DeepACE: Automated Chromosome Enumeration in Metaphase Cell Images Using Deep Convolutional Neural Networks

Li Xiao1*, Chunlong Luo1,3*, Yufan Luo1,3, Tianqi Yu2, Chan Tian2, Jie Qiao2, and Yi Zhao1

1 Key Laboratory of Intelligent Information Processing, Advanced Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
xiaoli@ict.ac.cn

2 Peking University Third Hospital, Beijing, China
tianchan_cdc@126.com, jie.qiao@263.net

3 School of Computer and Control Engineering, University of Chinese Academy of Sciences (UCAS), Beijing, China

Abstract. Chromosome enumeration is an important but tedious procedure in karyotyping analysis. In this paper, to automate the enumeration process, we developed a chromosome enumeration framework, DeepACE, based on the region based object detection scheme. Firstly, the ability of region proposal network is enhanced by a newly proposed Hard Negative Anchors Sampling to extract unapparent but important information about highly confusing partial chromosomes. Next, to alleviate serious occlusion problems, we novelly introduced a weakly-supervised mechanism by adding a Template Module into classification branch to heuristically separate overlapped chromosomes. The template features are further incorporated into the NMS procedure to further improve the detection of overlapping chromosomes. In the newly collected clinical dataset, the proposed method outperform all the previous method, yielding an mAP with respect to chromosomes as 99.45, and the error rate is about 2.4%.

1 Introduction

Karyotyping is a cytogenetic experiment method used to help cytologists to observe the structures and features of chromosomes presented on metaphase images [9]. Cytologists firstly need to pay attention to the problem of numerical abnormalities of chromosomes that may result in some genetic diseases, such as Down syndrome [8]. Counting chromosomes is performed manually now on at least 20 images per patient and needs 50-100 images more when chromosome mosaicism is explored. Considering that each human cell normally contains 46

* equal contribution
chromosomes, it is tedious and time consuming, therefore it is an urgent need to develop a computer-aided system for chromosomes enumeration.

However, despite that there are some methods have been developed to solve classification [7,10] and segmentation [2,6] problems of chromosomes, very few of the researches have tried to develop automated chromosomes enumeration method. Gajendran et.al [4] presented a study by combining a variety of pre-processing methods and counting algorithm based on topological analysis, but the error rate is high. Furthermore, some segmentation method may solve the problem indirectly such as Arora et.al [2] and Minaee et.al [6], but they only focused on segmenting touching or overlapping chromosomes, and the accuracy is not high enough. The challenges of chromosome enumeration mostly rely on two aspects: chromosomes in metaphase images usually contain severe occlusion and cross overlapping problem (Fig.1(a)); some partial chromosomes or two chromosomes connected head to head are similar to a whole chromosome (Fig.1(b,c)). Above two problems are often occurred simultaneously(Fig.1(d)), which make it difficult to detect all the chromosome objects accurately.

![Fig. 1. The green boxes are the ground truth bounding boxes of the chromosomes. (a) shows occlusion and cross overlapping problem, two ground truth boxes are very close to each other. (b)(c) shows self-similarity problem, the three connected chromosomes are likely to be classified as one chromosome and the deformed chromosomes are likely to be classified as two chromosomes connected to each other. (d) shows complex situation.](image)

In this paper, following region based objection detection scheme, we proposed a deep learning algorithm to address chromosomes enumeration directly on the entire G-band metaphase image. We firstly introduced a Hard Negative Anchors Sampling method on RPN [11] to learn more information about highly confusing partial chromosomes to solve self-similarity problem. Secondly, in classification branch, we proposed a weakly-supervised Template Module to heuristically separate touching and overlapping chromosomes. The features generated from the Template Module is further used to guide NMS procedure in order to avoid over deletion of overlapped chromosomes. Our model is trained and validated on a newly collected dataset with thousands of labeled metaphase images from cytogenetic laboratory, and achieves higher performances than previous studies on 47.6% of WCR, 2.389% of AER, 99.45% of mAP and 98.81% of F1-Score.
2 Method

As shown in Fig.2, the proposed framework consists of three main parts: (1) candidate chromosomes detection using RPN in which a Hard Negative Anchors Sampling procedure is proposed (Section 2.1); (2) an isolated classification branch with weakly-supervised Template Module (Section 2.2) and Feature-guided NMS (Section 2.3); (3) a parallel bounding box regression branch with $1 \times 1$ kernel for keeping identical feature map sizes.

2.1 Hard Negative Anchors Sampling in Region Proposal Network

The region-based object detection models such as Faster R-CNN [11] firstly introduce a region proposal network (RPN) to generate candidate proposals. Typically, RPN only focus on binary classification of integrated objects (eg. IoU $\geq 0.7$) and background (eg. IoU $< 0.3$), and the selected proposals are passed to Fast R-CNN for further fine classification and regression in which partial objects (eg. $0.3 \leq$ IoU $< 0.7$) are taken care of. However, unlike natural objects, chromosomes usually have various length and similar banding patterns, which confuse the network to discriminate partial and whole chromosomes, namely self-similarity problem. Meanwhile, the risk of irreversibly losing information in Fast R-CNN, such as cropping features and RoI Pooling, also make the network hard to distinguish partial chromosomes. To this end, we proposed a novel hard negative anchors sampling method during the RPN sampling procedure to better identify partial chromosomes and solve the self-similarity problem.

![Fig. 2. The framework of DeepACE: (a) shows Hard Negative Anchors Sampling procedure in the first stage. (b) illustrates the classification branch containing Template Module and Feature guided NMS. (c) represents the regression branch. The details of template module is depicted in Fig.3.](image)

For clarity, we regard those anchors that have the IoU in $[0.3, 0.7]$ as hard negative anchors and the original negative anchors with an IoU $< 0.3$ are named as easy negative anchors. Considering that RPN suffers from severe
inter-class imbalance (positive : negative≈ 1 : 2000) and intra-class imbalance (hard negative : easy negative≈ 1 : 25), a new Hard Negative Anchors Sampling method inspired by stratified sampling is then proposed. As shown in Fig.2(a), we divide all anchors into positive, hard negative and easy negative according to IoU overlap with ground truth box. We use mini-batches of size R=256 for training RPN and take 25% of the anchors from positive. The half of remaining are uniformly sampled from hard negatives (37.5%), and the rest are sampled from easy negatives (37.5%). Finally, positive anchors are labeled with a foreground object class, both hard and easy negative anchors are labeled as background, the loss function is the same as the original RPN. In this way, feature maps generated by RPN is enhanced by hard negative anchors informations, and the following stage is improved by these features.

### 2.2 Template Module for Disentangling Occlusion Chromosomes

Touching and overlapping chromosomes bring severe intra-class occlusion, in which network is unable to localize and classify the chromosomes correctly. To alleviate this problem, we insert a weakly-supervised Template Module into the classification branch to heuristically separate the touching or overlapping chromosomes. Specifically, although chromosomes are usually displayed with bending or deformation in metaphase images, they can be summarized into some regular schemes. Therefore, it is reasonable to introduce several general template masks to represent patterns of chromosomes. When two or more chromosomes are overlapped together within a selected proposal, particular chromosome is able to be extracted by the corresponding template mask, and thus facilitate the separation of overlapping chromosomes. The implementation details of the template module can be summarized in Fig.3.

![Fig. 3. The illustration of classification branch of DeepACE: (a) shows the whole classification branch. (b) illustrates the structure of weakly-supervised Template Module. (c) shows the Template Mask which select the most relevant template of the chromosome and output their features. (d) presents Template Residual Block.](image-url)
The design of template module is essential to influence the performance. Chromosomes are usually displayed as slender strips in metaphase images. Therefore, peak values located at the central part of feature map along the diagonal or horizontal or vertical direction, which leads to $T_D$, $T_{TD}$, $T_H$ and $T_V$. Besides, a circle-like template mask $T_C$ is introduced since there are also a few of seriously bending chromosomes as shown Fig.1(c). The feature map size is $7 \times 7$, we introduce $ID_{row} \in \{0, 1, 2, 3, 4, 5, 6\}$ and $ID_{col} \in \{0, 1, 2, 3, 4, 5, 6\}$ to indicate the positions of feature map, all the five template masks are designed as constant tensors with Gaussian distribution, where $x_{row} = ID_{row} - 3$, $y_{col} = ID_{col} - 3$:

$$T_D(ID_{row}, ID_{col}) = e^{-\frac{(x_{row} - y_{col})^2}{3}}, \quad T_H(ID_{row}, ID_{col}) = e^{-\frac{x_{row}^2}{3}}$$

$$T_{TD}(ID_{row}, ID_{col}) = e^{-\frac{(x_{row} + y_{col})^2}{3}}, \quad T_V(ID_{row}, ID_{col}) = e^{-\frac{x_{row}^2}{3}}$$

$$T_C(ID_{row}, ID_{col}) = e^{-\frac{|x_{row}^2 + y_{col}^2 - 5|}{3}} \quad (1)$$

After RoI Pooling layer, feature maps with shape of $7 \times 7 \times 512$ are sent into template module and a shortcut pathway at the same time. The shortcut pathway is composed of one ReLU unit and two $3 \times 3$ convolutional kernels with strides equal to 2 and 1. Meanwhile, as shown in Fig.3(b), the template module is composed of template mask and template residual block. Based on five template masks, the template mask component (as shown in Fig.3(c)) firstly extract features of specific locations followed by global average pooling. We select the $1 \times 1 \times 512$ tensor which has the maximum mean value in the five tensors as the template feature. The template residual block (as shown in Fig.3(d)) is composed of convolutional layers with ReLU and an identify mapping. Finally, the outputs of template module and shortcut pathway are combined together followed by ReLU layer and one 32-D fully connected layer to obtain the binary classification score.

We also design a weakly-supervised mechanism to guide the template feature selection procedure during training, in which the output of template module for ground truth is introduced as the label. The newly designed pull loss $L_{Pull}$ is composed of feature pull loss $L_{Pull_f}$ and mean pull loss $L_{Pull_m}$. The feature pull loss $L_{Pull_f}$ is to minimize the template feature distance between a proposal and its ground truth if they are fall into the same template mask, and maximize the distance if they are fall into different template mask. The square of L2 norm of the ground truth template feature is used for normalization, the definition is as:

$$L_{Pull_f} = \left\{ \begin{array}{ll}
\frac{\|f_g - f_p\|^2}{\|f_g\|^2}, & \text{if } T_g = T_p \\
\max(0, 1 - \frac{\|f_g - f_p\|^2}{\|f_g\|^2}), & \text{if } T_g \neq T_p
\end{array} \right. \quad (2)$$

where $f_p$ and $f_g$ represent template feature vector of proposals and corresponding ground truth. Both $T_p$ and $T_g$ come from $\{T_D, T_{TD}, T_H, T_V, T_C\}$ which represent template masks of selected proposals and corresponding ground truth. For training, we apply the feature pull loss $L_{Pull_f}$ at positive candidate proposals.
The mean pull loss $L_{Pull_m}$ provides a penalty when proposals and corresponding ground truths are fell into different template masks. The $L_{Pull_m}$ is defined as:

$$L_{Pull_m} = \sum_i (\bar{f}_g^i - \bar{f}_p^i)^2,$$

where $\bar{f}_g$ and $\bar{f}_p$ represent the mean value of template features, respectively. Similar to feature pull loss, we only apply the mean pull loss $L_{Pull_m}$ at the positive candidate proposals. The total pull loss $L_{Pull}$ is obtained as follows:

$$L_{Pull} = L_{Pull_f} + L_{Pull_m}$$

2.3 Feature-Guided NMS

In post-processing stage, IoU based algorithms like Non Maximum Suppression(NMS) [5] and Soft-NMS [3] are widely used in recent years. They suppress redundancy according to IoU metric, in which highly overlapped predicted bounding boxes are removed directly or inhibited through decaying its detection scores. However, over deletion still exists when severe occlusion occur, which is frequently happen between chromosomes. Thus, we propose a Feature-Guided NMS based on Soft-NMS, which introduce template features to optimize the score decay function. The basic idea is that if two bounding boxes are fell into different template masks or their template features are far away, they should represents two different chromosomes. Therefore, in Feature-Guided NMS, we compute the normalized distance $d$ between template features and assign a threshold value $\Delta$ (set as 0.5). We will lighter decay score if $d > \Delta$ and heavier decay score if $d < \Delta$. The overview of the algorithm is summarized in Algorithm 1.

3 Experiments Results

Data Set and Evaluation Metrics 1375 metaphase images with pixels of $1600 \times 1200$ containing 63026 objects were collected from the Peking University Third Hospital. All metaphase images are labeled by professional cytologists and are split into 3(training):1(validation):1(testing). Besides mean average precision (mAP) and F1-Score, we also introduced the Whole Correct Ratio (WCR) and Average Error Ratio (AER) as the metrics to evaluate the ensemble performance which has more clinical meanings. We called images that all chromosomes are correctly detected as all right images, and WCR is defined as the percentage of all right images in the whole testing set. We regarded a predicted bounding box as a false positive if it does not have an IoU greater than a threshold with any ground truth (0.5 in this work) or it has the max IoU with a ground truth that has already been detected. A ground truth that is not detected by any bounding box is regarded as a false negative. The AER is defined as the fraction of sum of false positives and false negatives divided by the number of ground truth. For fair comparisons, we perform the ablation study on validation set and report final results on the testing set.
Algorithm 1 Feature-Guided Non Maximum Suppression

Input:
The list of initial detection boxes $B = \{b_1, \ldots, b_N\}$;
The list of corresponding detection scores $S = \{s_1, \ldots, s_N\}$;
The list of corresponding template features $F = \{f_1, \ldots, f_N\}$;

Output:
The list of detection boxes with new order $D'$;
The list of corresponding detection scores which are decayed by function $S'$

1: Initialize $D' = \{\}$;
2: while $B \neq \{\}$ do
3: Sort all the detection boxes $B$ by scores $S$ in descending order, mark the first candidate as $b_{\text{max}}$, corresponding score $s_{\text{max}}$ and template feature $f_{\text{max}}$
4: Append the $b_{\text{max}}$ into $D'$ and pop it from $B$
5: Append the $s_{\text{max}}$ into $S'$ and pop it from $S$
6: for $b_i \in B$ do
7: Measure the normalized distance $d = \frac{\|f_{\text{max}} - f_i\|}{\|f_{\text{max}}\|}$
8: Compute new detection score $s_i$ by: $s_i = s_i e^{-\frac{\text{iou}(b_{\text{max}}, b_i)(2-\Delta+d)}{d}}$
9: end for
10: end while
11: return $D'$ and $S'$

Implementation Details During the pre-processing step, we perform one vertical and one horizontal flipping to augment data, and subtract the mean along each channel for zero centering. The $\sigma$ of Feature-Guided NMS are set as 0.2. Network is implemented on TensorFlow framework [1]. We trained the network for 100k iterations, with initialize learning rate set as 0.001 and decay by a factor of 10 at 60k iterations. Stochastic Gradient Descent (SGD) is adopted to optimize our network on a Nvidia Titan Xp GPU with momentum= 0.9.

3.1 Comparison

We verify the effectiveness of our proposed method by comparing with the chromosomes enumeration method proposed in [4], which is based on digital image analysis and evaluated by Cluster-Based Error criterion. The criterion only concerns the error of the chromosome number caused by cutting entire chromosome incorrectly or connecting individual chromosomes incorrectly. So, the criterion is much looser than AER criterion in this paper. Besides, some segmentation methods may also contain some potential connection with chromosomes enumeration task, which report detection accuracy about touching chromosomes ACC_T and overlapping chromosomes ACC_O, we list some of them. Nevertheless, as shown in Table 1, our method still greatly outperform the previous method [4] and achieve an AER of 2.389% and WCR of 47.63%. Furthermore, it is worth mention that although DeepACE does not involve any pretraining, it still significantly outperforms the Faster RCNN which has been pretrained on ImageNet.
Table 1. The comparison of chromosomes counting methods. Avg-CB-Error means Average Cluster-Based Error, $ACC_T$ means detection accuracy for Touching or partial overlapping chromosomes, $ACC_O$ means detection accuracy for overlapping chromosomes.

| Method            | Avg-CB-Error | $ACC_T$ | $ACC_O$ | WCR  | AER  | mAP  | F1-Score |
|-------------------|--------------|---------|---------|------|------|------|----------|
| Gajendran et. al [4] | 6.4%         | -       | -       | -    | -    | -    | -        |
| Minaee et. al [6]  | -            | 91.9%   | -       | -    | -    | -    | -        |
| Arora et. al [2]   | -            | 96.7%   | 81%     | -    | -    | -    | -        |
| Faster R-CNN       | -            | -       | -       | 39.64%| 2.44%| 99.03%| 98.79%   |
| DeepACE            | -            | -       | -       | 47.63%| 2.39%| 99.45%| 98.81%   |

3.2 Ablation Study

**Hard Negative Anchors Sampling** To verify HNAS’s contribution to performance, we train another network without HNAS. Table 2(a) shows that HNAS improves WCR greatly over baseline by 29.1%, AER by 2.7%, mAP by 2.18% and F1-Score by 1.37%. This is expected because self-similarity problem may happen on most of the chromosomes. Notice that the Faster R-CNN performs much better than DeepACE(w/o HNAS), this may because pretraining makes the network learn features more accurately thus the network is able to identify those subtle features which is essential to solve self-similarity problem.

**Template Module with Feature-Guided NMS** we use features came from the second conv layer of shortcut pathway as an alternative of template feature for experiment of Table 2(b). Results show that both Template Module and Feature-Guided NMS improves the performances significantly. Table 2(c) also represents that discriminative template feature can optimize the nms procedure, which improves WCR by 2.55%, AER by 0.31%, mAP by 0.41% and F1-Score by 0.15% in condition of high benchmark.

Table 2. Ablation Study on Different Component.

|                | WCR      | AER      | mAP      | F1-Score |
|----------------|----------|----------|----------|----------|
| (a)w/o HNAS   | 15.64%   | 5.03%    | 98.18%   | 97.46%   |
| (b)w/o Template Module | 36.73% | 2.76% | **99.53%** | 98.63%   |
| (c)w/o Feature-Guided NMS | 42.18% | 2.64% | 99.05% | 98.68%   |
| (d)DeepACE     | **44.73%** | **2.33%** | 99.46% | **98.83%** |

4 Conclusion

In this paper, we developed an automated chromosome enumeration algorithm, DeepACE. Hard Negative Anchors Sampling learns more about partial chromo-
somes and Template Module equipped with Feature-Guided NMS use weakly-supervised mechanism to heuristically identify overlapping chromosomes. Experiments on clinical datasets demonstrate its effectiveness. The future plan is to continue develop DeepACE to solve chromosomes detection and segmentation tasks on whole metaphase images.

5 Acknowledgements

This work was supported by the CAS Pioneer Hundred Talents Program (2017-074) to Li Xiao. We thank Professor S. Kevin Zhou, Professor Yang Wu and Professor Hui Li for critical comments during manuscript preparation.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: a system for large-scale machine learning. In: OSDI. vol. 16, pp. 265–283 (2016)
2. Arora, T., Dhir, R.: A novel approach for segmentation of human metaphase chromosome images using region based active contours. International Arab Journal of Information Technology (2016)
3. Bodla, N., Singh, B., Chellappa, R., Davis, L.S.: Soft-nms: improving object detection with one line of code. In: Computer Vision (ICCV), 2017 IEEE International Conference on. pp. 5562–5570. IEEE (2017)
4. Gajendran, V., Rodríguez, J.J.: Chromosome counting via digital image analysis. In: Image Processing, 2004. ICIP’04. 2004 International Conference on. vol. 5, pp. 2929–2932. IEEE (2004)
5. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
6. Minaee, S., Fotouhi, M., Khalaj, B.H.: A geometric approach to fully automatic chromosome segmentation. In: 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). pp. 1–6. IEEE (2014)
7. Munot, M.V.: Development of computerized systems for automated chromosome analysis: Current status and future prospects. International Journal of Advanced Research in Computer Science 9(1) (2018)
8. Patterson, D.: Molecular genetic analysis of down syndrome. Human genetics 126(1), 195–214 (2009)
9. Piper, J.: Automated cytogenetics in the study of mutagenesis and cancer. In: Advances in Mutagenesis Research, pp. 127–153. Springer (1990)
10. Qin, Y., Song, N., Zheng, H., Huang, X., Yang, J., Zhu, Y.M., Yang, G.Z.: Varifocalnet: A chromosome classification approach using deep convolutional networks. arXiv preprint arXiv:1810.05943 (2018)
11. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)