the bottom of the UNet, extracted deeper multi-scale features while retaining more spatial information, and constructed an end-to-end Context Encoder Network (CE-Net) [26]. It can be seen that most existing methods extract multi-scale features by designing multiple fixed-size receptive fields to improve the segmentation performance. However, regarding images with different input sizes, the receptive fields cannot match the target sizes. Direct adjustment or cropping of images may cause the loss of detailed information. Thus, it is necessary to construct an adaptive multi-scale feature extractor.

Besides, in the task of pixel-level classification, cross-entropy loss [30] or focal loss [31] are always adopted. The imbalance issue frequently occurs [32], which always leads to convergence difficulty. Thus, it is intuitive to increase the penalty of fewer samples when calculating loss. In this way, the convergence is improved to some extent. However, the weights of penalties in functions are usually determined by the grid search method, which is complicated. Furthermore, the universality of different data sets is not strong. Thus, it is necessary to design a loss function that can adjust the weights adaptively according to the distribution of positive and negative samples.

### 1.3. Contributions

Herein, we propose an adaptive multi-scale feature extraction module to construct an adaptive receptive field multi-scale network (ARMS Net) to enhance the segmentation of overlapping chromosomes, which may provide more reliable input for subsequent chromosome recognition for genetic diagnosis of diseases. The multi-scale feature extraction module includes the Adaptive Multi Atrous Convolution (AMAC) module and the Adaptive Same Stride Pooling (ASSP) module, both of which adjust the receptive fields adaptively according to feature map size. AMAC assigns atrous convolutions with different atrous rates, which are obtained adaptively according to the size of the feature map. It can achieve multi-scale feature extraction and fusion. ASSP sets the pooling sizes adaptively according to the feature map size to achieve the effective acquisition of context information.

The main contributions of this paper are as follows:

(1) ARMS Net uses the AMAC and ASSP modules to adaptively modify the atrous rates and pooling sizes. With more details of the input image remained, it also adaptively sets the receptive fields according to the size of the feature map, thus effectively extracting multi-scale features.
(2) Aiming at the category imbalance issue in pixel-level classification task, an adaptive smooth weighted cross-entropy loss function is designed to optimize the training process.

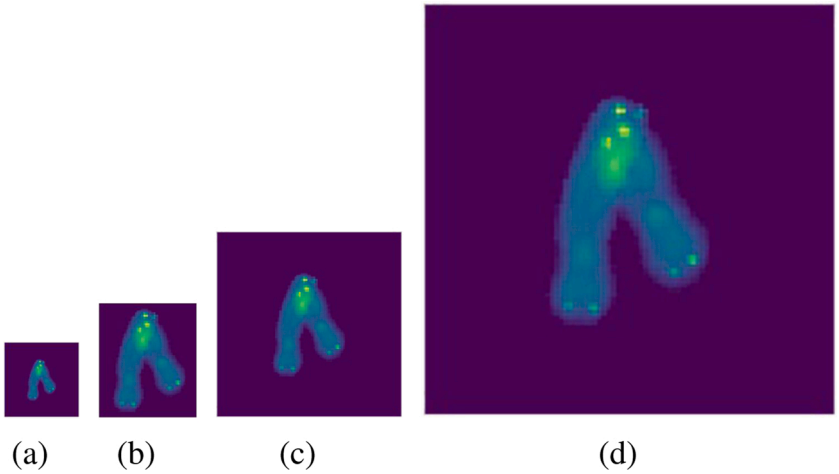


**Fig. 2.** The images that have been adjusted and cropped. (a) The zoomed out image, (b) the cropped image, (c) the input image, (d) the enlarged image.

## 2. Method

### 2.1. Problem formulation

Overlapping is common in chromosome images, and the overlapping regions are different in size and indistinct. Thus, for better generalization, an adaptive receptive field multi-scale network is constructed to achieve pixel-level classification of overlapping chromosome images. Let $X_{M\times N}$ represents the image of an overlapping chromosome, as shown in Fig. 1(a). Where $M$ and $N$ respectively represent the length and width of the input image. With the mapping model $f(\cdot)$, final output is $Y = f(X_{M\times N})$, where $Y$ is a $4 \times M \times N$ tensor, each $M \times N$ tensor is shown in Fig. 1(b)–(e).

The intermediate output of input image $X_{M\times N}$ in CNN is called feature map $I_{m\times n\times c}$. The feature map is obtained by $I_{m\times n\times c} = F(X_{M\times N})$. $F(\cdot)$ represents CNN operation, $m \times n$ is the feature map size, and $c$ is the number of feature map channels. This paper considers that most existing methods adopt fixed receptive fields, which leads to the result that receptive fields cannot match the target size. For example, a module designed with a feature map $I_{m\times n\times c}$, and its receptive field is $RF$. When the input image is modified to $X_{2M\times 2N}$, the feature map becomes $I_{2m\times 2n\times c}$, resulting in a mismatch between the receptive field ($RF$) and the feature map size ($2m \times 2n$). This problem is usually solved by adjusting and cropping the size of the input image, as shown in Fig. 2. It

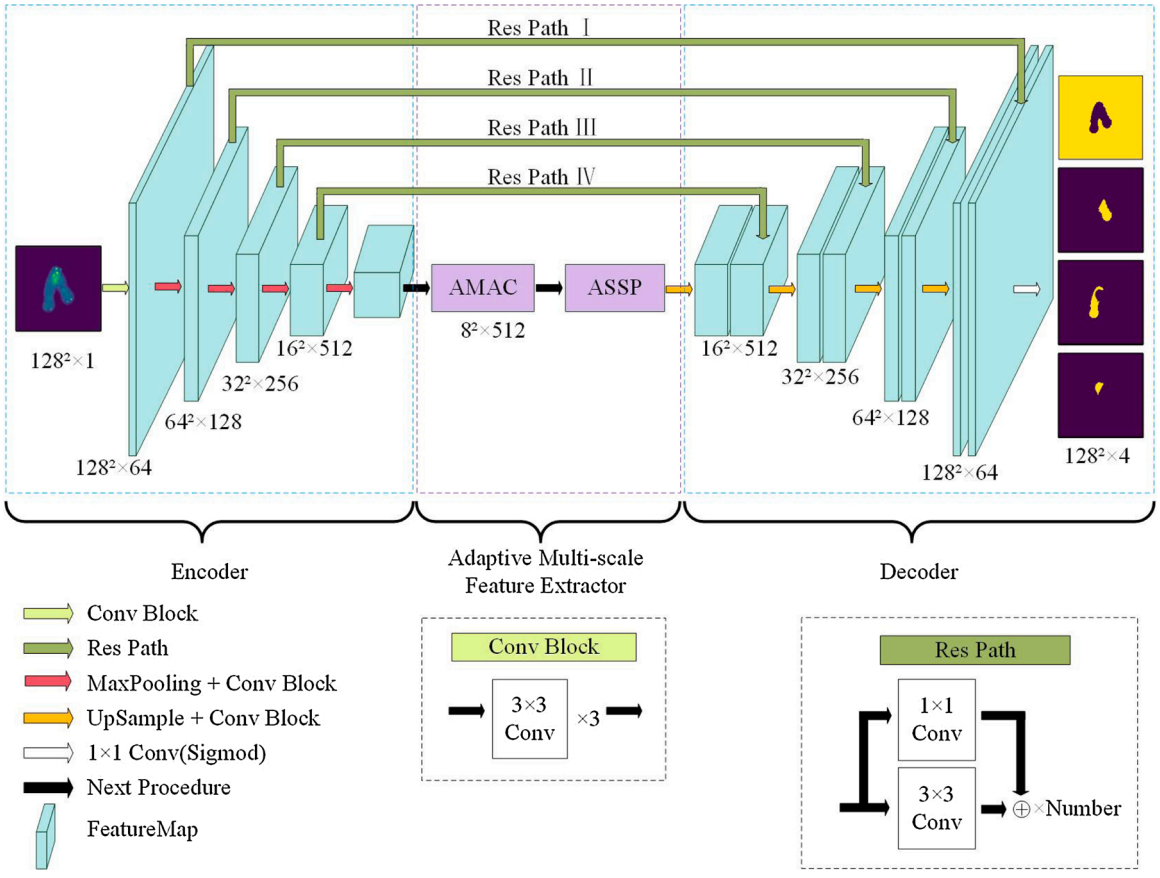


**Fig. 3.** The structure of the ARMS Net. It includes an encoder and decoder module, Res Path module and adaptive multi-scale feature extractor.

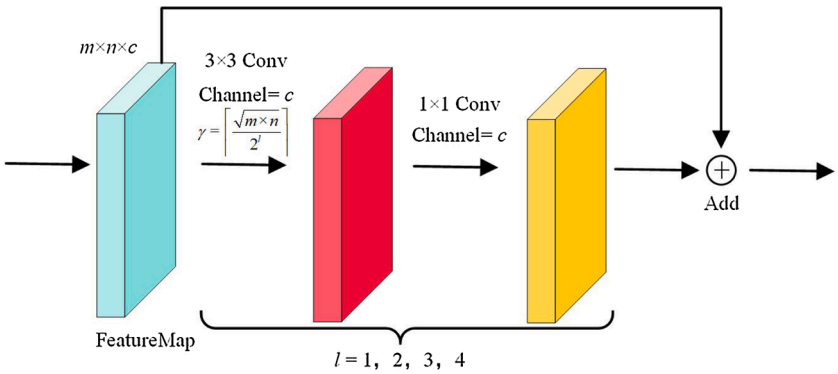


**Fig. 4.** AMAC module structure diagram. It contains four branches, which adaptively design atrous rates according to feature map sizes, then it adds each output and input. Thereby, it can extract features of different scales.

can be seen that all of the above operations will cause the image distortion and the loss of detailed information. This loss of information may have little effect on classification tasks, but it will seriously affect segmentation results.

Thus, we aim at the above problems and construct an adaptive multi-scale feature extractor so that the input images no longer need to be adjusted and cropped.

### 2.2. Experimental data

In this paper, independent chromosomes were segmented from FISH images in metaphase of real mitosis, then the overlapping chromosome data set was synthesized by rotation and overlap to evaluate the performance of ARMS Net [33]. This data set contains 13,434 overlapping chromosome images with 94 × 93 pixels. These images are first padded over round to 128 × 128 to meet the input requirements, and each image has a corresponding real label image, as shown in Fig. 1, (a) is a composite image made by overlapping chromosomes $\alpha$ and $\beta$, (b)–(e) are corresponding label images, (b) and (c) correspond to the non-overlapping regions of chromosomes $\alpha$ and $\beta$, respectively. (d) corresponds to the overlap regions, (e) corresponds to the background regions.

Taking the segmentation task of two overlapping chromosomes $\alpha$ and $\beta$ as an example, the target segmentation region includes 4 parts: the non-overlapping regions of chromosome $\alpha$ ($\alpha - \alpha \cap \beta$), the non-overlapping regions of chromosome $\beta$ ($\beta - \alpha \cap \beta$), the overlapping regions of $\alpha$ and $\beta$ ($\alpha \cap \beta$) and the background regions ($\overline{\alpha \cup \beta}$).

## 3. Analysis of network architecture

Aiming at the high-precision segmentation of different target region sizes, we construct an adaptive receptive field multi-scale network (ARMS Net). The structure of ARMS Net includes encoder and decoder module, Res Path module and adaptive multi-scale feature extractor. More specifically, the number of network pooling layers is 4 to balance the requirements of deep semantic information extraction and high precision detailed segmentation; Res Path module is used to replace original skip connection, which makes full use of the context information while extracting spatial features; Adaptive multi-scale feature extractor includes AMAC and ASSP, which design receptive fields according to the feature map size to extract and fuse the multi-scale features adaptatively, as shown in Fig. 3.

### 3.1. Choice of pooling layers

In semantic segmentation tasks, multi-layer pooling is an effective way to achieve deep convolution [34], but it will cause resolution to decrease while increasing receptive fields [35], which may lead to the loss of semantic information. However, this does not mean that we should abandon the deep semantic information and pursue the feature map resolution because the deep semantic information is also helpful for the network to achieve the segmentation of complex images. Thus, it is necessary to balance the requirements of deep semantic feature extraction and high precision detailed segmentation.

It has been shown that the number of pooling layers can be adjusted according to the size and complexity of input images. For example, GoogleNet [36], DenseNet [37], and DeepLab [38] all use five-layer pooling to achieve semantic segmentation of 224 × 224 size natural images; The MADCNN decrease the number of pooling layers to 3 based on DenseNet for 96 × 96 size brain nerve images [39]; UNet is only constructed by four-layer pooling for 572 × 572 size Drosophila ventral nerve cord (VNC) images [23]. Hu et al. and Saleh et al. set the number of pooling layers of UNet to 2 and 3 for the 88 × 88 size chromosome image. It can be seen that the number of network pooling layers cannot be derived from a rigorous formula. Compared with natural images, medical images have relatively single features and insufficient semantic information [40]. Thus, the selection of the number of pooling layers is related to not only the image size but also its feature complexity.

In this paper, the size of the overlapping chromosome image data set is 128 × 128. By comparing the segmentation performance of ARMS Net with different pooling layers, the number of pooling layers is finally determined to be four.

### 3.2. Adaptive multi-scale feature extraction module

In the overlapping chromosome segmentation task, the shape and size of the target regions are completely different. The adaptive multi-scale feature extraction module can adaptively design receptive fields according to the size of the feature map to extract and fuse multi-scale features, thereby automatically detecting target regions of different shape and size. This module includes AMAC and ASSP.

#### 3.2.1. Adaptive Multi Atrous Convolution

In the semantic segmentation task, deep semantic information is usually extracted through multi-layer pooling, but it will reduce feature map resolution, which is not conducive to detailed segmentation. To solve this problem, atrous convolution inserts holes between values of standard convolution kernel, which expands receptive field on the premise of keeping single convolution computation unchanged [41]. Thus, inserting with different numbers of holes will form convolution kernels with different atrous rates to obtain different receptive fields.

The DAC module in CE-Net is designed for a feature map with 14 × 14 pixels. By combining different atrous convolutions with different atrous rates, parallel branches of 3, 7, 9 and 19 receptive fields are formed respectively, then added to fuse multi-scale target spatial features [26]. Overlapping regions of chromosomes discussed in this paper are of different shapes and most are small. Thus, it is expected that

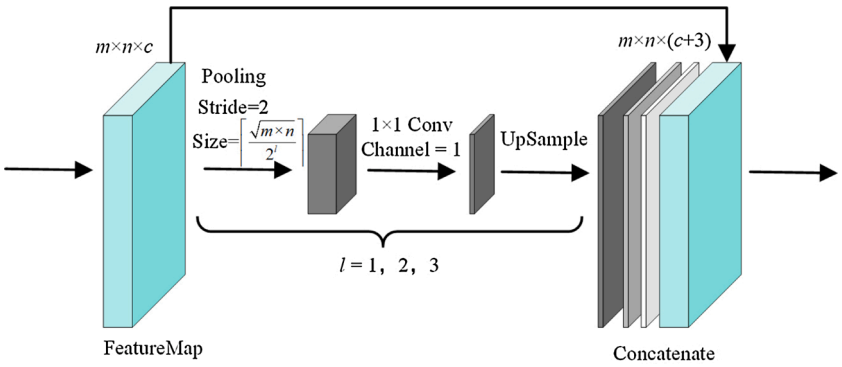


**Fig. 5.** ASSP module structure diagram. It contains three branches. The output of each branch is dimensionally reduced by 1 × 1 convolution, and then it goes through an up-sampling. Finally, it is stacked with an input feature map to achieve the fusion of multi-scale features.

the segmentation network will have a stronger performance on small-scale receptive fields. Considering that fixed atrous rates will cause receptive fields may not match the target regions when the size of the input image is too large or too small. However, if the size of the input image is fixed, much detailed information will be lost. Thus, this paper draws on the idea of DAC and constructs the AMAC module, as shown in Fig. 4. Specifically, the atrous rate of each branch is set adaptively according to feature map size to obtain a suitable receptive field and realize the extraction of adaptive multi-scale target features. More specifically, the atrous rate $\gamma = \left\lceil \frac{\sqrt{m \times n}}{2^l} \right\rceil$, where $m \times n$ indicates feature map size, $c$ indicates input channels of the feature map, Channel indicates output channels of feature map, $l$ indicates the $l$-th branch in AMAC, $\lceil \cdot \rceil$ indicates round-up operation, $\gamma$ indicates that $\gamma$-1 holes are inserted between convolution kernel values. For example, when the size of the feature map at the bottom of ARMS Net is 8 × 8 and the AMAC module with four branches is used, atrous rate $\gamma$ of each branch is set to 4, 2, 1, and 0.5. Considering that the fourth $\gamma$ is less than 1, which does not meet the operation rules, and in order to distinguish from the existing atrous rates, replace it with 3. For each branch with $\gamma$>1, the output is linearly corrected by 1 × 1 convolution. Finally, each branch's output is added with the input to achieve the fusion of different scales feature.

#### 3.2.2. Adaptive Same Stride Pooling

When setting the pooling size and step in different branches of the PPM module, the pooling step in each branch is usually not the common factor of feature map size, leading to semantic information loss in pooling and up-sampling operation. To solve this problem, we fix the pooling step to 2 in each branch of the ASSP module, ensuring that the pooling step is a common factor of all possible feature map sizes. What's more, we adaptively set multi-scale pooling sizes for feature map of different size to solve the problem that receptive fields does not match the target size after the input image size changed, which is conducive to the effective extraction of multi-scale feature, as shown in Fig. 5. The pooling size of each branch is set to $\gamma = \left\lceil \frac{\sqrt{m \times n}}{2^l} \right\rceil$, where $m \times n$ indicates feature map size, $c$ indicates input channels of the feature map, Channel indicates output channels of the feature map, $l$ indicates the $l$-th branch in ASSP, and $\lceil \cdot \rceil$ indicates round-up operation. For example, when the size of the feature map at the bottom of ARMS Net is 8 × 8, the pooling size of each branch in three-branch ASSP is 4, 2, and 1, respectively. However, considering that the pooling effect is lost when the pooling size is 1, the pooling size of the third branch is revised to 3. The output of each branch is dimensionally reduced by 1 × 1 convolution. Then, it goes through an up-sampling. Finally, it is stacked with the input feature map to achieve the fusion of multi-scale features.

**Table 1**
Res Path parameter configuration table in ARMS Net. Channels indicate the output channels of the Residual Block, and Number indicates the number of Residual Block.

| Name | Residual Block | Channels | Number |
|---|---|---|---|
| Res Path I | 3 × 3 Conv<br>1 × 1 Conv | 32 | 4 |
| Res Path II | 3 × 3 Conv<br>1 × 1 Conv | 64 | 3 |
| Res Path III | 3 × 3 Conv<br>1 × 1 Conv | 128 | 2 |
| Res Path IV | 3 × 3 Conv<br>1 × 1 Conv | 256 | 1 |

### 3.3. Res Path

Ibtehaz et al. added a different number of residual blocks to the simple skip connection and believed that this method could alleviate semantic gap through additional nonlinear transformations, and called it Res Path [27].

In order to make full use of the spatial information lost in the encoder module to compensate for the semantic gap between the encoder and decoder, we use four Res Path modules to replace the original four skip connections. The four modules are respectively denoted as Res Path I, II, III and IV, as shown in Fig. 3. Considering that there are more semantic gaps between the encoder and decoder in the shallower connection, the Res Path I module has the largest number of residual blocks. The configuration parameters of each connection path are shown in Table 1.

### 3.4. Loss function

The segmentation target regions of overlapping chromosomes $\alpha$ and $\beta$ can be divided into four parts, and the segmentation of each target region is realized by pixel-level classifications, as shown in Fig. 1. However, there is a serious category imbalance issue. To solve this problem, the weighted cross-entropy loss function is usually used to improve, as shown in Eq. (1).

$$L = -\frac{1}{4MN} \sum_{t=1}^{4} \sum_{i=1}^{M} \sum_{j=1}^{N} w \cdot y_{ij}^{t} log\left(p_{ij}^{t}\right) + \left(1 - y_{ij}^{t}\right) log\left(1 - p_{ij}^{t}\right) \quad (1)$$

Specifically, $y_{ij}^{t}$ indicates the label of the pixel $(i, j)$ in the $t$-th category, $p_{ij}^{t}$ indicates the probability of the pixel $(i, j)$ in the $t$-th category, and $w$ indicates the weight of the loss function.

However, in this issue of pixel-level classification, the number of pixels in each target region is different. That is to say, there are differences in the category imbalance between labels. Thus, methods using uniform weights are not suitable here. Considering that multiples of positive and negative sample numbers have the most direct relationship

with this weight. Furthermore, in order to prevent gradient explosion [42] caused by excessive multiples, we draw on the idea of smooth L1 loss function [43], the multiples of positive and negative samples in the four target regions of overlapping chromosomes are smoothed respectively, and are finally used as the weight of the loss function, as shown in Eq. (2).

$$w = \left(1 + \left|log_{1000}\frac{Neg^t}{Pos^t}\right|\right)^{-1^{\left(1-y_{ij}^t\right)}} \tag{2}$$

Specifically, $Pos^t$ indicates the number of pixels labeled as a chromosome region in the $t$-th category, $Neg^t$ indicates the number of pixels labeled as a background region in the $t$-th category. This loss function is suitable for multi-classification and imbalanced category scenarios and can effectively avoid gradient explosion when the difference between positive and negative samples is large and show stronger robustness.

## 4. Experiments and results

### 4.1. Experimental setting

**Experimental configuration**: The computer processor is Intel(R) Xeon(R) W-2175 CPU @2.50 GHz, 64 GB running memory, NVIDIA GeForce RTX 2080Ti GPU, Keras framework.

**Network training**: In this paper, a 5-fold cross-validation method is used to conduct experiments [44]. The proportion of the test set to all overlapping chromosome images is 0.2, and the remaining are divided into 5 parts, 4 parts are used as the training set and 1 part is used as the validation set in turn. The test set keeps separate from the training and validation set. Thus, each network needs to be trained five times.

ARMS Net uniformly uses the Adam optimizer to minimize the loss function [45], it can adaptively modify the learning rate according to the gradient descent situation, where the initial learning rate is set to the default value of 0.001. Moreover, the Batch Size is obtained by grid search. By comparing the segmentation results between the batch size of 8, 16, 32 and 64, 32 was finally selected as the batch size, as shown in Table 4. The pre-trained model was used to speed up the convergence. The maximum epoch is set to 300 with the early stopping method [46] adopted to avoid overfitting. Thus, the training converges after 30 consecutive epochs with validation loss keep stable.

### 4.2. Evaluation metrics

We equate the segmentation tasks of overlapping chromosomes as pixel-level classification tasks, which are achieved by classifying each pixel in the image into four categories.

(1) In overlapping chromosome segmentation tasks, the target segmentation regions are usually only a small part of the whole. That is to say, the segmentation accuracy of the background regions is usually high, which is easy to cause the illusion of excessive performance. Thus, common classification evaluation metrics such as precision and recall cannot fully evaluate the network performance [27]. Thus, we calculate the intersection of union (IoU) of chromosome overlapping regions ($\alpha \cap \beta$), chromosome $\alpha$ and chromosome $\beta$ as the Metrics of network performance [47], as shown in Eq. (3).

$$\text{IoU} = \frac{P \cap G}{P \cup G} \tag{3}$$

Where $P$ indicates the segmentation result, $G$ indicates the corresponding real label, and the closer the IoU is to 1, the better the segmentation performance is.

What's more, the overlapping region of chromosomes is a difficult-to-segment region. Taking the IoU of this region as the evaluation metrics can better reflect the network's segmentation performance. However, considering that the overlapping region of chromosomes is small, even only a few pixels, a small segmentation error will result in a significant reduction of evaluation metrics. Therefore, the IoU of the overlapping chromosome region and the complete chromosome is used as the evaluation metrics.

(2) To compare with the segmentation performance of existing models, we also calculate the segmentation accuracy, as shown in Eq. (4).

$$\text{Accuracy} = \frac{TP + TN}{ALL} \tag{4}$$

Where $TP$ indicates the number of pixels that are correctly predicted as chromosomes, $TN$ indicates the number of pixels that are correctly predicted as background, and $ALL$ indicates the number of pixels in the image.

**Table 2**
Network configuration table. Different networks are designed to have the same number of pooling layers.

| | Method | Feature extraction module | | Path | Loss function |
|---|---|---|---|---|---|
| 1 | UNet | Conv[a] | Conv | Skip Path | BCE[b] |
| 2 | UNet (R) | Conv | Conv | Res Path | BCE |
| 3 | UNet (PR) | Conv | PPM | Res Path | BCE |
| 4 | UNet (AR) | Conv | ASSP | Res Path | BCE |
| 5 | UNet (DAR) | DAC | ASSP | Res Path | BCE |
| 6 | UNet (AAR) | AMAC | ASSP | Res Path | BCE |
| 7 | ARMS Net | AMAC | ASSP | Res Path | AS_wBCE[c] |

[a] Conv indicates standard convolution.
[b] BCE indicates binary cross-entropy loss function.
[c] AS_wBCE indicates adaptive smooth weighted binary cross-entropy loss function.

**Table 3**
Comparison table of 3 networks under different evaluation metrics. Each network has a different number of pooling layers.

| Method | IoU (%)[b] | Chrom $\alpha$ IoU (%)[c] | Chrom $\beta$ IoU (%) | Accuracy (%)[d] | Time (ms)[e] |
|---|---|---|---|---|---|
| UNet (AAR) 3L[a] | 99.12 | 99.59 | 99.79 | 99.9923 | 3.9 |
| UNet (AAR) 4L | **99.22** | **99.70** | **99.85** | **99.9943** | 4.6 |
| UNet (AAR) 5L | 98.60 | 99.64 | 99.82 | 99.9932 | 4.8 |

Bold values denote the best performance.
[a] 3L indicates three pooling layers.
[b] IoU indicates the IoU score of the overlapping regions of chromosomes.
[c] Chrom $\alpha/\beta$ IoU indicates IoU scores of two complete chromosomes in overlapping chromosomes.
[d] Accuracy indicates the average accuracy of two complete chromosomes.
[e] Time indicates the prediction time of a single image.

### 4.3. Experimental analysis

A total of 6 networks (2–7) are designed to verify the effectiveness of each module, as shown in Table 2. Specifically, PPM and ASSP are designed with the same pooling sizes.

In order to verify the effectiveness of the AMAC module, the ASSP module and the Res Path module on the overlapping chromosome segmentation, we first configure the UNet (AAR) that contains the above mentioned three modules with three different pooling layers for comparison. The scores of the three networks in different metrics are shown in Table 3. Specifically, UNet (AAR) 3L has the most detailed feature map resolution, but the ability of deep semantic information extraction is insufficient. Thus, the network has a high accuracy for simple image segmentation, but the segmentation performance is poor in complex images such as those with small overlapping areas or serious overlapping situations. In contrast, UNet (AAR) 5L has a stronger semantic

**Table 4**
The weighted loss function compares the statistics of the experimental results.

| | Mean (var) of Training IoU (%)[a] | Mean (var) of Testing IoU (%) |
|---|---|---|
| BCE_bs8[b] | 99.43 (0.25) | 98.37 (0.41) |
| BCE_bs16 | 99.48 (0.28) | 98.98 (0.39) |
| BCE_bs32 | 99.63 (0.18) | 99.09 (0.29) |
| BCE_bs64 | 99.56 (0.21) | 99.01 (0.31) |
| wBCE (1.1)[c]_bs32 | 99.53 (0.25) | 99.15 (0.39) |
| wBCE (1.2)_bs32 | 99.57 (0.23) | 99.12 (0.38) |
| wBCE (1.3)_bs32 | 99.57 (0.26) | 99.10 (0.32) |
| wBCE (1.4)_bs32 | 99.52 (0.24) | 99.02 (0.31) |
| wBCE (1.5)_bs32 | 99.45 (0.25) | 99.04 (0.50) |
| wBCE (2.0)_bs32 | 99.55 (0.18) | 99.01 (0.44) |
| wBCE (3.0)_bs32 | 99.44 (0.23) | 98.73 (0.55) |
| AS_wBCE_bs32 | 99.33 (0.24) | **99.29 (0.27)** |

Bold value denotes the highest result.
[a] Mean (var) of Training/Testing IoU indicate the mean and variance of the IoU scores in the overlapping regions of the training set and the test set, respectively.
[b] bs8 indicate the value of batch size is 8.
[c] wBCE (1.1) indicates the weighted cross-entropy loss function shown in Eq. (1), 1.1 indicates the weight $w$ in Eq. (2) is 1.1.

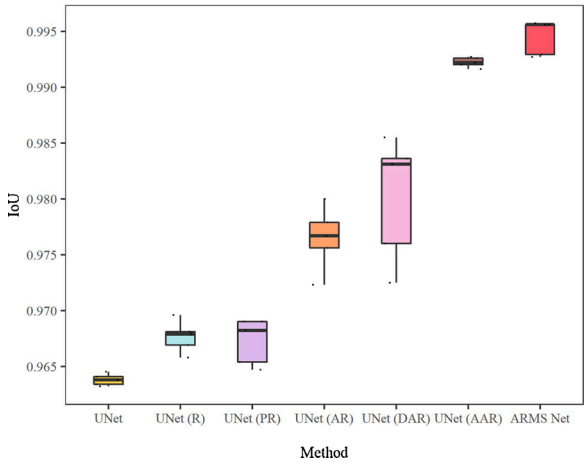


**Fig. 6.** Box plots of IoU scores in the overlapping areas of 7 networks. Each network is experimented by the 5-fold cross-validation method.

information extraction ability and can roughly segment complex images. However, due to the low resolution of the feature map, it cannot achieve detailed segmentation. UNet (AAR) 4L combines the advantages of both, it not only has a detailed feature map resolution but also can extract deep semantic information. Thus, its IoU score of the target overlapping regions on the test data set is the highest. Similarly, the network still achieves the highest scores of IoU and Accuracy on complete chromosomes. Therefore, in the following experiments, we construct networks with four-layer pooling to comprehensively balance the requirements of deep semantic information extraction and high precision detailed segmentation to improve the performance of overlapping chromosome segmentation networks.

In order to verify the effectiveness of the adaptive smooth weighted cross-entropy loss function, we carry out comparative experiments based on the UNet (AAR) 4L with different loss functions. The weights and results are shown in Table 4. It can be seen that the same weight shared by the four labels in the weighted cross-entropy loss function is not suitable for all labels, which is caused by the massive difference in the number of positive and negative samples in different labels. However, the adaptive smooth weighted cross-entropy loss function can adaptively calculate each cross-entropy weight in the overall loss according to the category imbalance of the four labels, then smooths it as

**Table 5**
Comparison table of 9 networks under different evaluation metrics.

| Input size | Method | IoU (%) | Chrom $\alpha$ IoU (%) | Chrom $\beta$ IoU (%) | Accuracy (%) | Time (ms) |
|---|---|---|---|---|---|---|
| 128 × 128 | UNet | 96.38 | 98.38 | 99.16 | 99.9690 | 3.8 |
| 128 × 128 | UNet (R) | 96.77 | 98.85 | 99.41 | 99.9780 | 4.0 |
| 128 × 128 | UNet (PR) | 96.73 | 98.84 | 99.40 | 99.9778 | 4.4 |
| 128 × 128 | UNet (AR) | 98.29 | 99.12 | 99.55 | 99.9832 | 4.4 |
| 128 × 128 | UNet (DAR) | 98.36 | 99.27 | 99.62 | 99.9860 | 4.7 |
| 128 × 128 | UNet (AAR) | 99.22 | 99.70 | 99.85 | 99.9943 | 4.6 |
| 256 × 256 | UNet (AAR)_128[a] | 97.26 | 99.03 | 99.51 | 99.9817 | 4.7 |
| 256 × 256 | UNet (AAR)_256 | 98.00 | 99.30 | 99.65 | 99.9869 | 4.7 |
| 128 × 128 | ARMS Net | **99.45** | **99.77** | **99.88** | **99.9955** | 4.5 |

Bold values denote the highest result.
[a] 128 indicates the adaptive multi-scale feature extraction module to design the atrous rates combination and pooling sizes according to the input image size as 128.

the corresponding weight of the loss function. It can reduce the negative effects caused by category imbalance in various labels and prevent the segmentation performance degradation caused by excessive weight. Thus, the loss function helps optimize the training process, prevent gradient explosion, and further improve the IoU score of the overlapping region, showing stronger robustness.

Next, we conduct ablation experiments for the seven networks in Table 2, and each network is designed with four-layer pooling. Fig. 6 shows the IoU score box plot of the seven networks for the overlapping regions of chromosomes. It can be seen that the various modules involved in the ablation experiment are beneficial to improve the segmentation performance, and the segmentation performance of ARMS Net is the best. Specifically, the UNet (R) introduces Res Path to alleviate the semantic gap between the encoder and decoder, and the segmentation performance is slightly improved compared to UNet. On this basis, the UNet (PR) introduces the PPM module for a feature map with 8 × 8 pixels. Its pooling steps and sizes are 2, 3, and 4. However, due to the pooling operation with the pooling step of 3, much semantic information is lost during up-sampling, which deteriorates the network's performance. The UNet (AR) reduces the semantic information loss in the up-sampling by fixing the pooling step of ASSP module pooling to 2 and adaptively designs the pooling sizes to achieve better segmentation performance. Thus, based on UNet (AR), we further introduce the DAC module to construct the UNet (DAR). Although the introduction of the DAC module effectively improves the IoU score of overlapping chromosome region, as shown in Fig. 6, the stability of this model is not ideal enough. This is caused by the DAC module being able to extract multi-scale feature information, but its receptive fields may not match the feature map size. Thus, we finally construct the UNet (AAR) with an adaptive multi-scale feature extraction module to adaptively design a combination of continuous and small atrous rates and pooling sizes for the multi-scale overlapping regions of chromosomes to achieve multi-scale feature extraction and fusion. The segmentation performance and the generalization ability are further improved. On this basis, the loss function in UNet (AAR) is upgraded to an adaptive smooth weighted cross-entropy loss function, which optimizes the training process by weakening the negative effects caused by category imbalance in various labels to improve the segmentation performance further. The scores of nine networks under different evaluation metrics are compared in Table 5. By comparing UNet (AAR)_128 and UNet (AAR)_256, we can see that adjusting the size of the input image will cause image distortion

**Table 6**
Comparison table of ARMS Net and the latest research progress using this data set.

| | IoU (%) | Chrom $\alpha$ IoU (%) | Chrom $\beta$ IoU (%) | Accuracy (%) |
|---|---|---|---|---|
| Hu et al.[24] | 94.70 | 88.20 | 94.40 | 99.22 |
| Hu et al.+ TTA[a] | ~ | ~ | ~ | 99.27 |
| Saleh et al. [25] | ~ | ~ | ~ | 99.68 |
| Sun et al.[48] | 96.32 | ~ | ~ | 99.86 |
| CE-Net [26] | 90.35 | 96.04 | 97.76 | 99.92 |
| ARMS Net | **99.45** | **99.77** | **99.88** | **99.99** |

Bold values denote the highest result.
[a] TTA indicates Test Time Augmentation.

and is not conducive to improving the segmentation result. Thus the adaptive multi-scale feature extraction module provides a better segmentation performance for input images of different sizes without adjusting the image size than the fixed receptive field method. And it can be seen that the performance of ARMS Net is better than other networks regardless of the IoU of overlapping chromosome regions or the IoU and Accuracy of complete chromosomes, and the increase in prediction time due to the introduction of AMAC, ASSP and Res Path modules is not much.

The segmentation performance of ARMS Net and networks in other papers are compared in Table 6. It can be seen that whether the IoU of overlapping regions or the IoU and Accuracy of complete chromosomes are both used as the evaluation metrics, the performance of ARMS Net is greatly improved compared with those in other work, far better than the current popular network CE-Net for medical image segmentation. The main reason is that excessive pooling operations in CE-Net cause the receptive fields in DAC to be much larger than the feature map size, which is not conducive to the extraction of multi-scale features. Moreover, the pooling sizes designed by RMP are also not conducive to the restoration of the original image size after the up-sampling, resulting in a large amount of loss of semantic information and ultimately resulting in deterioration of segmentation performance.

Finally, this paper presents the segmentation results of ARMS Net, UNet and CE-Net in three kinds of typical overlapping situations (a,b,c), as shown in Fig. 7. The IoU score of the overlapping regions is listed in the right bottom corner. (a): overlapping region of two chromosomes is tiny; (b): two chromosomes are clearly cross; (c): two chromosomes are overlapped dramatically. It can be seen that ARMS Net performs better in segmentation scenarios with multi-scale overlapping areas. CE-Net and UNet obviously perform poorly in scenarios with tiny target regions.

## 5. Conclusion

In order to achieve the detailed segmentation of overlapping regions with different sizes and the adaptive extraction of multi-scale features, we construct an ARMS Net that adaptively extracts multi-scale features and alleviates semantic information gaps to achieve high performance of overlapping chromosome segmentation. It may provide more reliable input for subsequent chromosome recognition for genetic diagnosis of diseases. ARMS Net sets an appropriate number of pooling layers to balance the requirements of deep semantic information extraction and high precision detailed segmentation; It can adaptively correct the receptive fields without scaling or cropping the image size to preserve more image details and improve the ability to extract multi-scale features; Res Path module alleviates the semantic gap between the encoder and decoder instead of skip connection; In order to optimize the training process, an adaptive smooth weighted cross-entropy loss function is specially designed for the category imbalance issue.

Of course, the ARMS Net proposed in this paper still has shortcomings: ARMS Net, as a semantic segmentation network, only realized the segmentation of two overlapping chromosomes. Therefore, applying ARMS Net as a backbone network to end-to-end chromosome segmentation and identification will be the main research direction of this subject in the future.

## CRediT authorship contribution statement

**Guangjie Wang**: Conceptualization, Methodology, Software, Writing.

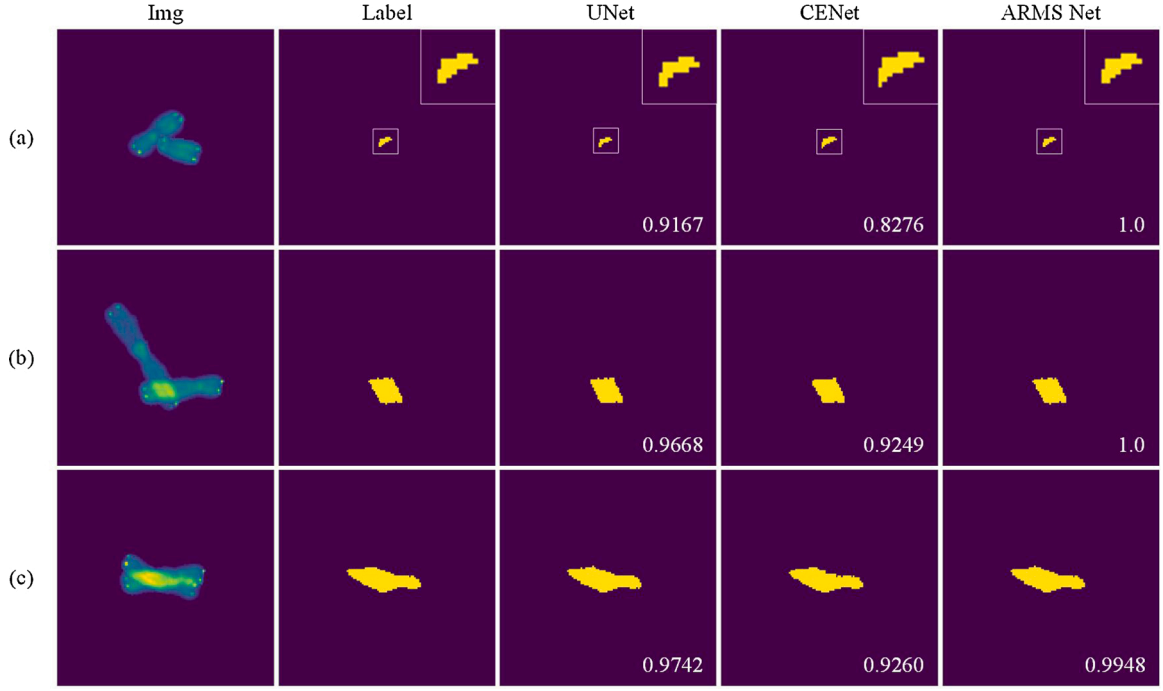


**Fig. 7.** Example of sample segmentation. (a) the overlapping region of two chromosomes is tiny; (b) the two chromosomes are clearly cross; (c) the two chromosomes are overlapped dramatically. The IoU score of the overlapping regions is listed in the right bottom corner.

**Hui Liu**: Supervision, Project administration.
**Xianpeng Yi**: Software, Conceptualization.
**Jinjun Zhou**: Data curation.
**Lin Zhang**: Review & editing.

## Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities (No. 2019ZDPY15).

*Declaration of Competing Interest*

The authors report no declarations of interest.

## References

[1] J.H. Tjio, A. Levan, The chromosome number of man, Am. J. Obstet. Gynecol. 130 (1978) 723–724.
[2] D. Wells, J.D.A. Delhanty, Comprehensive chromosomal analysis of human preimplantation embryos using whole genome amplification and single cell comparative genomic hybridization, Mol. Hum. Reprod. 6 (2000) 1055–1062.
[3] E. SchrCk, T. Veldman, H. Padilla-Nash, Y. Ning, J. Spurbeck, S. Jalal, L.G. Shaffer, P. Papenhausen, C. Kozma, C. Mary, Human Genetics Phelan, Spectral karyotyping refines cytogenetic diagnostics of constitutional chromosomal abnormalities, Hum. Genet. 101 (3) (1997) 255–262.
[4] Abid, Faroudja, Hamami, Latifa, A survey of neural network based automated systems for human chromosome classification, Artif. Intell. Rev. 49 (2016) 41–56.
[5] S. Jahani, S.K. Setarehdan, E. Fatemizadeh, Automatic identification of overlapping/touching chromosomes in microscopic images using morphological operators, 2011 7th Iranian Conference on Machine Vision and Image Processing (2011) 1–4.
[6] R.R. Schreck, C.M. Disteche, Chromosome Banding Techniques. Current Protocols in Human Genetics, 2001. Chapter 4:Unit4.2-Unit4.2.
[7] Elisa Garimberti, Sabrina Tosi, Fluorescence in situ hybridization (fish), basic principles and methodology, Methods Mol. Biol. (Clifton, N.J.) 659 (2010) 3–20.
[8] F. Altinordu, L. Peruzzi, Y. Yu, X.J. He, A tool for the analysis of chromosomes: karyotype, Taxon 65 (2016) 586–592.
[9] T. Arora, R. Dhir, A review of metaphase chromosome image selection techniques for automatic karyotype generation, Med. Biol. Eng. Comput. 54 (2016) 1147–1157.
[10] S. Almagro, S. Dimitrov, T. Hirano, M. Vallade, D. Riveline, Individual chromosomes as viscoelastic copolymers, Europhys. Lett. 63 (2003) 908–914.
[11] Devaraj Somasundaram, Machine learning approach for homolog chromosome classification, Int. J. Imaging Syst. Technol. 29 (2019) 161–167.
[12] Marina Yusoff, Muhamad Nazreen Shah Bin Mohd Ikram, Norjansalika Janom, Task assignment optimization for crowdsourcing using genetic algorithm, Adv. Sci. Lett. (2018) 8205–8208.
[13] M. Sharma, O. Saha, A. Sriraman, R. Hebbalaguppe, L. Vig, S. Karande, Crowdsourcing for chromosome segmentation and deep classification, 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (2017) 786–793.
[14] W. Srisang, K. Jaroensutasinee, Mullica of Science Jaroensutasinee, and Technology, Segmentation of overlapping chromosome images using computational geometry, Int. J. Hum. Genet. 3 (2006) 653.
[15] X.Y. Xu, S.Z. Xu, L.H. Jin, E.M. Song, Characteristic analysis of otsu threshold and its applications, Pattern Recognit. Lett. 32 (2011) 956–961.
[16] A. Cachia, J.F. Mangin, D. Riviere, D. Papadopoulos-Orfanos, F. Kherif, I. Bloch, J. Regis, A generic framework for the parcellation of the cortical surface into gyri using geodesic voronoi diagrams, Med. Image Anal. 7 (2003) 403–416.
[17] M.V. Munot, M.A. Joshi, Mandhawkar, Priyanka, Semi automated segmentation of chromosomes in metaphase cells, Conference on Image Processing (2012) 1–6.
[18] V.S. Balaji, S. Vidhya, Separation of touching and overlapped human chromosome images. Advancements of Medical Electronics, 2015, pp. 59–65.
[19] S. Javed, A. Mahmood, M.M. Fraz, N.A. Koohbanani, K. Benes, Y.-W. Tsang, K. Hewitt, D. Epstein, D. Snead, N. Rajpoot, Cellular community detection for tissue phenotyping in colorectal cancer histology images, Med. Image Anal. 63 (2020) 20.
[20] D. Somasundaram, S. Palaniswami, R. Vijayabhasker, V. Venkatesakumar, G-band chromosome segmentation, overlapped chromosome separation and visible band calculation, Int. J. Hum. Genet. 14 (2014) 73–81.
[21] F. Yang, F. Kruggel, Automatic segmentation of human brain sulci, Med. Image Anal. 12 (2008) 442–451.
[22] I.C. Yilmaz, J. Yang, E. Altinsoy, L. Zhou, Ieee, An improved segmentation for raw g-band chromosome images, 2018 5th International Conference on Systems and Informatics (2018) 944–950.
[23] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention, Pt III, 2015, pp. 234–241.
[24] R. Hu, J. Karnowski, R. Fadely, J.-P. Pommier, Image Segmentation to Distinguish Between Overlapping Human Chromosomes, 2017 (arXiv e-prints, page), arXiv: 1712.07639.
[25] H.M. Saleh, N.H. Saad, N.A. Mat Isa, Overlapping chromosome segmentation using u-net: convolutional networks with test time augmentation, Proc. Comput. Sci. 159 (2019) 524–533.
[26] Z.W. Gu, J. Cheng, H.Z. Fu, K. Zhou, H.Y. Hao, Y.T. Zhao, T.Y. Zhang, S.H. Gao, J. Liu, Ce-net: context encoder network for 2d medical image segmentation, IEEE Trans. Med. Imaging 38 (2019) 2281–2292.
[27] N. Ibtehaz, M.S. Rahman, Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation, Neural Netw. 121 (2020) 74–87.
[28] Y. Ruan, D. Li, H. Marshall, T. Miao, T. Cossetto, I. Chan, O. Daher, F. Accorsi, A. Goela, S. Li, Mb-fsgan: joint segmentation and quantification of kidney tumor on ct by the multi-branch feature sharing generative adversarial network, Med. Image Anal. 64 (2020) 1361–8415.
[29] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network. 30th IEEE Conference on Computer Vision and Pattern Recognition (2017) 6230–6239.
[30] S.N. Xie, Z.W. Tu, Ieee, Holistically-nested edge detection, 2015 IEEE International Conference on Computer Vision (2015) 1395–1403.
[31] T.Y. Lin, P. Goyal, R. Girshick, K.M. He, P. Dollar, Ieee, Focal loss for dense object detection, 2017 IEEE International Conference on Computer Vision (2017) 2999–3007.
[32] Y.L. Zhang, L.G. Shuai, Y.L. Ren, H.L. Chen, Ieee, Image classification with category centers in class imbalance situation, Proceedings 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC) (2018) 359–363.
[33] J.P. Pommier, Generating Images of Overlapping Chromosomes, 2016. https://www.kaggle.com/jeanpat/overlapping-chromosomes/data/.
[34] Z. Yu, S. Dai, Y. Xing, Ieee, Adaptive salience preserving pooling for deep convolutional neural networks, 2019 IEEE International Conference on Multimedia & Expo Workshops (2019) 513–518.
[35] R.X. Wang, M.M. Gong, D.C. Tao, Receptive field size versus model depth for single image super-resolution, IEEE Trans. Image Process. 29 (2020) 1669–1682.
[36] C. Szegedy, W. Liu, Y.Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Ieee, Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (2015) 1–9.
[37] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, 30th IEEE Conference on Computer Vision and Pattern Recognition (2017) 2261–2269.
[38] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 834–848.
[39] C. Meng, K. Sun, S.Y. Guan, Q. Wang, R. Zong, L. Liu, Multiscale dense convolutional neural network for dsa cerebrovascular segmentation, Neurocomputing 373 (2020) 123–134.
[40] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. M. Liang, Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging 35 (2016) 1299–1312.
[41] George Papandreou, Florian Schroff, Hartwig Adam, Liang-Chieh Chen, Rethinking Atrous Convolution for Semantic Image Segmentation, 2017 (arXiv e-prints, page), arXiv:1706.05587.
[42] F. Zhang, N.A. Cai, J.X. Wu, G.D. Cen, H. Wang, X.D. Chen, Image denoising method based on a deep convolution neural network, Iet Image Process. 12 (2018) 485–493.
[43] S.Q. Ren, K.M. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 1137–1149.
[44] L. Prechelt, Automatic early stopping using cross validation: quantifying the criteria, Neural Netw. Off. J. Int. Neural Net. Soc. 11 (1998) 761–767.
[45] P. Poudel, BongseogJang, Comparison of different deep learning optimizers for modeling photovoltaic power, J. Chosun Nat. Sci. 11 (2018) 204–208.
[46] Y.Z. Wang, J. Li, S. Zhang, B. Huang, G. Yao, J. Zhang, An RNA scoring function for tertiary structure prediction based on multi-layer neural networks, Mol. Biol. 53 (2019) 132–141.
[47] McGuinness, N.E.K. O'Connor, A comparative evaluation of interactive segmentation algorithms, Pattern Recognit. 43 (2010) 434–444.
[48] X. Sun, J. Li, J. Ma, H. Xu, T. Feng, Segmentation of overlapping chromosome images using u-net with improved dilated convolutions, J. Intell. Fuzzy Syst. 4 (2020) 1–16.