END-TO-END CHROMOSOME KARYOTYPING WITH DATA AUGMENTATION USING GAN

Yirui Wu[‡†] *Yisheng Yue*[†] *Xiao Tan* [§] *Wei Wang*[§] *Tong Lu*[†*]

[‡] College of Computer and Information, Hohai University
[†] National Key Lab for Novel Software Technology, Nanjing University
[§] Nanjing iRobsen Intelligent Technology Company
{wuyirui@hhu.edu.cn;abelyys@foxmail.com;{tanxiao,wangwei}@irobsen.com;lutong@nju.edu.cn}

ABSTRACT

Classifying human chromosomes from input cell images, *i.e.*, karyotyping, requires domain expertise and quantity of manual effort to perform. In this paper, we propose an end-to-end chromosome karyotyping method, which can automatically detect, segment and classify chromosomes from cell images. During detection, we explore Extremal Regions (ER) to obtain chromosome candidates in input images. During segmentation, we segment overlapping chromosome candidates by approximating chromosome shapes with eclipses. In classification, we first propose Multiple Distribution Generative Advertising Network (MD-GAN) to effectively cover diverse data modes and generate more labeled samples for data augmentation. Then, we finetune pre-trained convolutional neural network (CNN) to classify chromosomes with samples generated by MD-GAN. We demonstrate the accuracy of the proposed end-to-end method in detecting, segmenting and classifying by experiments on a self-collected dataset. Experiments also prove data augmentation with MD-GAN could improve classification performance of CNN.

Index Terms— Chromosome karyotyping, Data Augmentation, Generative Advertising Network

1. INTRODUCTION

Karyotyping refers to the process of classifying 23 pairs of human chromosomes from cell images. Conventionally, karyotyping is performed where the condensed chromosome images are Giemsa stained. Efficiently karyotyping is required due to its widely use in cytogenetics analyzing chromosome images to diagnose genetic disorders, birth defects and cancers [1]. However, even after years of expertise, doctors still need pay considerable manual effort and time to produce desirable karyotyping results [2].

Automatically karyotyping is challenging due to the unpredictable shapes and appearances caused by non-rigid nature of chromosomes. Karyotyping usually consists of chromosome detection, overlap segmentation and category classification. The most popular methods for detection are based on binarization using either the Otsu method [3] or a re-thresholding scheme [4]. However, uneven Giemsa staining may cause the fail of detection, since binarization is highly related with thresholds. Inspired by [5] which explores Extremal Regions (ER) to detect character candidates, we detect chromosomes candidates by utilizing ER. Due to the fact that chromosomes may touch and overlap, research attempts have been made to segment clusters of either touching [6] or overlapping chromosomes [7], where geometric and intensity based features have been used. During category classification, earlier methods would first extract manually designed features [8] and then apply learning structures [9] for chromosome classification. However, traditional schemes result in low accuracy due to the loss of useful information by manually designing features and feature selection.

Deep neural networks have demonstrated the state-of-the-art power for vision tasks [5, 10, 11], so that they have been applied to chromosome karyotyping [12]. However, these methods require a large amount of labeled data. Therefore, it is a difficult task to classify images with multiple class labels using only a small number of labeled examples. Chromosome classification is a typical example, since the labeled chromosome images are difficult to achieve due to the privacy of individual. Therefore, researchers have proposed several architectures to handle limited data scenarios [13, 14]. One of the most successful methods is data augmentation using GAN [15]. The key idea of GAN stems from the two-player game designed by GAN, *i.e.*, generator and discriminator, which provides a simple but powerful way to estimate target distribution and generate novel image samples. With this power for distribution modeling, GAN is extremely suitable to increase the size of training set for more efficient deep learning [16, 17]. However, training GAN for particular usage is challenging as it can be easily trapped into the mode collapsing problem where the generator only concentrates on producing samples lying on a few modes instead of the whole data space [18, 19].

Based on these considerations, we propose a novel end-to-end chromosome karyotyping method. Fig. 1 gives the overview of the proposed method. Chromosome candidates

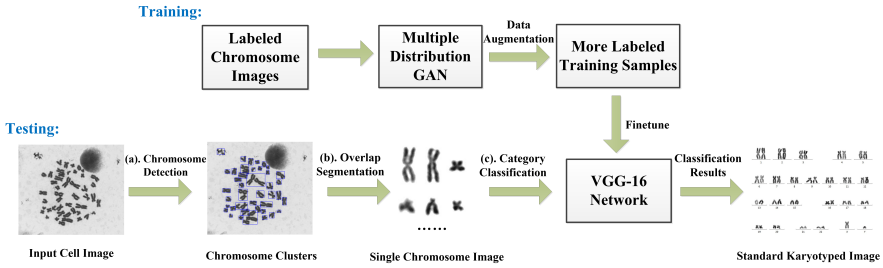

Fig. 1. The proposed framework for end-to-end chromosome karyotyping.

are obtained from the input cell image by Extremal Regions (ER) with several geometric and intensity filters. Next, we segment overlapping and touching chromosome clusters to individual ones. During classification, MD-GAN is proposed to diverse data modes by utilizing multiple data distributions. Based on the generated training samples, we finetune VGG-16 network to achieve desirable classification results, which are sorted into a standard karyotyping image.

The main contribution of the paper is to propose an end-to-end chromosome karyotyping method. We explore ER algorithm to detect chromosomes, which solves the problem of false detection caused by uneven staining and avoids time-consuming computation brought by adopting complex features for detection. The proposed MD-GAN employs a mixture of data distributions to generate diverse training samples instead of using multiple generators, which not only overcomes the mode collapsing problem, but also saves computation and reduces complexity. By adopting generated samples for training, we further prove data augmentation with MD-GAN could improve classification accuracy.

2. METHODOLOGY

In this section, we describe the proposed method by chromosome candidate detection, cluster segmentation and classification with data augmentation.

2.1. Chromosome Candidate Detection

It is true that chromosome in different cell images have characteristics which play a prominent role in representing chromosomes, namely, shape, contrast, uniform staining color inside the chromosome. To exploit such features, we explore Extremal Regions (ER) [21] to detect chromosome candidates. For an input cell image, we generate chromosome candidates, say $\{e_i\}$ of the input image I by detecting ER in gray cell image. Due to background variations, ER cann't detect chromosome candidates accurately. We thus propose the geometric and intensity filters for the output of ER.

1) Filter using geometric properties: Since nucleus or small points are the main noise objects for chromosome detection, we utilize Hough algorithm to detect whether there exists an eclipse. In addition, this filter also uses Euler number and area to delete false candidates.

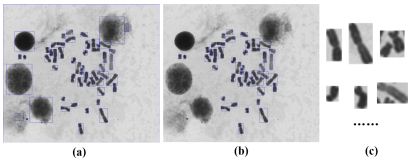

Fig. 2. Chromosome candidates detection, where (a) is the ER detection result, (b) are the chromosome candidates after filters and (c) are the samples of chromosome candidates.

2) Filter using intensity distribution: Inspired by the fact that chromosomes have uniform intensity values, we propose to discard the chromosome candidates which have high variation in intensity values. Supposed that ER with chromosome inside should contain chromosome and background, we first perform histogram operation on intensity values, then adopt mean of the maximal and second-maximal values to calculate the intensity variance V_i of ER by

$$V_i = \frac{n_c \cdot \sum\limits_{j \in e_{i,c}} (I(j) - A_{i,c})^2 + n_b \cdot \sum\limits_{j \in e_{i,b}} (I(j) - A_{i,b})^2}{n_c + n_b} \quad (1)$$

where $e_{i,c}$ and $e_{i,b}$ represent the chromosome and background of e_i respectively, n is the number of pixels and M represents the average value. The sample results of ER and filtering are illustrated in Fig. 2, where we could notice the touching and overlap chromosome candidates emerge as clusters.

2.2. Chromosome Clusters Segmentation

In this subsection, we aim to segment overlap and touching chromosome candidates into individual ones.

Following [22], we propose to segment clustered and partially overlapping chromosome candidates with a shape that can be approximated with an eclipse. We present sampled segmentation results in Fig. 3. Due to the adaptive design, we find individual chromosome candidates remain the same

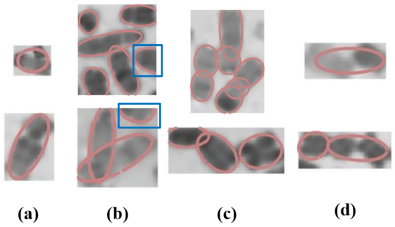

Fig. 3. (a), (b), (c) and (d) represent the segmentation for individual chromosome candidates, successful segmentation for chromosome clusters, failure cases requiring modification tool, and failure cases requiring manual work, respectively. Note blue rectangles indicate eliminated eclipses.

as shown in (a) and overlapping or touching chromosome clusters are successfully segmented as shown in (b). Note that we discard eclipses labeled by blue rectangles, if the ratio between area of eclipse in image and area of the same eclipse is smaller than a pre-defined threshold. We also find some failure cases due to the complexity of touching and overlapping as shown in (c) and (d). For failure cases in (c) caused by over segmentation, we develop an intuitive modification tool, which enables clicks among over-segment parts to form as an individual chromosome. As for cases in (d), we regard it as total failure ones and recommend doctors to segment by hand. After segmentation, we could achieve a set of chromosomes to be classified.

2.3. Chromosome Classification with Data Augmentation

Data augmentation with GAN could complement and complete the data manifold, and find better margins between neighboring classes. Essentially, chromosome classification is a typical classification task with multiple category labels and inadequate data. Data augmentation is thus appropriate to effectively help enlarge the original dataset and increase classification accuracy. Given the discriminator S and generator G, the training of the original GAN as shown in Fig. 4, can be explained to minimax the following objective function:

$$\min_G \max_S E_{X\sim P_d(X)}[\log S(X)] + E_{Z\sim P_Z}[\log(1-S(G(Z)))] \quad (2)$$

where x represents real data drawn from distribution P_d, Z is drawn from a prior distribution P_Z (usually normal distribution) and the function $G(Z)$ induces a generator distribution to be utilized for data augmentation. GAN alternatively optimizes S and G using stochastic gradient-based learning. After discrimination on generated data, the optimization order in Eq. 2 can be reversed, causing the minimax formulation to become maximin. The reverse optimization thus force G to map every Z to a single X that is most likely to be classified as true data, leading to mode collapsing problem.

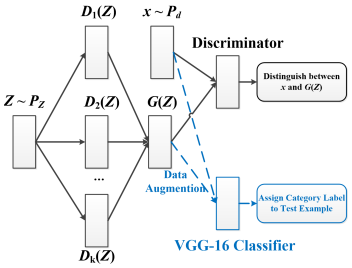

Fig. 4. The architectures of data augmentation with MD-GAN and classification with VGG-16 network.

To avoid mode collapsing problem, the most intuition idea is to employ multiple generators instead of using a single one as in the original GAN, which have been proved extremely effective [20]. However, optimizing multiple generators are complicated and cost large computations. We thus propose to improve this idea by generating with multiple distributions, which is the key idea of the proposed MD-GAN. Due to the property of theoretically fitting any kind of distribution, gaussian mixture model M is adopted to construct the distribution generator D as follows:

$$D_j(Z_j) = P_Z(Z_j) + M_j(Z_j), j = 1...K \quad (3)$$

where K and j are the number of index of distribution generators, P_Z refers to the normal distribution, and Z_j represents the vectors of random samples values between 0 and 1. Note the size of Z_j fits the size of the chromosome images, which is a determined value after re-scaling to the same size. Gaussian mixture distribution M_j is defined as:

$$M_j(Z_j) = \Sigma_{k=1}^{n_j} \frac{1}{n_j} \cdot \phi(Z_j; \mu_{j,k}, \sigma_{j,k}) \quad (4)$$

where ϕ represents the gaussian distribution, n_j is the preset number of formed gaussian distributions, $\mu_{j,k}$ and $\sigma_{j,k}$ are mean and variance parameters for the kth gaussian distribution. We believe more distribution generators would lead to more diverse generation results. However, such construction requires huge computation. We thus need keep balance between computation and generation diversity. By experiments, we set both K and n_j as 8 for each class of chromosome. Note that μ and σ are randomly sampled from normal distribution but keep in a reasonable range.

After data augmentation, we propose to fine-tune the VGG-16 network for classification based on training sets of real data and generated data, which is presented as blue in Fig. 4. Note that we keep earlier layers fixed and only fine-tune some higher-level portion of the network. This is motivated by the truth that earlier features of a CNN contain more generic features that should be useful to many tasks and prevent overfitting, while later layers should become progressively more specific to chromosome classification by learning information from training set.

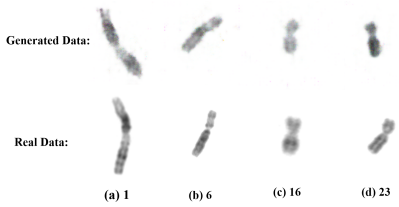

Fig. 5. The comparison results of generated chromosomes and real chromosomes. Note that numbers below represent the class labels corresponding to the chromosome images.

3. EXPERIMENTS

To conduct experiments, we have collected 120 cell images and chromosome images of 119 persons (5474 labeled chromosome images in total) from a hospital. We randomly divided chromosome images into a group of 4600 and 874 for training and test sets, respectively. Essentially, the size of real data is relatively a small number of training samples for classification of 24 classes using deep learning. We thus utilize MD-GAN to generate more training samples. Note that we generate all classes of chromosome images using only one specific MD-GAN. By comparison between generated chromosomes and real ones as shown in Fig. 5, we could find generated samples maintain acceptable visual fidelity meanwhile share diverse modes.

We first compare the detection accuracy of the proposed method with detection using Otsu binarization method. Due to the uneven staining of cell images and the adoption of a global threshold for binarization, we find Otsu method fails in some cases of chromosome detection. Therefore, the detection accuracy and recall achieved by the proposed method is 95.9% and 94.8%, while Otsu method achieves 86.3% and 87.2% for detection accuracy and recall, respectively. For segmentation, we find the failure cases requiring manual work takes only 2.6% of the total chromosome images, which proves the efficiency and ensures the automatical ability of the proposed method.

Table.1 gives the detailed statistics of classification results on the collected chromosome images. We implement Multi-Layer Perceptron (MLP) [23] with 2 and 5 layers as two baselines for comparative study. To show the performance improvement of data augmentation using MD-GAN, we design experiments that we first utilize MD-GAN to generate chromosome images of several persons, and then fuse generated and real chromosomes for classification. Such experiments are expressed as CNN+xGAN in Table. 1, where x is the number of persons. Note that P^x and P^T refer to the precision of classification for one specific class of chromosomes and for all chromosomes, respectively.

From Table.1, we can notice deep neural network achieves much higher precision than two comparative methods, which proves the powerful distinguish ability of deep neural network. By utilizing data augmentation for training, we find the largest improvement in total precision 4.6% achieved by CNN+50GAN comparing with the original CNN. However, more generated examples not always help improve classification accuracy, which could be proved by decrement of precision when comparing among CNN+50GAN, CNN+150GAN and CNN+250GAN. For classification on a specific class of chromosomes, we could view the decrement of precision as well. Note that the inconsistent precision of the fourth chromosome could be explained by the small size of testing examples. Above all, we could draw a conclusion that the size of generated data should be fit with size of original data. If not, the generated data will be noises to confuse the classifier. In our case of chromosome classification, the most suitable size of generated chromosomes is 50, which is nearly half of the size of the original training set, *i.e.*, 100. Combined with the above quantitative results, it is convincing that our MD-GAN bring benefits for both training stability and mode variety without the loss of sample quality.

Method	$P^4(\%)$	$P^{18}(\%)$	$P^{22}(\%)$	$P^T(\%)$
CNN	68.4	60.0	60.0	58.9
CNN+50GAN	69.6	**72.0**	**62.5**	**63.5**
CNN+150GAN	**86.7**	70.8	53.3	62.8
CNN+250GAN	63.6	60.0	50.0	60.5
Two Layer MLP	58.3	54.2	52.9	51.3
Five Layer MLP	62.1	55.3	53.9	53.1

Table 1. Performance of chromosome classification on the collected dataset.

4. CONCLUSION

We propose an end-to-end chromosome karyotyping method by utilizing data generated by MD-GAN for classification. Our further work includes utilizing MD-GAN for other applications, such as emotion classification and person identification.

Acknowledgment

This work was supported by the Natural Science Foundation of China under Grant 61702160, Grant 61370091, Grant 61672273, the Fundamental Research Funds for the Central Universities under Grant 2016B14114, the Science Foundation of JiangSu under Grant BK20170892, the Science Foundation for Distinguished Young Scholars of Jiangsu under Grant BK20160021 and the open Project of the National Key Lab for Novel Software Technology in NJU under Grant KFKT2017B05.

5. REFERENCES

[1] A. P. Britto and G. Ravindran, "A review of cytogenetics and its automation," *Journal of Medical Sciences*, vol. 7, pp. 1–18, 2007.

[2] X. Wang, B. Zheng, S. Li, J. J. Mulvihill, M. C. Wood, and H. Liu, "Automated classification of metaphase chromosomes: Optimization of an adaptive computerized scheme," *Journal of Biomedical Informatics*, vol. 42, no. 1, pp. 22–31, 2009.

[3] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[4] L. Ji, "Fully automatic chromosome segmentation," *Cytometry Part A*, vol. 17, no. 3, pp. 196–208, 1994.

[5] Y. Wu, W. Wang, S. Palaiahnakote, and T. Lu, "A robust symmetry-based method for scene/video text detection through neural network," in *Proceedings of IAPR International Conference on Document Analysis and Recognition*, 2017, pp. 1249–1254.

[6] S. Xia, W. Xu, and Y. Shen, "Two intelligent algorithms applied to automatic chromosome incision," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003, pp. 697–700.

[7] C. Urdiales, A. Bandera, F. Arrebola, and F. S. Hernández, "A curvature-based multiresolution automatic karyotyping system," *Mach. Vis. Appl.*, vol. 14, no. 3, pp. 145–156, 2003.

[8] R. M Nair, R. Remya, and K. Sabeena, "Karyotyping techniques of chromosomes: a survey," *Int J Comput Trends Technol*, vol. 22, no. 1, 2015.

[9] F. Abid and L. Hamami, "A survey of neural network based automated systems for human chromosome classification," *Artificial Intelligence Review*, vol. 49, no. 1, pp. 41–56, 2018.

[10] L. Wei, Y. Wu, W. Wang, and T. Lu, "A novel 3d human action recognition framework for video content analysis," in *Proceeding of International Conference on MultiMedia Modeling*, 2018, pp. 42–53.

[11] H. Li, J. Liu, G. Zhang, Y. Gao, and Y. Wu, "Multi-glimpse LSTM with color-depth feature fusion for human detection," *CoRR*, vol. abs/1711.01062, 2017.

[12] Swati, G. Gupta, M. Yadav, M. Sharma, and L. Vig, "Siamese networks for chromosome classification," in *Proceedings of ICCV Workshops*, 2017, pp. 72–81.

[13] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proceedings of NIPS*, 2016, pp. 3630–3638.

[14] X. Zhu, Y. Liu, and Z. Qin, "Data augmentation in classification using gan," *arXiv preprint arXiv:1711.00648*, 2017.

[15] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, Da. W. Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of NIPS*, 2014, pp. 2672–2680.

[16] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *Proceedings of CVPR*, 2017, pp. 3754–3762.

[17] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein, "Unrolled generative adversarial networks," *CoRR*, vol. abs/1611.02163, 2016.

[18] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Proceedings of NIPS*, 2016, pp. 2226–2234.

[19] T. Che, Y. Li, A. Paul Jacob, Y. Bengio, and W. Li, "Mode regularized generative adversarial networks," *CoRR*, vol. abs/1612.02136, 2016.

[20] Q. Hoang, T. D. Nguyen, T. Le, and D. Phung, "Multi-generator gernerative adversarial nets," *arXiv preprint arXiv:1708.02556*, 2017.

[21] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vision Comput.*, vol. 22, no. 10, pp. 761–767, 2004.

[22] S. Zafari, T. Eerola, J. Sampo, H. Kälviäinen, and H. Haario, "Segmentation of overlapping elliptical objects in silhouette images," *IEEE Trans. Image Processing*, vol. 24, no. 12, pp. 5942–5952, 2015.

[23] R. Collobert and S. Bengio, "Links between perceptrons, mlps and svms," in *Proceedings of ICML*, 2004.