

A New Convolutional Neural Network Architecture for Automatic Segmentation of Overlapping Human Chromosomes

Sifan Song · Tianming Bai · Yanxin Zhao ·
Wenbo Zhang · Chunxiao Yang · Jia Meng ·
Fei Ma* · Jionglong Su*

Received: date / Accepted: date

Abstract In clinical diagnosis, karyotyping is carried out to detect genetic disorders due to chromosomal aberrations. Accurate segmentation is crucial in this process that is mostly operated by experts. However, it is time-consuming and labor-intensive to segment chromosomes and their overlapping regions. In this research, we look into the automatic segmentation of overlapping pairs of chromosomes. Different from standard semantic segmentation applications that mostly detect object regions or boundaries, this study attempts to predict not only non-overlapping regions but also the order of superposition and opaque regions of the underlying chromosomes. We propose a novel convolutional neural network called Compact Seg-UNet with enhanced deep feature learning capability and training efficacy. To address the issue of unrealistic images in use characterized by over-

Sifan Song
Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

Tianming Bai
School of Mathematics, the University of Edinburgh, UK

Yanxin Zhao
Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

Wenbo Zhang
Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

Chunxiao Yang
Suzhou Sano Precision Medicine Ltd., Suzhou 215123, China

Jia Meng
Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

Fei Ma*
Department of Applied Mathematics, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China E-mail: fei.ma@xjtlu.edu.cn

Jionglong Su*
School of AI and Advanced Computing, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
E-mail: jionglong.su@xjtlu.edu.cn

lapping regions of higher color intensities, we propose a novel method to generate more realistic images with opaque overlapping regions. On the segmentation performance of overlapping chromosomes for this new dataset, our Compact Seg-UNet model achieves an average IOU score of $93.44\% \pm 0.26$ which is significantly higher than the result of a simplified U-Net reported by literature by around 6.08%. The corresponding $F1$ score also increases from 0.9262 ± 0.1188 to 0.9596 ± 0.0814 .

Keywords Convolutional neural networks · Deep learning · Automatic segmentation · Overlapping chromosomes · Compact Seg-UNet

1 Introduction

In eucaryotes, the genetic information is encoded and packaged into a set of chromosomes. For example, the genome of human beings has been divided into 23 pairs of chromosomes. The first 22 pairs are homologous chromosomes which means they are common to both females and males. The last pair is called sex chromosomes and it is the only nonhomologous pair (two X in females, one X and one Y in males) (Alberts et al. 2014). Chromosomes are typically colored by fluorescent dyes at the metaphase of mitosis. At this stage, cells undergo nuclear division and their chromosomes are highly organized and compacted. To unambiguously display the 46 chromosomes in a human cell, researchers artificially rearrange them in numerical order (King 1974). Cytogeneticists can detect abnormalities associated with inherited defects. An original microphotograph and its corresponding karyotype are shown as Figure 1-a and b respectively.

In the medical profession, chromosome abnormalities (Munné et al. 1995; Lee et al. 2007; Marshall et al. 2008) are significant evidence in the diagnosis of genetic disorder such as Down syndrome (Roizen and Patterson 2003), Williams syndrome (Tassabehji et al. 1999), and cancer (MacLeod et al. 2011; Wan 2014). Some chromosome abnormalities, including atypical number and structural abnormalities, are distinct and visible in metaphase cells. Therefore, analyzing karyotypes segmented from metaphase cell images occupies a key role in cytogenetics and cancer studies (MacLeod et al. 2011; Wan 2014). Accurate segmentation is crucial to karyotyping which involves organizing and ordering pairs of homologous chromosomes in terms of their entire bands and features. Nonrigid shapes, overlapping chromosomes, stain debris and other noises increase the difficulty for chromosome segmentation (Minaee et al. 2014). As professional discretion is generally required in karyotyping, it is inevitably time-consuming and expensive (Srisang et al. 2011).

Convolutional neural networks (CNNs) refer to a class of feed-forward artificial networks, which have been proved to be efficient in image analysis. The success of CNNs advances the development in computer vision area, significantly improving the performance on many computer vision tasks, such as label prediction (Abdulnabi et al. 2015), neural style transfer (Gatys et al. 2016), object detection (Girshick et al. 2014; Ren et al. 2017), and image semantic segmentation (Girshick et al. 2014). In the field of semantic segmentation, specifically in scene understanding, U-Net (Ronneberger et al. 2015) and SegNet (Badrinarayanan et al. 2017) are two efficient and practical CNN architectures that have achieved many breakthrough results and showed advantages in end-to-end applications.

In this research, we focus on the automatic segmentation of overlapping chromosomes. Most semantic segmentation applications emphasize detecting object

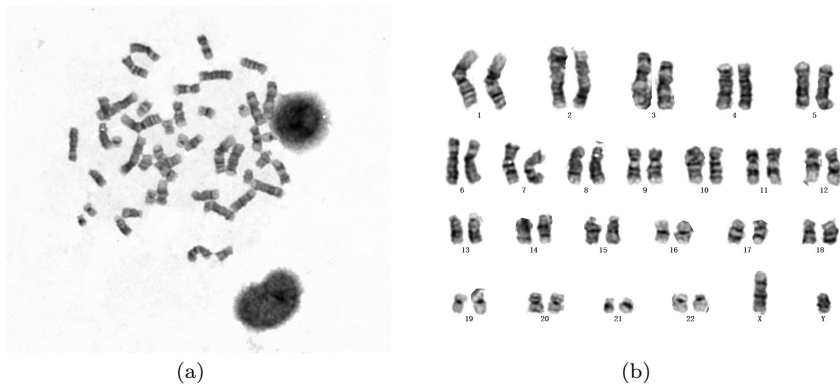


Fig. 1 (a) An example of the real microphotographs; (b) An example of the karyotypes.

regions or boundaries, e.g. Fast/Faster/Mask R-CNN (Girshick 2015; Ren et al. 2017; He et al. 2017) and pyramid CNN (Morris 2018). In recent years, several works concentrate more on the occlusion detection problem in many application scenarios (Kowal et al. 2020; Chen et al. 2020; Lazarow et al. 2020). In the medical segmentation field, Kowal et al. (2020) combine a CNN model and seeded watershed algorithm for addressing the problem of aggregated cell nuclei segmentation. BANet (Chen et al. 2020) focuses on handling occlusion by assigning occluded pixels to the correct object in panoptic segmentation tasks. OCFusion (Lazarow et al. 2020) resolves the occlusion problem by adding an additional “head”, named as the “occlusion head”, to the Mask R-CNN architecture for classifying pixels. However, these current works still focus on separating the boundaries rather than predicting the underlying regions between overlapped objects. This study attempts to predict not only non-overlapping regions but also the order of superposition and opaque regions of underlying chromosomes.

There are two key contributions in this study. First, a new neural network architecture called Compact Seg-UNet, a hybrid of U-Net (Ronneberger et al. 2015) and SegNet (Badrinarayanan et al. 2017), is proposed in this study to predict non-overlapping and overlapping segments separately. Second, to evaluate the segmentation performance of overlapping chromosomes, we modified the dataset construction method used in (Hu et al. 2017) to generate more realistic images with opaque overlapping regions.

The rest of this paper is organized as follows. In Section 2, we review several chromosome segmentation methods available in literature. Section 3 first provides details of dataset preparation, then the proposed approaches and the architecture of Seg-UNet & Compact Seg-UNet are introduced. Subsequently, we provide detailed results of the experiments in Section 4. Finally, conclusions and future work are given in Section 5.

2 Related work

2.1 Architecture of U-Net and SegNet

Before deep CNNs gain popularity in its application in computer vision, researchers regularly work on object recognition and edge detection. The deep learning methods can achieve significant improvements not only in edge detection but also in pixel-wise semantic segmentation. Different from Fully Convolutional Networks (Long et al. 2015), U-Net and SegNet (Ronneberger et al. (2015), Badrinarayanan et al. (2017)) are both designed as encoder-decoder architectures. The encoder path extracts and integrates the interior features of images, while the decoder ensures that the output details and sizes are identical to the input images.

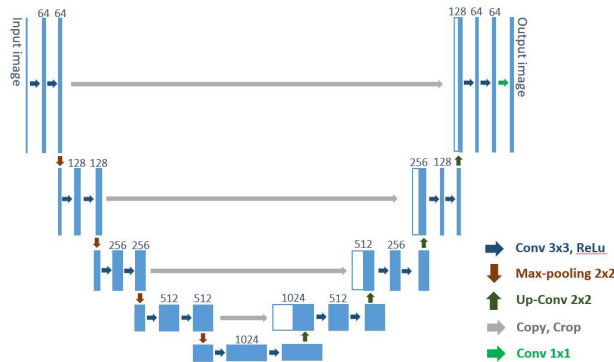


Fig. 2 Architecture of U-Net (Ronneberger et al. 2015)

The U-Net is initially applied to biomedical image segmentation (Ronneberger et al. 2015). It has a unique U-shaped architecture, comprising of a contracting path and an expansive path (Figure 2). The contracting path contains 4 blocks, each consisting of two convolutional layers with respective rectified linear units (ReLU) as activation functions, and one max-pooling layer for downsampling. In the expansive path, upsampling is achieved through transposed convolution operators. Each upsampling output is concatenated with a corresponding high-resolution feature map of the contracting path. Thus, high-resolution features are maintained and inherited across layers. (Ronneberger et al. 2015) utilizes the U-Net for cell tracking and segmentation in biomedical images. The U-Net achieves the best performance of 77.5% intersection over union (IOU) score which is 46% higher than its nearest competitor in ISBI challenge.

As a semantic segmentation architecture, SegNet is initially applied to road and indoor scene understanding scenarios which also require the neural networks to detect different objects, e.g., pedestrians, vehicles, doors and office chairs (Badrinarayanan et al. 2017). The SegNet consists of an encoder network (contracting path), a decoder network (expansive path), and a pixel-wise classifier in its final layer (Figure 3). The encoder network is identical to the first 13 convolutional layers of the VGG16 network (Simonyan and Zisserman 2014), in which each convolutional layer is supplemented by batch normalization and ReLU. The upsampling

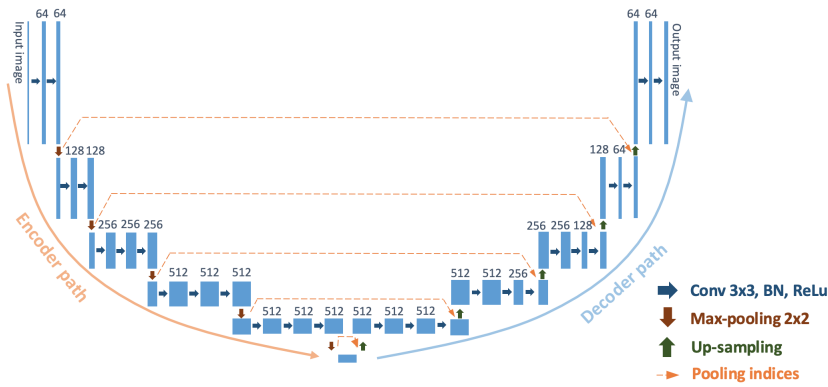


Fig. 3 Architecture of SegNet (Badrinarayanan et al. 2017)

step of SegNet is computed by corresponding *pooling indices* of the encoder path, and it differs from the transposed convolution method in U-Net.

2.2 Previous Work on Chromosome Overlapping Segmentation

Several methods and algorithms have been proposed for the automatic segmentation of overlapping chromosomes in metaphase images. (Karvelis et al. 2010) utilizes the watershed transform which decomposes images into watershed regions and gradient paths, and then merges adjacent regions to generate chromosome areas. Furthermore, a hybrid of fuzzy C-means and the watershed algorithm has been proposed to detect overlapping region (Manohar and Gawande 2017). Methods focusing to detect cut-points have also been studied. (Ranjan et al. 2012) proposes a novel method to detect pale path for chromosome images, which obtains an optimum number of cut-points and minimized grayscale intensity by self-adaptive searching windows. Delaunay triangulation is utilized to identify the number of overlapping chromosomes by detecting the optimal cut-points (Munot et al. 2013). Most of these geometric analyses can detect and segment overlapping chromosomes, but have weak performances when chromosomes are merely touching or partially overlapping. In practice, these methods may require a lot of human interventions which are time-consuming.

Although CNNs have been developed for over 20 years, they are seldom applied to the field of chromosome related image analysis. In 2017, (Sharma et al. 2017) and (Gupta et al. 2017) design pipelines for the automation of chromosome segmentation and classification. However, their chromosome segmentation is implemented by crowdsourcing method which is carried out manually. This is different from our purpose which is the automatic segmentation of overlapping chromosomes. A simplified U-Net (abbreviated as Sim U-Net) for automatic segmentation of overlapping chromosome pairs is proposed by (Hu et al. 2017), by retaining the first two downsampling and the corresponding upsampling blocks of the regular U-Net (Figure 2). The maximum channel size of Sim U-Net is 256. By training on randomly overlapped chromosome pairs, it quantitatively evalu-

ates the segmentation accuracies of overlapping and non-overlapping regions at pixel level. For improving the performance, a method is proposed by (Saleh et al. 2019) which combines a medium-size U-Net (abbreviated as Med U-Net) and test time augmentation (TTA). By retaining the first three downsampling and corresponding upsampling blocks, the maximum channel size of Med U-Net is 512. The depth of bottleneck layers of Med U-Net is 512 which is deeper than that of Sim U-Net (256) and shallower than that of the original U-Net (1024). It achieves IOU accuracies between 90.63%-99.94% which is significantly better than the range of 78.93%-99.93% (Hu et al. 2017) in its reproduced experiments. TTA is a method generally used after the training stage to improve the performance of test sets. It is not employed in this study because our focus is to compare the efficacies of different architectures.

3 Proposed Method

In this section, we provide a detailed description of our approach, organized as follows: a) The construction of datasets, b) pre-processing of data, c) architecture of Compact Seg-UNet and d) evaluation metrics.

3.1 Datasets

In this research, two datasets are used. Dataset 1 comprises 13434 pairs of overlapping chromosomes. The dataset is obtained by extracting 46 individual chromosomes from two microphotographs, one containing DAPI stained human metaphase chromosomes and the other including Cy3 labeled telomeres. The areas of chromosomes are calculated and ordered. After that, 12 chromosomes are picked by selecting every fourth chromosome from a sequence of chromosomes ordered by sizes (Lines 1 to 2 of Algorithm 1). They are combined to produce $\binom{12}{2} = 66$ chromosome pairs (Line 3 of Algorithm 1). With random rotation, 13434 overlapping chromosome images are generated (Dataset 1) (Line 4 of Algorithm 1). Dataset 1 is also used in (Hu et al. 2017). Figure 4-a gives four examples of the remained 13434 chromosome pairs. Their corresponding ground truth information is illustrated in Figure 4-b whose orange and green regions correspond to the non-overlapping regions of underlying and top chromosomes respectively, while blue regions correspond to the overlapping region.

A shortcoming of Dataset 1 is that, during image generation, the pixel-wise summation of greyscale values occurs when chromosomes are superimposed, resulting in lighter luminance (Figure 5-a). In this case, the non-overlapped regions of top and underlying chromosomes are indistinguishable and the lighter overlapping regions are distinct for recognition. This phenomenon will not happen to physical objects including chromosomes (Figure 5-c). For generating more realistic overlapping chromosomes, we modify the image generation method. In overlapping regions, only the pixel values of top chromosomes remained. A new dataset (Dataset 2) is then constructed in which the top chromosomes are opaque with respect to the underlying ones (Figure 5-b) (Lines 5 of Algorithm 1). The overlapping images and their order of the two datasets are identical except for

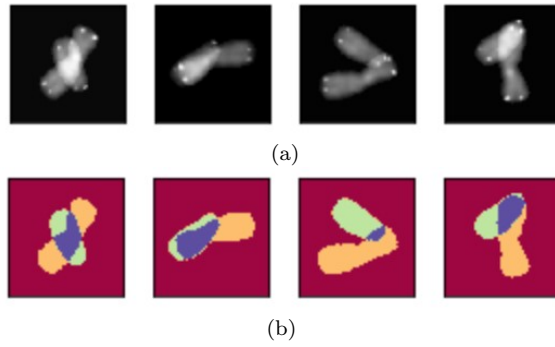


Fig. 4 (a) Four examples of generated overlapping chromosomes; (b) their corresponding ground truth (orange and green colors are used to distinguish between two chromosomes; blue color is used to represent the overlapping region)

the pixel values of overlapping regions. Dataset 1 and 2 are available at https://github.com/SifanSong/Chr_overlapping_datasets.

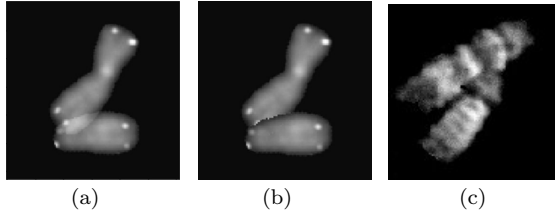


Fig. 5 (a) The overlapping region in Dataset 1 is lighter than non-overlapping areas; (b) We modify the overlapping method, making Dataset 2 appears more similar to real images; (c) In real microphotographs, opaque overlaps are more common than translucent ones

3.2 Pre-processing

The ground truth (Figure 6-GT) of Datasets 1 and 2 is initially transformed to four one-hot images for assessing the accuracies of different target regions. It is then denoised using a label correction method (Hu et al. 2017) to eliminate mislabeled pixels (Figure 6-i to iv). After denoising, Figure 6-v shows the label of background. Figure 6-vi and vii illustrate smoothed non-overlapping regions of the underlying and the top chromosomes respectively. Figure 6-viii shows the overlapping region. Pixel values of target regions are labeled as 1 (the black regions) and those of the rest are labeled as 0 (the white regions in Figure 6). Before being fed into CNNs, the background of all overlapping chromosome images and one-hot images are extended to 96×96 pixels (Lines 6 to 10 of Algorithm 1). It is notable that the pre-processing steps are identical in all experiments in this study.

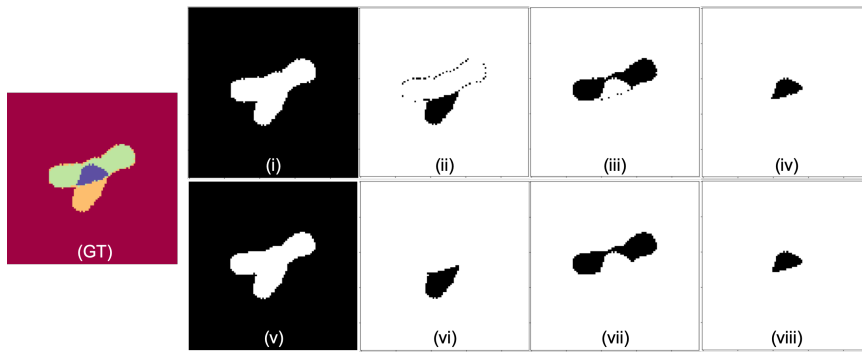


Fig. 6 (GT) An example of ground truth labels; (i) to (iv) four one-hot images without denoising; (v) to (viii) four one-hot images after denoising

Algorithm 1 Dataset preparation and pre-processing

Input: Extracted 46 chromosome digital images. The background of the image is black (0 pixel values).

Output: Two datasets for experiments.

- 1: The areas (non-zero pixel number) of chromosome images are counted and sorted in increasing order.
 - 2: An image set including 12 chromosomes are collected by choosing every fourth chromosome.
 - 3: Then $\binom{12}{2} = 66$ chromosome pairs are combined.
 - 4: By using different degrees of rotating, these 66 pairs generate 14850 images of which 1416 images that do not exist overlapping regions then are removed (13434 images remaining).
 - 5: In order to eliminate the deficiency of the pixel-wise summation of greyscale values in Dataset 1, we generate a more realistic dataset (Dataset 2). Compared to Dataset 1, Dataset 2 contains 13434 exactly identical chromosome overlaps but their overlapping regions are opaque.
 - 6: **for** each image $i \in \{1, 2, \dots, 13434\}$ of Dataset 1 and 2 respectively **do**
 - 7: Four-layer one-hot labeling (Figure 6 (i) to (iv)))
 - 8: Denoising (Figure 6 (v) to (viii)))
 - 9: Extending the image size to 96×96
 - 10: **end for**
 - 11: For fair comparison, all experiments implemented to Dataset 1 and 2 are with identical hyperparameters in this research. These are further described in **Section 4**.
-

3.3 Architecture of Compact Seg-UNet

Seg-UNet is a hybrid convolutional neural network combining the main characteristics of U-Net (Ronneberger et al. 2015) and SegNet (Badrinarayanan et al. 2017). It consists of an encoder path and a decoder path (Figure 7). The SegNet architecture is used as a framework, that is, each convolutional block contains a convolutional layer, a batch normalization and an activation function ReLU. In encoder path, 2×2 max-poolings with stride 2 are conducted to filter deep features when downsampling, and their *pooling indices* are saved. The architecture of the decoder path is almost symmetrical to that of the encoder. The *pooling indices* are utilized to perform upsampling (Figure 8). The upsampling guided by *pooling indices* restores double feature sizes by keeping positions with the best logits recorded in max-pooling (Badrinarayanan et al. 2017). In order to deliver low-level features, we concatenate the results of upsampling to corresponding layers with

the same feature sizes from the encoder path (Figure 8). These skip connections decrease the loss of features during upsampling in different scales and preserve more contour details (Ronneberger et al. 2015), (Chung et al. 2020).

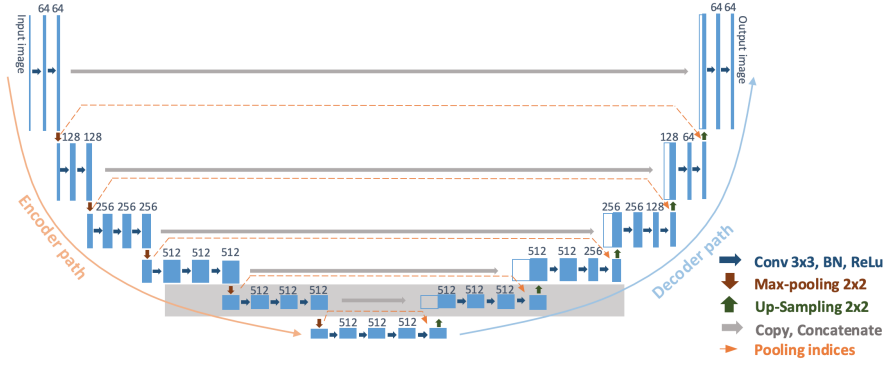


Fig. 7 The architecture of Seg-UNet and its compact version

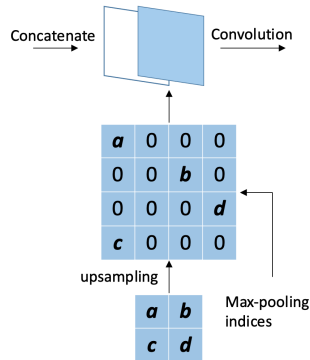


Fig. 8 upsampling of Compact Seg-UNet

The narrow and thick bottleneck layers introduce excessive trainable parameters in the Seg-UNet, leading to overfitting when applied to small datasets (Wu et al. 2017). To avoid this pitfall, we customize the depth of Seg-UNet by removing these layers (the grey box in Figure 7). We denote this architecture with a smaller network size as the Compact Seg-UNet.

3.4 Evaluation metrics

To evaluate the performance of segmentation in pixel-level, we first apply three measurement metrics, IOU, precision and recall:

$$IOU = \frac{Intersection}{Union} \times 100\% = \frac{TP}{TP + FP + FN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

where *Intersection* is equal to *TP* (True Positive) representing existing segments and being predicted correctly; *FP* (False Positive) is non-existing and wrongly predicted segments; *FN* (False Negative) represents segments that exist but not predicted; and *Union* is a summation of *TP*, *FP* and *FN*. For every one-hot image, we use IOU, precision and recall to assess the performance of segmentation. We overlay two corresponding one-hot images of ground truth and predicted segmentation to demonstrate them (Figure 9). In Figure 9, the colors black, blue and yellow represent *TP*, *FP* and *FN* respectively. We further use *F1* score to combine the precision and recall and to provide a balanced measurement of these two metrics:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

where *Precision* and *Recall* are previously defined in (2) and (3) respectively.

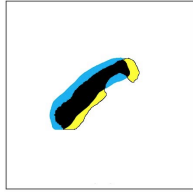


Fig. 9 An illustration of True Positives (black color), False Positives (blue color), and False Negatives (yellow color).

4 Experiments and Results

4.1 Implementation Details

Due to the fact that the one-hot images are composed of 0 and 1 pixel values, we apply sigmoid and binary cross-entropy as the last activation function and the loss function respectively. Our experiments are coded by PyTorch 3.7.5, and run on a NVIDIA RTX 2080Ti GPU and an Intel(R) Core(TM) i9-9900K CPU. We list all detailed hyperparameters in Table 1. For fair comparison, we apply the configuration described in Table 1 to every experiment and each dataset. We

Table 1 The Detailed Hyperparameters of All Experiments.

Hyperparameters	Details
Training:Validation Sets	9548:2386 (5-fold cross-validation)
Testing Set	1500 images
Optimizer	Adam
Fixed Random Seed	152
Batch Size	8
Learning Rate	0.00002
Decay Rate	0.5
Patience for Learning Rate Decay	3
Patience for Stop Training	9

randomly choose 1500 images as a testing set. The order of remained images is randomized by a fixed random seed = 152. These remained images are split in the ratio of 4:1 (9548:2386 images) for training and validation (batch size = 8). We then perform 5-fold cross-validation to prevent overfitting issues. The training process is optimized using Adam optimizer. We utilize the early stopping training strategy in this study, and the initial learning rate lr is 0.00002. The validation performance is checked once per epoch, and model weights with the best validation performance (the lowest loss of validation sets) are saved. If the best validation performance is not updated after 3 consecutive epochs, the lr is decreased by 0.5 decay rate. If the best validation performance is not updated after 9 consecutive epochs, the training process is terminated for avoiding overfitting.

Table 2 The IOU scores (mean \pm 2 standard deviation) of one-hot images (5-fold cross-validation)

Exp. ^a	Architectures	DS ^b	IOU_0 ^c	IOU_1 ^d	IOU_2 ^e	IOU_3 ^f
A	Sim U-Net	1	99.88% \pm 0.02	72.94% \pm 1.48	85.07% \pm 0.92	78.31% \pm 0.72
B	Med U-Net	1	99.82% \pm 0.02	82.74% \pm 1.76	91.80% \pm 0.78	75.91% \pm 1.52
C	Seg-UNet	1	99.97%\pm0.00	93.63%\pm0.42	98.11%\pm0.24	88.21%\pm0.58
D	Compact Seg-UNet	1	99.97%\pm0.00	93.82%\pm0.32	98.20%\pm0.22	88.51%\pm0.56
E	Sim U-Net	2	99.86% \pm 0.02	87.58% \pm 1.36	89.68% \pm 1.08	72.30% \pm 2.36
F	Med U-Net	2	99.84% \pm 0.02	90.23% \pm 1.10	92.03% \pm 0.72	74.34% \pm 2.30
G	Seg-UNet	2	99.97%\pm0.00	95.35%\pm0.30	95.70%\pm0.52	81.99%\pm2.60
H	Compact Seg-UNet	2	99.97%\pm0.00	95.48%\pm0.06	95.81%\pm0.26	82.49%\pm0.74

^a Experiments^b Datasets^c IOU scores of the background region^d IOU scores of the non-overlapping region of the underlying chromosome^e IOU scores of the non-overlapping region of the top chromosome^f IOU scores of the overlapping region

4.2 Results

To better assess the segmentation performance of overlapping chromosomes, we employ four CNN architectures (Sim U-Net, Med U-Net, Seg-UNet and Compact Seg-UNet) on the two datasets. Since Sim U-Net has the shallowest architecture among the four CNNs, we use its results as baselines. As shown in Table 2 and Table 3, Experiments A to D are conducted using Dataset 1 and Experiments E to H are carried out using Dataset 2. Specifically, Experiment A and E utilize Sim U-Net; both Experiments B and F use Med U-Net; both Experiments C and G utilize Seg-UNet; and both Experiment D and H use Compact Seg-UNet. Table 2 lists the architectures, datasets, and four IOU scores whose meanings are consistent with Figure 6-b-v to viii respectively. To further evaluate the segmentation performance of these architectures, we calculate sizes of GPU memory usage (GMU), average IOU, precision, recall and F1 scores (Table 3). We note that all the results of Table 2 and Table 3 are average scores ± 2 standard deviation ($2std.$) of test sets of 5-fold cross-validation.

Table 3 The average scores (± 2 standard deviation) of the experiments (5-fold cross-validation)

Exp. ^a	GMU ^b (MiB)	Avg. ^c IOU	Avg. Precision	Avg. Recall	Avg. F1
A	1443	84.05% \pm 0.78	89.51% \pm 0.64	90.80% \pm 0.48	0.9015 \pm 0.1286
B	1609	87.57% \pm 1.02	91.68% \pm 1.08	92.72% \pm 0.72	0.9220 \pm 0.1180
C	2455	94.98%\pm0.32	96.58%\pm0.28	96.88%\pm0.24	0.9673\pm0.0608
D	2251	95.12%\pm0.28	96.70%\pm0.36	96.93%\pm0.24	0.9681\pm0.0596
E	1443	87.36% \pm 1.20	92.57% \pm 0.74	92.68% \pm 0.74	0.9262 \pm 0.1188
F	1609	89.11% \pm 1.04	93.50% \pm 1.20	93.58% \pm 0.76	0.9354 \pm 0.1124
G	2455	93.25%\pm0.86	95.93%\pm1.26	95.80%\pm0.68	0.9586\pm0.0832
H	2251	93.44%\pm0.26	96.08%\pm0.44	95.84%\pm0.54	0.9596\pm0.0814

^a Experiments

^b GPU Memory Usage

^c average scores

For the Experiments (A to D) conducted on Dataset 1 and the Experiments (E to H) conducted on Dataset 2, Table 2 indicates that all IOU scores of Seg-UNet and Compact Seg-UNet have a significantly better performance than those of Sim U-Net and Med U-Net, since their confidence intervals (average scores $\pm 2std.$) do not have overlap with those of Sim U-Net and Med U-Net. At the same time, their $2std.$ scores are much lower than those of Sim U-Net and Med U-Net. The segmentation performance of Med U-Net is generally better than that of Sim U-Net. It indicates that Sim U-Net is too shallow to achieve a good performance in this study, so a deeper architecture (Med U-Net) improves the capacity to learn deep features. In Table 2, the comparisons of Seg-UNet and Compact Seg-UNet (Experiments C to D and G to H) show that results of IOU_0 to IOU_3 are almost similar, but the results of Seg-UNet may be affected by overfitting due to this excessively complex architecture. In contrast, the results of Compact Seg-UNet are slightly better than those of Seg-UNet. Compact Seg-UNet achieves the best IOU scores in the range of

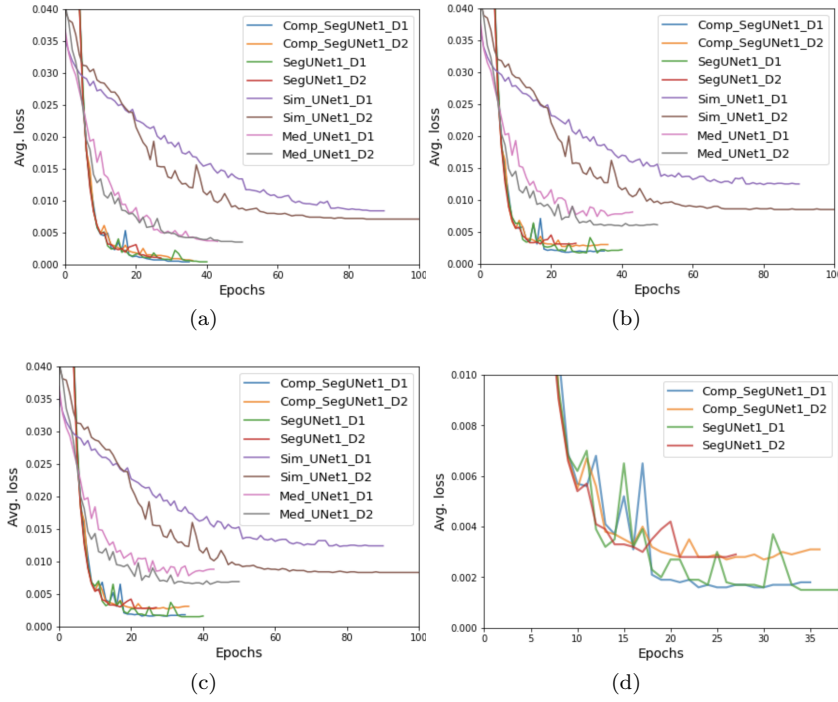


Fig. 10 (a) (b) (c) Examples of average losses of training, validation and test sets versus epochs respectively, where D1 and D2 represent Dataset 1 and Dataset 2; (d) the zoom-in plot of Seg-UNet and Compact Seg-UNet experiments of test sets.

88.51%(±0.56)-99.97%(±0.00) on Dataset 1 and 82.49%(±0.74)-99.97%(±0.00) on Dataset 2. Especially, Experiments D and H show that the robustness of Compact Seg-UNet model is superior to that of Seg-UNet with respect to lower $2std.$ scores. Although IOU₃ scores of Dataset 2 are lower than those of Dataset 1, these IOU scores demonstrate the flaws of Dataset 1. The directly summed pixel values cause unrealistic lighter overlapping regions as we described above, and this phenomenon cannot be observed in the real-world. Therefore, these lighter regions not only are distinct to neural networks, but also do not appear which chromosomes are underlying. Comparing Experiments (A-D) to (G-H), these flaws have been reflected by higher scores of the IOU₃ and poorer ability to distinguish the non-overlapping regions of underlying chromosomes (IOU₁). Although Experiments (C and D) on Dataset 1 show higher average scores in both Table 2 and Table 3, we still recommend setting the results on Dataset 2 as benchmarks in future research and evaluations, since the underlying regions are not transparent under microscopes and CNNs should learn to predict the extend of overlapping.

In Table 3, the average IOU is the mean of IOU₀ to IOU₃ of Table 2. We observe that relationships of other average $\pm 2std.$ results (precision, recall and F1) of four CNN architectures are consistent with IOU scores. Although Sim U-Net and Med U-Net only have 1443 and 1609 MiB GMU respectively, their segmentation results have been significantly improved by altering architectures and may

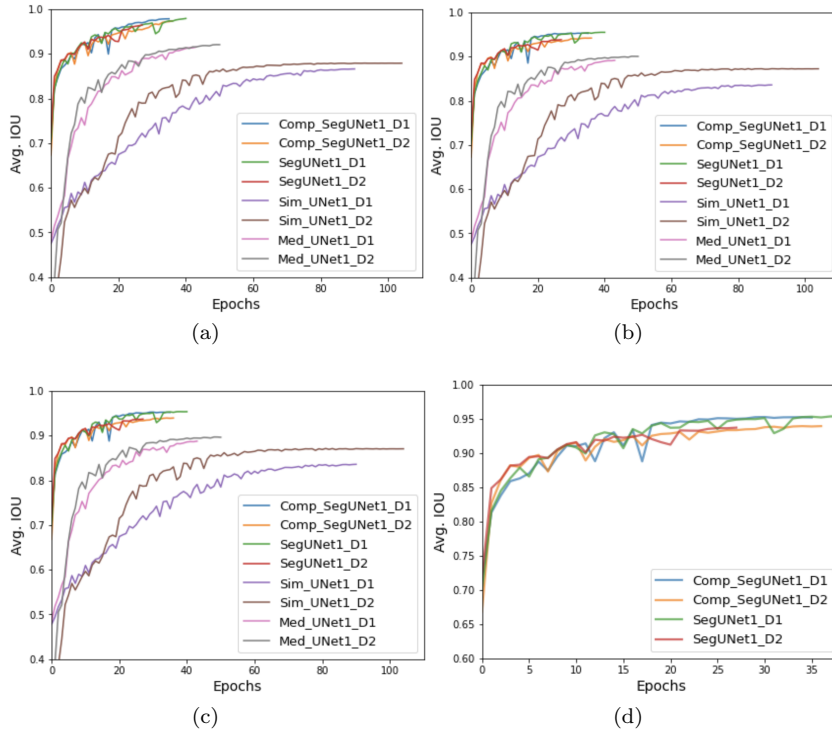


Fig. 11 (a) (b) (c) Examples of average IOU scores of training, validation and test sets versus epochs respectively, where D1 and D2 represent Dataset 1 and Dataset 2; (d) the zoom-in plot of Seg-UNet and Compact Seg-UNet experiments of test sets.

not satisfy the requirement of further research on automatic segmentation of overlapping chromosomes. Compact Seg-UNet (Experiments D and H) also achieves the best performance in overlapping chromosome segmentation (an average IOU of $93.44\% \pm 0.26$ and a $F1$ score of 0.9596 ± 0.0814 on Dataset 2) and its GMU has been reduced from 2455 (Seg-UNet) to 2251 MiB. These results highlight the fact that the removal of bottleneck layers of Seg-UNet not only reduces the load of training but also exhibits superiority in this study.

To further explore the segmentation performance in training processes, we present average losses of test sets versus epochs (Figure 10) and average IOU accuracies of test sets versus epochs (Figure 11). The results of the first fold of cross-validation are recorded and illustrated as examples. The curves of Sim UNet and Med UNet are distinctly separated from those of Seg-UNet and Compact Seg-UNet. Figure 10-a and Figure 11-a demonstrate the continuous overfitting of models is terminated by our early stopping strategy when validation performances are not updated for a while (Figure 10-b and Figure 11-b). Figure 10-c and Figure 11-c are average scores of test sets, and they also show that both Seg-UNet and Compact Seg-UNet significantly improve average results and decrease total epochs before early stopping. Figure 10-d and Figure 11-d are zoomed-in plots of the curves of Seg-UNet and Compact Seg-UNet. We observe that the experiments

conducted using each dataset show similar trends and are highly consistent. They highlight the superiority of the Compact Seg-UNet and Seg-UNet over Sim U-Net and Med U-Net in this task with high efficacies.

5 Conclusions and Future Work

In this work, a novel deep learning neural network architecture, Compact Seg-UNet, is proposed to segment images with overlapping chromosomes. With the removal of several low-resolution layers, Compact Seg-UNet requires relatively less GPU memory usage. The scarcity of real-world overlapping chromosome pairs motivates the construction of a dataset with generated overlapping chromosomes. To evaluate the performance of the proposed method, we also compare Compact UNet with Sim U-Net, Med U-Net and Seg-UNet. With the measurement metrics of IOU, precision, recall and F1 scores, the proposed Compact Seg-UNet is superior to other architectures in terms of segmentation performance.

For the public dataset (Dataset 1), there are lighter overlapping regions due to pixel-wise summation of the respective greyscale values. Such feature is not only unrealistic but can be directly recognized by neural networks. In view of this, we propose a modified approach to generate images with opaque overlapping regions that are more commonly seen in the real-world (Dataset 2). On Dataset 2, our proposed Compact Seg-UNet achieves the best average IOU score ($93.44\% \pm 0.26$) and the highest average $F1$ score (0.9596 ± 0.0814). Those significantly outperform the previous work by large margins.

When CNNs have identified and learned interior textures from chromosomes, the trained models can predict the shape, size and obstructed overlapping region of chromosomes in this study. This differs from most semantic segmentation applications whose emphases are merely detecting object regions or boundaries, e.g. Fast/Faster/ Mask R-CNN and pyramid CNN.

The achievement in this research is the first step towards the segmentation of chromosomes with higher degrees of overlapping. The research shows that Compact Seg-UNet can be used to segment overlapping chromosomes and predict nonrigid shapes. Compact Seg-UNet is designed for alleviating the overfitting problem. It not only improves the segmentation performance but also reduces computational costs. In future research, apart from customizing the size of CNN architectures, methods such as, dropout (Srivastava et al. 2014), weight initialization (He et al. 2015) and stochastic weight averaging (Izmailov et al. 2018) may be integrated into our architecture to further improve the generalization and robustness.

Despite the encouraging results obtained with Compact Seg-UNet in this study, Datasets 1 and 2 are artificially generated from merely 12 individual chromosomes. This is the main weakness impeding the practicability of models to segment overlapping chromosomes in more complicated scenarios. To achieve robust segmentation performance on more realistic images with overlapping chromosomes, we would augment training sets with real-world data and various chromosome shapes for constructing a great variety of overlapping conditions and chromosome individuals. In the next step of the research, we would also label and use real overlapping chromosome images to train our proposed neural network for improving model robustness, as well as to test fine-tuned models on real chromosome images for an assessment of its efficacy.

Acknowledgment

This study is supported by the National Science and Foundation of China (NSFC) under Grant 61501380, Key Program Special Fund in XJTU (KSF-T-01), the Key Program Special Fund in XJTU (KSF-A-22), the Key Programme Special Fund in XJTU (KSF-E-21), Jiangsu Society and Science Development Program, Project No. BE2016678, and the Open Program of Neusoft corporation, item number SKLSAOP1702.

References

- Abdulnabi, Abrar H, Gang Wang, Jiwen Lu, and Kui Jia. 2015. Multi-task CNN Model for Attribute Prediction. *IEEE Transactions on Multimedia* 17 (11): 1949–1959.
- Alberts, B., A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter. 2014. *Molecular Biology of the Cell. 500 tips*. Garland Science.
- Badrinarayanan, V., A. Kendall, and R. Cipolla. 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12): 2481–2495.
- Chen, Yifeng, Guangchen Lin, Songyuan Li, Omar Bourahla, Yiming Wu, Fangfang Wang, Junyi Feng, Mingliang Xu, and Xi Li. 2020. BANet: Bidirectional Aggregation Network with Occlusion Handling for Panoptic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3793–3802.
- Chung, Minyoung, Jinyoung Lee, Minkyung Lee, Jeongjin Lee, and Yeong-Gil Shin. 2020. Deeply Self-supervised Contour Embedded Neural Network Applied to Liver Segmentation. *Computer Methods and Programs in Biomedicine*.
- Gatys, Leon A, Alexander S Ecker, and Matthias Bethge. 2016. Image Style Transfer using Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2414–2423.
- Girshick, Ross. 2015. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, 1440–1448.
- Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.
- Gupta, Gaurav, Mohit Yadav, Monika Sharma, Lovekesh Vig, et al. 2017. Siamese Networks for Chromosome Classification. In *Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on*, 72–81.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving Deep into Rectifiers: Surpassing Human-level Performance on Imagenet Classification. In *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034.
- He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask R-CNN. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2980–2988.
- Hu, R Lily, Jeremy Karnowski, Ross Fadely, and Jean-Patrick Pommier. 2017. Image Segmentation to Distinguish between Overlapping Human Chromosomes. In *2017 Machine Learning for Health (NIPS)*. Long Beach, CA: Long Beach, CA.
- Izmailov, Pavel, Dmitrii Podoprikin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. 2018. Averaging Weights Leads to Wider Optima and Better Generalization. *arXiv preprint arXiv:1803.05407*.
- Karvelis, Petros, Aristidis Likas, and Dimitrios I Fotiadis. 2010. Identifying Touching and Overlapping Chromosomes using the Watershed Transform and Gradient Paths. *Pattern Recognition Letters* 31 (16): 2474–2488.
- King, Robert C. 1974. *A Dictionary of Genetics: Rev.* Oxford University Press.
- Kowal, Marek, Michał Żejmo, Marcin Skobel, Józef Korbicz, and Roman Monczak. 2020. Cell Nuclei Segmentation in Cytological Images using Convolutional Neural Network and Seeded Watershed Algorithm. *Journal of Digital Imaging* 33 (1): 231–242.
- Lazarow, Justin, Kwonjoon Lee, Kunyu Shi, and Zhuowen Tu. 2020. Learning Instance Occlusion for Panoptic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10720–10729.

- Lee, Charles, A John Iafrate, and Arthur R Brothman. 2007. Copy Number Variations and Clinical Cytogenetic Diagnosis of Constitutional Disorders. *Nature Genetics* 39: 48.
- Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.
- MacLeod, Roderick AF, Maren Kaufmann, and Hans G Drexler. 2011. Cytogenetic Analysis of Cancer Cell Lines. In *Cancer Cell Culture*, 57–78. Springer.
- Manohar, Ruchita, and Jayanand Gawande. 2017. Watershed and Clustering Based Segmentation of Chromosome Images. In *Advance Computing Conference (IACC), 2017 IEEE 7th International*, 697–700.
- Marshall, Christian R, Abdul Noor, John B Vincent, Anath C Lionel, Lars Feuk, Jennifer Skaug, Mary Shago, Rainald Moessner, Dalila Pinto, Yan Ren, et al.. 2008. Structural Variation of Chromosomes in Autism Spectrum Disorder. *The American Journal of Human Genetics* 82 (2): 477–488.
- Minaee, Shervin, Mehran Fotouhi, and Babak Hossein Khalaj. 2014. A Geometric Approach to Fully Automatic Chromosome Segmentation. In *Signal Processing in Medicine and Biology Symposium (SPMB), 2014 IEEE*, 1–6.
- Morris, Daniel D. 2018. A Pyramid CNN for Dense-Leaves Segmentation. In *2018 15th Conference on Computer and Robot Vision (CRV)*, 238–245. doi:10.1109/CRV.2018.00041.
- Munné, Santiago, Mina Alikani, Giles Tomkin, Jamie Grifo, and Jacques Cohen. 1995. Embryo Morphology, Developmental Rates, and Maternal Age are Correlated with Chromosome Abnormalities. *Fertility and sterility* 64 (2): 382–391.
- Munot, Mousami V, Jayanta Mukherjee, and Madhuri Joshi. 2013. A Novel Approach for Efficient Extrication of Overlapping Chromosomes in Automated Karyotyping. *Medical & Biological Engineering & Computing* 51 (12): 1325–1338.
- Ranjan, Rajeev, Akila Subasinghe Arachchige, Jagath Samarabandu, Peter K Rogan, and Joan Knoll. 2012. Automatic Detection of Pale Path and Overlaps in Chromosome Images using Adaptive Search Technique and Re-thresholding. In *VISAPP (1)*, 462–466.
- Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. Faster R-CNN: towards Real-time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*.
- Roizen, Nancy J, and David Patterson. 2003. Down’s Syndrome. *The Lancet* 361 (9365): 1281–1289.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, 234–241.
- Saleh, Hariyanti Mohd, Nor Hidayah Saad, and Nor Ashidi Mat Isa. 2019. Overlapping Chromosome Segmentation using U-Net: Convolutional Networks with Test Time Augmentation. *Procedia Computer Science* 159: 524–533.
- Sharma, Monika, Oindrila Saha, Anand Sriraman, Ramya Hebbalaguppe, Lovekesh Vig, and Shirish Karande. 2017. Crowdsourcing for Chromosome Segmentation and Deep Classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 786–793.
- Simonyan, Karen, and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* abs/1409.1556.
- Srisang, Wacharapong, Krisanadej Jaroensutasinee, and Mullica Jaroensutasinee. 2011. Segmentation of Overlapping Chromosome Images using Computational Geometry. *Walailak Journal of Science and Technology (WJST)* 3 (2): 181–194.
- Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research* 15 (1): 1929–1958.
- Tassabehji, Mayada, Kay Metcalfe, Annette Karmiloff-Smith, Martin J Carette, Julia Grant, Nick Dennis, W Reardon, Miranda Splitt, Andrew P Read, and Dian Donnai. 1999. Williams Syndrome: Use of Chromosomal Microdeletions as a Tool to Dissect Cognitive and Physical Phenotypes. *The American Journal of Human Genetics* 64 (1): 118–125.
- Wan, Thomas SK. 2014. Cancer Cytogenetics: Methodology Revisited. *Annals of Laboratory Medicine* 34 (6): 413–425.
- Wu, Bingzhe, Zhichao Liu, Zhihang Yuan, Guangyu Sun, and Charles Wu. 2017. Reducing Overfitting in Deep Convolutional Neural Networks using Redundancy Regularizer. In *International Conference on Artificial Neural Networks*, 49–55. Springer. Springer.