Journal Pre-proof

KaryoXpert: An accurate chromosome segmentation and classification framework for karyotyping analysis without training with manually labeled metaphase-image mask annotations

Siyuan Chen, Kaichuang Zhang, Jingdong Hu, Na Li, Ao Xu, Haoyang Li, Juexiao Zhou, Chao Huang, Yongguo Yu, Xin Gao

PII: S0010-4825(24)00686-3
DOI: https://doi.org/10.1016/j.compbiomed.2024.108601
Reference: CBM 108601

To appear in: *Computers in Biology and Medicine*

Received date : 25 February 2024
Revised date : 8 April 2024
Accepted date : 11 May 2024

Please cite this article as: S. Chen, K. Zhang, J. Hu et al., KaryoXpert: An accurate chromosome segmentation and classification framework for karyotyping analysis without training with manually labeled metaphase-image mask annotations, *Computers in Biology and Medicine* (2024), doi: https://doi.org/10.1016/j.compbiomed.2024.108601.

Computers in Biology and Medicine (2024)

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

KaryoXpert: An Accurate Chromosome Segmentation and Classification Framework for Karyotyping Analysis without Training with Manually Labeled Metaphase-Image Mask Annotations

Siyuan Chen[a,b,*], Kaichuang Zhang[c,*], Jingdong Hu[d], Na Li[d], Ao Xu[d], Haoyang Li[a,b], Juexiao Zhou[a,b], Chao Huang[e], Yongguo Yu[c,**], Xin Gao[a,b,**]

[a]Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia
[b]Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia
[c]Department of Pediatric Endocrinology and Genetic Metabolism, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai Institute for Pediatric Research, Room 801, Science and Education Building, Kongjiang Road 1665, Shanghai, China.
[d]Smiltec (Suzhou) Co., Ltd., Room 401B, Building B6, No. 218 Xinghu Street, Suzhou Industrial Park, Suzhou, Jiangsu, China.
[e]Ningbo Institute of Information Technology Application, Chinese Academy of Sciences (CAS), Ningbo, China

ARTICLE INFO

Article history:
Received 1 May 2013
Received in final form 10 May 2013
Accepted 13 May 2013
Available online 15 May 2013

Communicated by S. Sarkar

2000 MSC: 41A05, 41A10, 65D05, 65D17

Keywords:
Cytogenetics
Automated Karyotyping
Chromosome Segmentation
Chromosome recognition
Deep Learning

ABSTRACT

Automated karyotyping is of great importance for cytogenetic research, as it speeds up the process for cytogeneticists through incorporating AI-driven automated segmentation and classification techniques. Existing frameworks confront two primary issues: Firstly the necessity for instance-level data annotation with either detection bounding boxes or semantic masks for training, and secondly, its poor robustness particularly when confronted with domain shifts. In this work, we first propose an accurate segmentation framework, namely KaryoXpert. This framework leverages the strengths of both morphology algorithms and deep learning models, allowing for efficient training that breaks the limit for the acquirement of manually labeled ground-truth mask annotations. Additionally, we present an accurate classification model based on metric learning, designed to overcome the challenges posed by inter-class similarity and batch effects. Our framework exhibits state-of-the-art performance with exceptional robustness in both chromosome segmentation and classification. The proposed KaryoXpert framework showcases its capacity for instance-level chromosome segmentation even in the absence of annotated data, offering novel insights into the research for automated chromosome segmentation. The proposed method has been successfully deployed to support clinical karyotype diagnosis.

1. Introduction

Cytogenetics is a specialized field devoted to exploring the intricate connection between chromosomal alterations and genetic diseases in humans. It involves the analysis of an individual's chromosomes to identify structural or numerical abnormalities that may be associated with various diseases. It plays a vital role in genetic counseling, prenatal diagnosis, postnatal di-

*contributed equally
** Co-Corresponding author: Yongguo Yu, Xin Gao
E-mail addresses:
yuyongguo@shsmu.edu.cn;
xin.gao@kaust.edu.sa

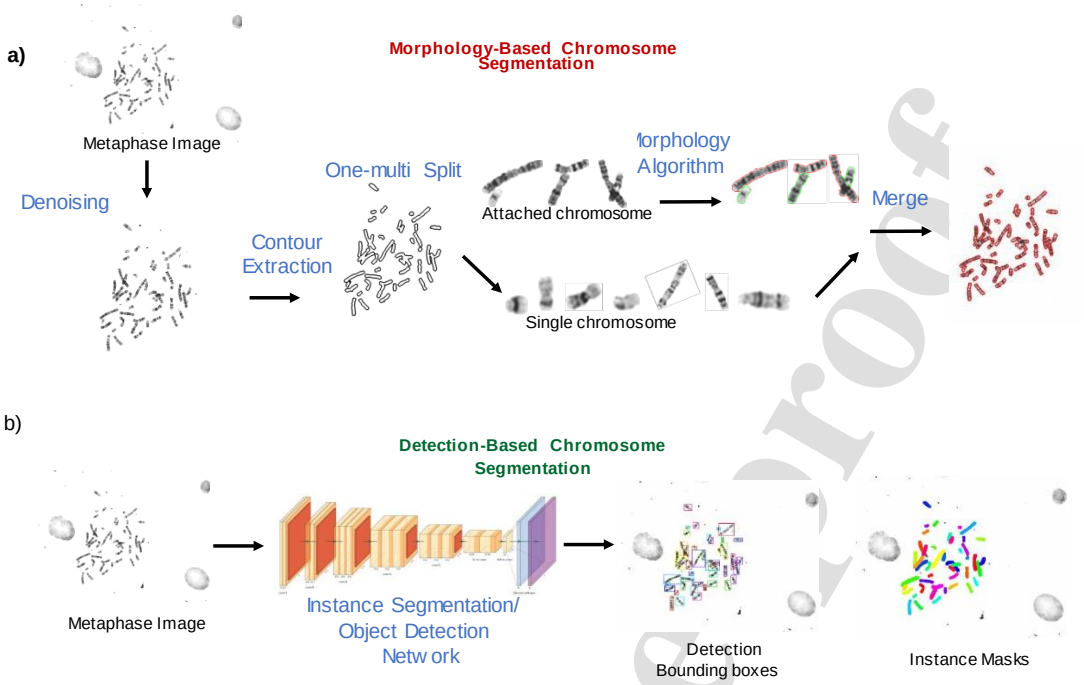

Fig. 1. Brief illustration for morphology-based and detection-based frameworks for chromosome segmentation. (a). Morphology-based chromosome segmentation exhibits robustness in most scenarios but encounters difficulties when dealing with overlapping chromosome segmentation. (b). Detection-based segmentation excels in challenging segmentation tasks but may experience robustness issues when confronted with domain shifts.

agnosis, and genetic research. Some common genetic disorders, including Down syndrome, Turner syndrome, Klinefelter syndrome, and cancers, can be diagnosed and studied through clinical cytogenetics. Karyotyping is one of the fundamental techniques and the **gold standard** in cytogenetic research. It offers insights into the number, size, shape, and structural integrity of an individual's chromosomes, enabling the detection of various genetic anomalies. Following the sequential steps of colchicine treatment, hypotonic treatment, fixation, sectioning, and staining, specimens are then subsequently examined through microscopy. Metaphase images exhibiting distinct chromosome boundaries are systematically enumerated, segmented, and categorized to compile a comprehensive karyotype report. In clinical practice, metaphase chromosomes for noninvasive prenatal testing (NIPT) are primarily obtained from peripheral blood. In this study, our primary focus is on metaphase chromosomes derived from both peripheral blood and amniotic fluid, which are utilized in the majority of prenatal testing.

During clinical studies, cytogeneticists should manually select 20 metaphase images from more than 200 samples for chromosome enumeration. Five of these images will undergo further processing to generate one karyotype report. Despite efforts by software solutions such as Ikaros Rose et al. (2019); Vajen et al. (2022), CytoVision Micci et al. (2001); Yang et al. (2010), and ASI Hiband Fan et al. (2000) to automate the process, multiple challenges still remain in chromosome segmentation and classification, including mask annotation requirements for chromosome segmentation, poor robustness against domain shift, batch effects, and inter-class similarity for chromosome classification.

To tackle these challenges, in this paper, we propose KaryoXpert, an accurate chromosome segmentation and classification framework that efficiently addresses the challenges listed above. The contributions can be summarized as follows:

- Training for KaryoXpert does not require manually labeled metaphase-image-level mask annotations. We introduce a highly effective data simulation method, complemented by a five-way segmentation stream. To the best of our knowledge, this is the first segmentation framework with high accuracy that is applicable without the requirement for manual mask annotation on metaphase images during training.

- KaryoXpert combines the advantages of morphology-based, and deep-learning-based segmentation methods, showcasing strong robustness and high precision across multiple datasets during rigorous statistical benchmarking and clinical evaluation.

- We propose a metric-learning-based classification model in KaryoXpert that eliminates the batch effect and inter-class similarity challenges caused by variant environmental conditions.

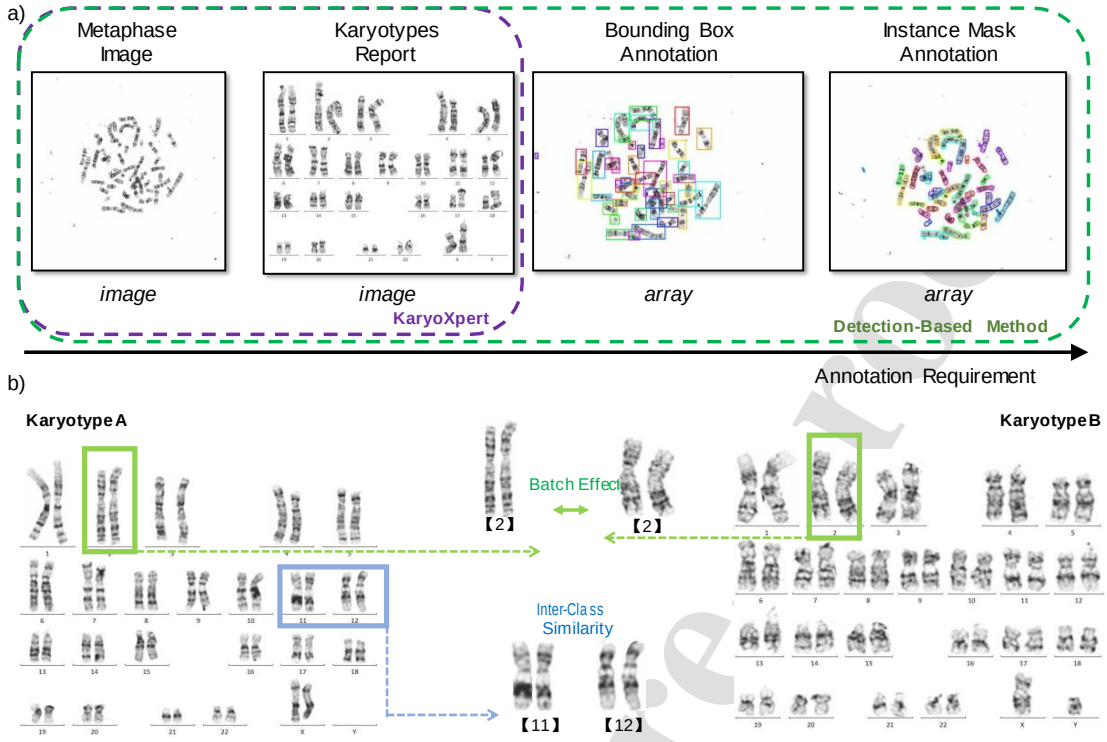

Fig. 2. (a). Data annotation challenges for automated karyotype analysis. The detection-based method requires comprehensive annotation involving instance masks for training, whereas KaryoXpert only relies on karyotype reports as training input. (b). Batch effect and inter-class similarity challenges in the chromosome identification problem.

2. Related Work

2.1. Chromosome Segmentation

For instance-level chromosome segmentation, there are mainly two types of segmentation frameworks, namely morphology-based segmentation, and detection-based segmentation (Fig 1). Each of them possesses distinct strengths and limitations under various scenarios:

In the context of morphology-based chromosome segmentation Devaraj et al. (2022); Minaee et al. (2014); Wu et al. (2020), the initial step involves binarization and contour detection, which serves to distinguish chromosomes from the background. Devaraj et al. (2022) employed curvature analysis techniques to identify concave and convexpoints along the chromosome contours, aiding in the separation of closely located chromosomes. Similarly, Minaee et al. (2014) detected convex hull within chromosome clusters, and applied the sum of distances among total points (SDTP) metric to segregate connected chromosomes. Wu et al. (2020) employed the fast radial symmetry (FRS) transform on the original input images to obtain seed points, and then utilized the ring radius transform (RRT) for overlapping segmentation. The key advantage of morphology-based methods is their independence from the need for ground truth annotations during training. However, they tend to exhibit limitations in scenarios involving densely packed chromosome clusters, characterized by instances where more than two chromosomes are intricately connected and exhibit extensive overlap. In such complex situations, their segmentation often fails.

The recently emerged detection-based chromosome segmentation framework Xiao et al. (2020); Kang et al. (2022); Ren et al. (2015); Xie et al. (2019); Al-Kharraz et al. (2020); Mei et al. (2022); Saleh et al. (2019); Tseng et al. (2023) leveraged a deep neural network as a powerful backbone for object detection, and showed improved detection accuracy on overlapped chromosomes. DeepACEv2 Xiao et al. (2020), building upon Faster R-CNN Ren et al. (2015), employed a Hard Negative Anchors Sampling strategy within its Region Proposal Network (RPN) to enhance detection performance. Tseng et al. (2023) provided an open-source dataset with YOLOv4 Bochkovskiy et al. (2020) as their detection backbone. While accurate detection with bounding boxes does not guarantee instance-level segmentation, since overlapping chromosomes will be framed into the same detection box. Furthermore, leveraging Mask-RCNN He et al. (2017), Xie et al. (2019), and Al-Kharraz et al. (2020) managed to conduct instance-level segmentation on the meta-phase image. Their approaches were trained using extensive datasets comprising thousands of metaphase images, each meticulously annotated with instance-level chromosome masks. Lin et al. (2021) utilized ResNeXt WSL (Weakly Supervised Learning) to distinguish chromosomes from overlapping, touching, and clustered categories, however, this approach did not provide specific executable segmentation solutions. UNet-based approaches Mei et al. (2022); Saleh et al. (2019) are also widely adopted for overlapping chromosome semantic segmentation, while they are not instance-level and only focus on over-

lapping or touching chromosome segmentation. Chromosome segmentation remains a formidable challenge for algorithm developers since obtaining instance-level mask annotations is notably more challenging than acquiring raw metaphase images and karyotype reports (see Fig. 2). Additionally, the limited availability of publicly accessible datasets makes it difficult to comprehensively evaluate the robustness of deep learning models across diverse datasets with notable domain shifts.

2.2. Chromosome Classification

Chromosome classification is another specialized cytogenetic technique. Typically, a healthy human somatic cell possesses a total of 46 chromosomes, comprising 22 pairs of autosomes and 1 pair of sex chromosomes (either XY or XX). Accurately distinguishing between those 24 classes requires several years of training for proficient operators. Such differentiation is based on chromosome characteristics such as length, centromere position, the ratio of long to short arms, as well as chromosome banding features. Therefore, manual chromosome identification in karyotype analysis is a time-consuming and tedious task that heavily relies on expert knowledge. Over the past few years, there has been a proliferation of computer-assisted methods for chromosome classification, aimed at alleviating the manual identification workload. The proposed classification algorithms are mainly morphology-based approaches Lerner et al. (1995); Ming and Tian (2010); Abid and Hamami (2018); Biyani et al. (2005); Markou et al. (2012) and deep-learning based approaches Xia et al. (2022); Xiao and Luo (2021); Haferlach et al. (2020); Zhang et al. (2021); Wei et al. (2022); Wu et al. (2018); Jindal et al. (2017); Sharma et al. (2017); Vajen et al. (2022); Sharma et al. (2018); Qin et al. (2019); Peng et al. (2021). Band features, gray profiles, and shape profiles were utilized by Ming and Tian (2010) to conduct geometric classification and achieved 85.6 % accuracy in classification. Markou et al. (2012) proposed a support vector machine (SVM) based classifier by extracting the chromosome medial axes. In recent years, with the advances in convolutional neural networks (CNNs) and their applications in various fields, especially computer vision and medical imaging, Deep learning-based approaches have been widely applied to computer-assisted chromosome recognition and demonstrate higher classification accuracy. Xiao and Luo (2021) proposed a method based on a siamese network, incorporating a group inner-adjacency loss for chromosome classification. Qin et al. (2019) and Vajen et al. (2022) emphasized the utilization of prior knowledge to reassign prediction labels derived from convolutional neural networks (CNNs). KaryoNet Xia et al. (2023), on the other hand, enhanced prediction accuracy by leveraging global contextual information through the innovative Masked Feature Interaction Module (MFIM). The above-mentioned studies highlight the critical challenges encountered in automated chromosome segmentation and classification primarily concentrated in the following aspects:

- **Annotation Requirements (Segmentation)**: Training an instance-level segmentation network requires mask annotation on thousands of metaphase images (Fig. 2). Publicly available datasets are rarely accessible due to privacy issues. Whereas, no existing deep-learning-based algorithm can be trained without ground truth annotation for each image.

- **Robustness (Segmentation)**: Deep-learning-based segmentation exhibits limited robustness when transferring to metaphase images captured in diverse environmental conditions, or domain shift. On the other hand, morphology-based segmentation algorithms are more robust, while they demonstrate low precision in scenarios involving overlapping and closely touching chromosomes.

- **Batch Effect (Classification)**: Chromosomes exhibit significant variations in length, banding depth, and thickness from various batches of metaphase images. Such variation can also be caused by taking metaphase images at different stages of cell division, using different types of microscopes under variant environmental conditions (Fig. 2).

- **Inter-class Similarity (Classification)**: Chromosome classification involves a detailed process in which visually similar chromosomes may exhibit only minor local differences that require careful comparison for differentiation.

3. Materials and Methods

Table 1. Data Statistics for 2 Benchmark Datasets

Dataset	Attributes	
Ikaros Dataset (in-House)	Number of Samples	1,500
	Banding	G-Band
	Image Size	1360*1024
	Resolution	63X
	Data Source	Shanghai Xinhua Hospital
	Image Format	TIF
	Detection Annotations	1,500
	Mask Annotations	1,500
	Karyotype Reports	1,500
Tseng Dataset (Public) Tseng et al. (2023)	Number of Samples	5,000
	Banding	G-Band
	Image Size	Various
	Resolution	13X
	Data Source	Taichung Veterans General Hospital
	Image Format	JPG
	Detection Annotations	5,000
	Mask Annotations	N/A
	Karyotype Reports	N/A

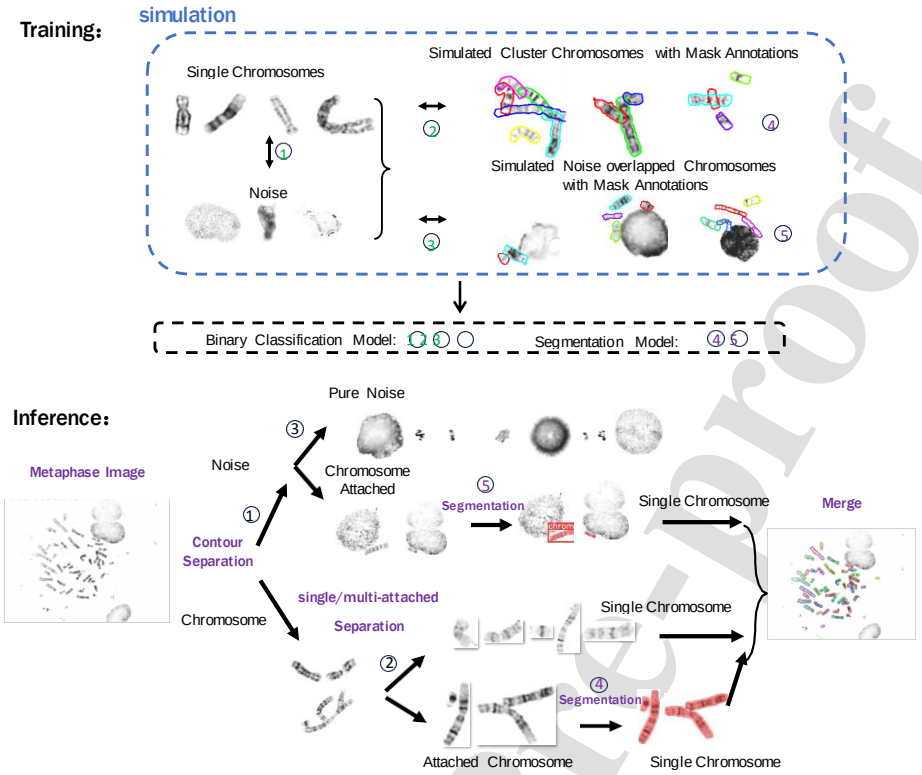

Fig. 3. The detailed structure for the five-way segmentation stream and the data simulation process of KaryoXpert.

3.2. *Datasets and Imaging Protocol*

We collected 2 datasets for automated chromosome segmentation and classification. In the Ikaros Dataset, we collected 1,500 G-band metaphase images sourced from 605 patients between 2022 and 2023 in Shanghai Xinhua Hospital, China. These metaphase images were captured by CoolCube 1 camera (MetaSystems, Germany) attached to the AXIO IMAGER Z2 at 63x (Carl Zeiss, Germany), and subsequently exported to the Ikaros system (MetaSystems, Germany). Each metaphase image in this dataset is thoughtfully paired with its corresponding chromosome mask annotation, delineating the boundaries of all 46 chromosomes, and accompanied by a karyotype report with classification labels. Additionally, we have categorized the samples collected into two distinct groups: one from peripheral blood and the other from amniotic fluid.

In the publicly available Tseng Dataset Tseng et al. (2023), 5,000 G-band metaphase images are collected from prenatal chromosomal studies conducted between 2014 and 2021 at the Cytogenetic Laboratory, Department of Women's Medicine, Taichung Veterans General Hospital. However, it is important to note that mask annotations are not available for this dataset. As a result, instance-level segmentation evaluation and chromosome classification assessments cannot be performed on the Tseng dataset, primarily due to the potential presence of overlapping chromosomes within bounding boxes. Detailed dataset statistics can be referred to Table 1. To validate the models' robustness, we conduct the training process on 1,400 samples from the Ikaros dataset. We directly tested the remaining 100 samples from the Ikaros dataset and applied them to 500 test samples of the Tseng dataset.

3.3. *Instance-Level Chromosome Segmentation without trained with manually labeled Mask Annotations*

3.2.1. *Contour Separation*

The initial stage of our proposed method leverages morphology-based contour separation (first step Fig. 3). This process utilizes a watershed transform Vincent and Soille (1991); Beucher (1979) after applying binarization to raw metaphase images with a specified threshold. We then eliminate contours that fall below a certain size threshold. Such an approach ensures the robustness of our model across diverse metaphase image environments. This transformation effectively breaks down the raw metaphase image into manageable components, which are subsequently channeled into the five-way segmentation stream.

3.2.2. *Five-Way Segmentation Stream*

Utilizing a divide-and-conquer methodology, we partition the entire segmentation stream into five trainable models, each amenable to training with pure data simulation (see Fig. 3). Firstly, we employ three binary classifiers for the initial separation of contour elements into four distinct categories: Model ⓪ is dedicated to classifying contours consisting of pure chromosomes (either single or attached) from background noise.

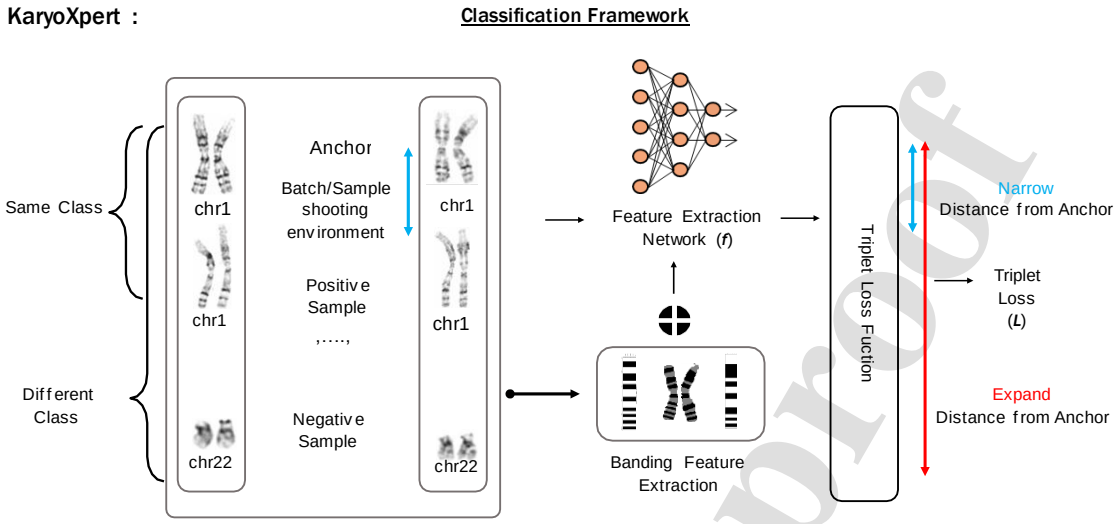

Fig. 4. Detailed Structure of the training process for the Deep Metric Learning framework for chromosome classification.

Model ❷ focuses on distinguishing single chromosomes from cases involving multiple connected chromosomes. Model ❸ is specialized in isolating noise regions that are attached to chromosomes from purely noisy regions. Afterward, Model ❹ functions as a detection-based segmentation network, to further disentangle chromosome clusters into single ones. Meanwhile, Model ❺ is designed with the extraction of chromosomes that are conjoined with noise—a notably challenging scenario in our pipeline that necessitates careful consideration.

3.2.3. *Data Simulation*

The previously mentioned five-way segmentation stream is able to be trained in a pure simulation format (Fig. 3 upper panel), requiring only raw metaphase images and their corresponding karyotype reports. In Model ❶, we manually annotated any instances of noise that appeared in metaphase images, contrasting them with chromosomes extracted from the karyotype reports. For model ❷❸, we employed a "copy and paste" methodology to generate attached chromosomes, varying the number of chromosomes, rotations, and flipping conditions randomly. All images are normalized to 256 × 256 before feeding into the deep-metric learning network. The simulated masks are obtained according to their counters relative to the background. Utilizing the data generated from these previous simulations, we also obtained the corresponding chromosome masks, which were subsequently utilized to train the segmentation networks for Model ❹ and Model ❺. Overall, we extracted approximately 60,000 single chromosome images and 8,000 noise samples from 1,500 karyotypes. Utilizing these, we were able to generate around 100,000 simulated overlapped chromosomes, accompanied with instance masks, for the training of Model ❹ and Model ❺.

3.4. *Deep Metric Learning with Triplet Loss*

Previously, we mentioned that the major challenges for automated chromosome classification lie in data insufficiency, inter-class similarity, and batch effect. Here we propose a chromosome classification method based on deep metric learning that is trained with predefined triplet loss to effectively tackle these challenges.

In Fig. 4, a feature mapping network (f) is firstly designed to learn a mapping from the original input chromosome image to a low-dimensional dense embedding space. The feature mapping network consists of two components: The first part employs a convolutional neural network to extract and characterize features from the input chromosome image, translating spatial information into vectors. The second part conducts banding extraction on the original chromosome image to derive its banding feature vector. These two vectors are then concatenated and ultimately transformed into a 128-dimensional vector space: $x = (x_1, x_2, x_3, \ldots, x_{128})$. Such embedding is further normalized into (0-1) range under Equation (1).

$$x_i^{\wedge} = \frac{x_i - \mu}{\sqrt{(\sigma^2) + \epsilon}}, (i \in [1, 128]) \quad (1)$$

, where μ and σ are the mean and standard deviation of the batch, and ϵ is an arbitrarily small constant added to the denominator to ensure numerical stability. Subsequently, a distance function is formulated to ensure that similar chromosomes exhibit relatively narrow distances from each other, whereas objects of distinct classes exhibit relatively expanded distances. With previously calculated feature vectors, we employ cosine similarity to quantify the distances between pairs of feature vectors. Take vectors $\vec{A}_{1\times128}$, and $\vec{B}_{1\times128}$ as examples. The cosine distance is calculated in Equation 2

$$Cos(A, B) = \frac{\vec{A} \cdot \vec{B}}{|\vec{A}| \cdot |\vec{B}|} \quad (2)$$

, where $\vec{A} \cdot \vec{B}$ is the inner product of the two 128-dimensional vectors, and $|\vec{A}|, |\vec{B}|$ represent their modules.

We then define the triplet with 3 samples: Anchor (x), Positive(x^+), and Negative(x^-) sample. Positive samples are chromosomes of the same category with anchor, although they may originate from different batches. Negative examples consist of chromosomes from different categories. With the triplet, we define the triplet loss function:

$$L(x, x^+, x^-) = max(0, ||f - f^+|| - ||f - f^-|| + m) \quad (3)$$

, where f (feature mapping network), m (margin) is a constant greater than zero. Our ultimate optimization objective is to minimize the distance between x and x^+ while simultaneously maximizing the distance between x and x^-. In our specific case, we set the margin, m, to 0.2. Detailed explanations for triplet loss can be found in Supplementary Materials.

3.5. Implementation Details

3.4.1. Network Training

Model **O-O** are trained with ResNet50He et al. (2016) network architecture, with fully connected layers for binary classification incorporated into the architecture. During the training process, data augmentation techniques such as random rotation, flipping, scaling, and color jittering are systematically applied to diversify the dataset. Model **O** and **O** are trained with the Yolov7 Wang et al. (2023) network architecture with fully synthetic data. We select the Yolo family backbone due to its real-time detection speed and flexibility. This versatility ensures that KaryoXpert can be utilized in a wide range of applications, from cloud-based solutions to personal computer applications. Detailed training parameters can be found in Supplementary Materials.

3.4.2. Benchmarks

For segmentation, all models are trained using the training set of the Ikaros dataset and evaluated against its test set to establish benchmarks. Additionally, to assess their robustness, the models are directly applied to the test subset of the Tseng dataset without further adaptation. Due to concerns related to data privacy and ethical approval, few chromosomal segmentation programs are publicly available for head-to-head comparison. Therefore, we extracted the segmentation backbone for the state-of-the-art (SOTA) algorithm Xie et al. (2019); Tseng et al. (2023) and conducted detailed statistical evaluations. Detection-based frameworks of MaskRCNN He et al. (2017) (Backbone of Xie et al. (2019)), Yolov4 (Backbone of Tseng et al. (2023)), Yolov7 Wang et al. (2023), YOLACT Bolya et al. (2019), and SOLO Wang et al. (2020) are trained under the same training set from the Ikaros dataset with Open MMLab Detection Toolbox Chen et al. (2019).

For classification, we benchmarked with the state-of-the-art (SOTA) chromosome classification algorithm KaryoNet Xia et al. (2023), and classical classification backbones, such as ResNet50 He et al. (2016), EfficientNetV2 Tan and Le (2021), ConvNeXt Liu et al. (2022), and Vision Transformer (ViT)Dosovitskiy et al. (2020). All models are fine-tuned under the same chromosome identification dataset from the Ikaros dataset with torchvision v0.15 Marcel and Rodriguez (2010) pre-trained weights.

4. Experiments

4.1. Evaluation Metrics

Segmentation: In the evaluation of chromosome segmentation at the instance level, we utilize the Average Precision (AP) metric across various IoU thresholds, specifically AP@0.5 and mAP(AP@0.5:0.95), with reference to both ground truth segmentation masks (Seg) and bounding boxes (BBox). This evaluation is conducted using the COCO instance segmentation API Lin et al. (2014). It offers a comprehensive assessment of a model's ability to detect objects across different scales within an image. IoU, or Intersection over Union, serves as a measure to estimate how well the algorithm's predictions align with the actual regions of interest for objects in an image, as defined in the Equation (4):

$$IntersectionoverUnion(IoU) = \frac{IntersectionArea}{UnionArea} \quad (4)$$

Then we are able to calculate $Precision_{seg}$ and $Recall_{seg}$ with Equation 5 based on various IoU thresholds:

$$\begin{aligned} Recall_{seg} &= \frac{TP}{TP + FN} \\ Precision_{seg} &= \frac{TP}{TP + FP} \end{aligned} \quad (5)$$

where TP, FN, and FP stand for true positives, false negatives, and false positives under a certain IoU threshold $\in [0.5, 0.95]$. Then, AP can be calculated from 11 points from the Precision-Recall curve from levels of [0, 0.1, 0.2,...,1] under Equation 6:

$$AP = \frac{1}{11} \sum_{r \in 0, 0.1, ...1} Precision(Recall_i) \quad (6)$$

We also report the Average Recall (AR@IoU0.5:0.95) value for each benchmark, which provides an overall assessment of the model's ability to recall objects across various scales.

Enumeration: For stability analysis, we conduct metaphase image chromosome enumeration across two datasets. Different from instance-level segmentation, the evaluation criteria are Accuracy (Acc) and Average Error Ratio (AER), which are defined under Equation 7:

$$\begin{aligned} Acc &= \frac{\sum_k TP_k}{\sum_k (TP_k + FP_k + FN_k)} \\ AER &= \frac{\sum_k (FP_k + FN_k)}{\sum_k (TP_k + FN_k)} \end{aligned} \quad (7)$$

, where True Positive (TP_k) is defined as the instance when a predicted bounding box is correctly matched to a ground truth

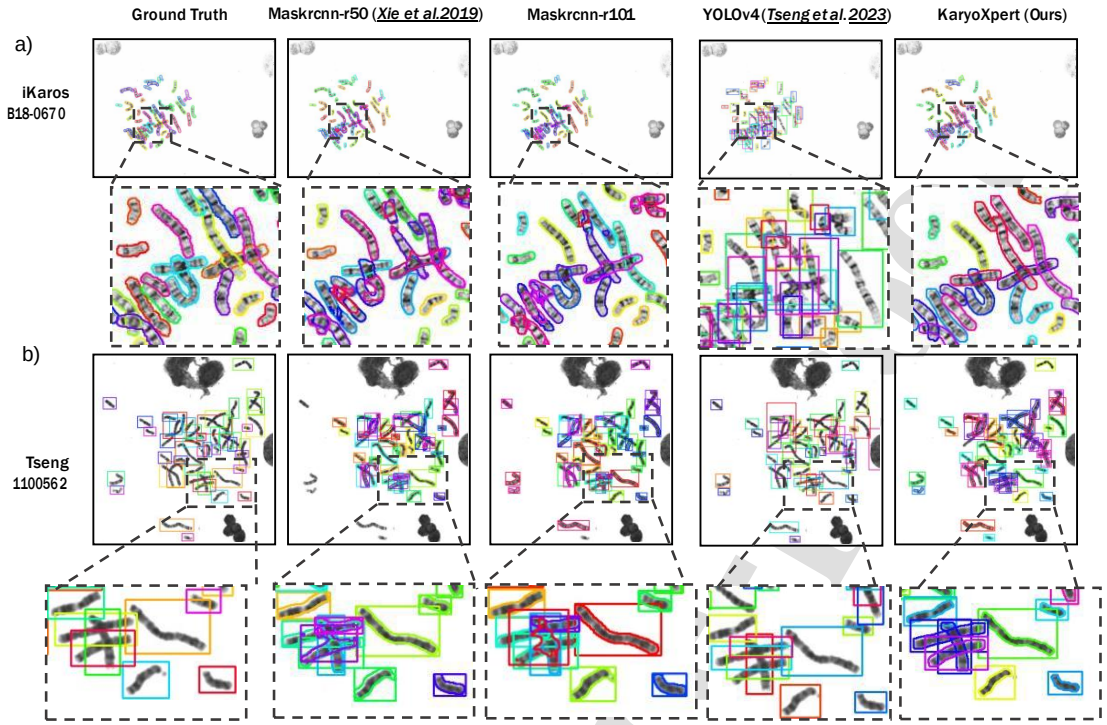

Fig. 5. Intuitive segmentation performance of KaryoXpert and multiple benchmarks on the Ikaros (a) and Tseng (b) test samples. A particular zoom-in region is shown on challenging areas exhibiting overlapping and touching chromosomes.

bounding box in the k_th image over a given IoU threshold (0.5 in this study). False Positive (FP_k) denotes that a predicted bounding box does not have a matched ground truth above the IoU threshold. False Negative (FN_k) here denotes the ground truth that is not detected by any predicted bounding box.

We also include a rather strict enumeration metric, named Whole Correct Ratio (WCR). It is calculated as the percentage of 100% correct enumeration across the whole testing set. The metric is calculated as the ratio N^+/N, where N^+ represents the number of ground truth samples correctly identified (FN_k=0, and FP_k = 0), and N is the size of the test set.

Classification: For a comprehensive evaluation, we employ both Accuracy @1 and Accuracy @5 metrics. The two metrics assess the proportion of instances in which either the top-1 or top-5 prediction results correctly match the ground truth label.

4.2. *Instance-Level Metaphase Image Segmentation Performance*

In this section, we present a comparative analysis of instance-level metaphase image segmentation performance on the test sets from two distinct datasets. Initially, we assess the performance using metrics such as bounding box average precision (AP) and segmentation average precision (AP), within a detection-based framework.

Table 2 provides a summary of the segmentation results. Notably, when examining KaryoXpert's performance on the Ikaros dataset, it's important to note that KaryoXpert was trained using entirely simulated data, we showed 79.9% bonding box mAP with 71.9% segmentation mAP, compared with algorithms trained with paired mask annotations. Due to the effective adaptation and beneficial overfitting, detection-based algorithms (MaskRCNN ResNet-101+RPN He et al. (2017)) demonstrate good performance in terms of segmentation mean average precision (mAP). In terms of recall, KaryoXpert demonstrated a strong ability to accurately identify and capture all relevant positive chromosomes within the metaphase image, achieving 85.3% in bounding box mean average recall (mAR) and 81.8% in segmentation mAR.

On the other hand, when tested on different domains (Tseng dataset), existing detection-based methods Xie et al. (2019); Tseng et al. (2023) faced huge performance drops due to domain shifts under dataset variations (Fig. 5). This domain shift can lead to a lack of model generalization, as the learned features may not be directly applicable to the new dataset. Regarding the bonding-box mean average precision and recall, KaroXpert demonstrates state-of-the-art segmentation robustness with 60.2% mAP and 71.5% mAR (Table 2). It is worth noticing that mask annotations are not available for the Tseng dataset, therefore no segmentation evaluations can be conducted on this dataset. Furthermore, segmentation evaluation is also not applicable to the Yolov4 object detection engine applied from Tseng et al. (2023).

While the morphology-based segmentation algorithm exhibits robustness in its initial step of separating the contours, it demonstrates lower segmentation precision when dealing with overlapping and touching chromosomes. Fig. 6 is an illustra-

Table 2. Performance Metrics of KaryoXpert and benchmarks on the test sets of 2 datasets are presented for both bounding box and segmentation evaluations. mAP refers to mean average precision@IoU0.5:0.95, mAR refers to mean average Recall@IoU0.5:0.95

Dataset	Method	Train w/ Mask Ann	BBox AP@ 0.5	BBox mAP	BBox mAR	Seg AP@0.5	Seg mAP	Seg mAR
Ikaros Dataset (in-House)	Mask R-CNN Xie et al. (2019)	Y	0.988	0.789	0.829	0.978	0.622	0.672
	Mask R-CNN ResNet-101+RPN	Y	**0.988**	**0.812**	0.842	**0.978**	**0.746**	0.777
	YOLACT ResNet-50	Y	0.901	0.707	0.652	0.896	0.505	0.555
	SOLO ResNet-50-1x	Y	0.752	0.563	0.638	0.807	0.525	0.577
	YOLOv4 Tseng et al. (2023)	Y	0.726	0.448	0.548	-	-	-
	YOLOv7	Y	0.731	0.454	0.554	0.786	0.468	0.480
	KaryoXpert	N	0.982	0.799	**0.853**	0.962	0.719	**0.818**
Tseng Dataset (Public) Tseng et al. (2023)	Mask R-CNNXie et al. (2019)	Y	0.783	0.420	0.498	-	-	-
	Mask R-CNN ResNet-101+RPN	Y	0.829	0.470	0.555	-	-	-
	YOLACT ResNet-50	Y	0.752	0.456	0.538	-	-	-
	SOLO ResNet-50-1x	Y	0.728	0.313	0.429	-	-	-
	YOLOv4 Tseng et al. (2023)	Y	0.564	0.256	0.290	-	-	-
	YOLOv7	Y	0.664	0.286	0.365	-	-	-
	KaryoXpert	N	**0.863**	**0.602**	**0.715**	-	-	-

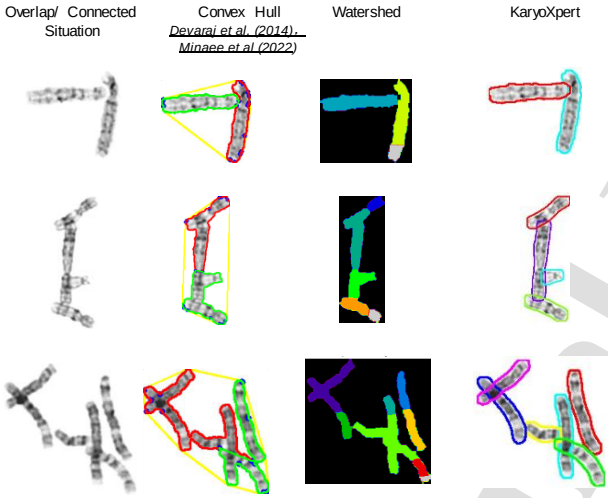

Fig. 6. Segmentation performance of KaryoXpert compared with Morphology-based algorithms.

Table 3. Stability Analysis on Chromosome Enumeration

Dataset	Method	WCR (%) ↑	AER (%) ↓	Acc (%) ↑
Ikaros Dataset	Xie et al. (2019)	25.0	12.89	89.19
	MaskR-CNN-R101	41.0	9.27	90.31
	YOLACT-R50	25.0	13.06	87.92
	SOLO-R50	18. 0	27.41	74.53
	Tseng et al. (2023)	17.0	25.84	79.80
	YOLOv7	34.0	14.3	85.12
	KaryoXpert	**62.0**	**8.08**	**92.3**
Tseng Dataset	Xie et al. (2019)	7.31	40.89	67.17
	MaskR-CNN-R101	13.98	30.07	75.65
	YOLACT-R50	15.95	17.57	73.94
	SOLO R50	12.68	42.43	61.34
	Tseng et al. (2023)	10.6	48.75	68.62
	YOLOv7	31.91	34.69	68.91
	KaryoXpert	**33.51**	**13.28**	**88.09**

tion of the overlapping situations regarding touching (1st row), overlapping (2nd row), and chromosome clusters (3rd row). We visualized the segmentation performance with the convex-hull-based algorithm Minaee et al. (2014) and watershed-based algorithm Vincent and Soille (1991); Devaraj et al. (2022). The morphology-based algorithm effectively handles simple scenarios where less than three chromosomes are present without overlapping. However, in the presence of multiple overlapping regions, particularly in cases involving clusters, the morphology algorithms struggle to provide accurate segmentation results. Detailed overlapping variances and resource consumption can be found in Supplementary Materials.

4.3. Stability Analysis in Chromosome Enumeration

Table 3 presents the stability analysis of KaryoXpert across various benchmarks. Given the fact that deep-learning models are quite sensitive to domain shift during cross-dataset transferring, existing approaches experience a notable performance drop in WCR(%). This effect is further illustrated in Figure 5, where an increase in false negatives (FN) is more prevalent due to domain shift. Owning to a sophisticated algorithmic design that combines the strengths of morphological algorithms and deep-learning neural networks, KaryoXpert demonstrates state-of-the-art stability and robustness performance compared to previous works. It notably decreased the Average Error Rate (AER%) to 8.08% on the iKaros dataset and further reduced it to 13.28% on the Tseng dataset.

4.4. Chromosome Classification

Table 4. Performance Metrics of KaryoXpert on the classification. KaryoNet Xia et al. (2023) takes 46 images as input, therefore no Acc@5 can be reported.

Dataset	Methods	Acc@1	Acc@5
Ikaros Dataset (in-House)	ResNet50	0.9476	0.9957
	EfficientNetV2	0.9591	0.9946
	ConvNeXt	0.9497	0.9946
	ViT	0.9533	0.9943
	Xia et al. (2023)	0.9680	-
	KaryoXpert	**0.9706**	**0.9963**

We then evaluated the quantification performance of deep-learning-based chromosome recognition methods in contrast to the proposed deep-metric learning framework integrated within KaryoXpert. The benchmark algorithm consists of SOTA chromosome classification algorithm KaryoNet Xia et al. (2023), and standard classification backbones: ResNet50 He et al. (2016), EfficientNetV2 Tan and Le (2021), ConvNeXt Liu et al. (2022), and Vision Transformer (ViT) Dosovitskiy et al. (2020). As displayed in Table 4, we computed identification accuracy for both top-1 and top-5 predictions for the 24 potential G-band chromosome classes. All models demonstrate satisfactory performance (above 90% in Acc@1, and Acc@5) when provided with an adequate number of training samples. Different from a supervised learning algorithm, deep metric learning acquires knowledge about the internal connections and class distribution characteristics of chromosomes within a karyotype report, leading to its improved classification performance of 97.06% on the test sets.

4.5. Runtime and Resources Comparison

Table 5. Segmentation Inference Runtime & GPU Memory Comparison

Methods	Inference Run-time per image	GPU Memory
Mask-RCNN Xie et al. (2019)	0.325 sec	4,819Mb
Mask-RCNN ResNet-101 +RPN	0.395 sec	4,385Mb
YOLACT ResNet-50	0.580 sec	8,100Mb
SOLO ResNet-50-1x	0.523 sec	5,413Mb
YOLOv4 Tseng et al. (2023)	**0.191** sec	1,874 Mb
YOLOv7	0.232 sec	2,225Mb
KaryoXpert	0.916 sec	**1,439Mb**
Meta-phase Image	≈7.2 sec (Production Time)	

Another crucial criterion for autonomous karyotyping is the execution time and computational efficiency. As indicated in Table 5 and 6, a karyotyping system must generate automated segmentation results within a 7-second timeframe, parallel with the production time of the metaphase image. Detection-based frameworks exhibit rapid inference speed but consume substantial memory. Since KaryoXpert conducts a five-way segmentation stream, its inference time is longer than that of the detection-based framework. Nevertheless, it still delivers a robust and precise segmentation result within 1 second, utilizing less than 2 GB of GPU memory. One clear conclusion can be drawn regarding the prediction time for KaryoXpert's segmentation: it ensures robustness and delivers satisfactory segmentation performance while providing a good tradeoff between prediction time and the availability of training annotations. On the other hand, in chromosome classification, due to the fact that KaryoNet Xia et al. (2023) incorporates global information in its prediction process. As a result, it contributes to a comparatively slower inference speed for chromosome classification.

Table 6. Classification Inference Runtime & GPU Memory Comparison

Methods	Inference Runtime (per 100 reports)	GPU Memory
ResNet50	16.1 sec	1,687Mb
EfficientNetV2	19.2 sec	1,265Mb
ConvNeXt	15.2 sec	1,511Mb
ViT	15.1 sec	1,843Mb
Xia et al. (2023)	75.4 sec	9,733 Mb
KaryoXpert	**14.4 sec**	**1,105Mb**

4.6. Training Efficiency Analysis

Table 7 presents an efficiency analysis of the training process for the proposed KaryoXpert and benchmark methods, comparing both training duration and GPU memory utilization. The benchmark Vision Transformer necessitates approximately 11.5 GPU hours and approximately 7GB of GPU memory for training. In contrast, the proposed KaryoXpert consumes less amount of GPU memory and approximately 1.6 GPU hours for model training.

Table 7. Classification Training Time Comparison

Methods	Parameter Size	Training Time	GPU Memory
ResNet50	22.4M	6.5h	5,451Mb
EfficientNetV2	19.27M	7.3h	8,119Mb
ConvNeXt	47.18M	13h	9,071Mb
ViT	81.84M	11.5h	6,385Mb
Xia et al. (2023)	28.19M	12h	8,105Mb
KaryoXpert	**1.14M**	**1.6h**	**1,547Mb**

4.7. Clinical Evaluation

In order to evaluate the clinical value of KaryoXpert in the segmentation process, we invited a highly experienced cytogeneticist with more than 10 years of karyotyping experience from Shanghai Xinhua Hospital to give the clinical evaluation for the segmentation results. In this evaluation, we rigorously examined the segmentation outcomes by randomly selecting 20 samples derived from metaphase images sourced from both peripheral blood and amniotic fluid. The aim was to confront KaryoXpert's segmentation results with a meticulous clinical perspective, setting an extremely rigorous criterion where

only segmentation that exhibited a 100% correspondence with the ground truth was deemed true positives. We report the clinical evaluation for KaryoXpert, MaskRCNN-R101He et al. (2017), and commercial software Ikaros V5.5.10 (MetaSystems GmbH) Rose et al. (2019); Vajen et al. (2022) in their segmentation respectively. The results of this evaluation are summarized in Table 8. The results illustrate that KaryoXpert outperforms Ikaros V5.5.10 by enhancing segmentation performance by approximately 30% in precision, and 50% in recall. Furthermore, the false positive rate has been effectively reduced to nearly 4%, significantly accelerating the CRUD (Create, Delete, Read, Update) process through the utilization of KaryoXpert.

Table 8. Clinical evaluation on the Ikaros Dataset separately for peripheral blood and amniotic fluid samples, FPR stands for False positive rate

Sample	Method	Precision (%) ↑	Recall (%) ↑	FPR (%) ↓
Peripheral Blood	KaryoXpert	**96.31**	**93.47**	**3.68**
	Mask R-CNN-R101	88.28	91.28	11.71
	iKaros V5.5.10	66.11	47.43	52.56
Amniotic Fluid	KaryoXpert	**95.84**	**97.51**	**4.15**
	Mask R-CNN-R101	89.13	94.41	10.58
	iKaros V5.5.10	59.66	39.13	60.87

5. Conclusions

Automated karyotyping holds great significance in advancing cytogenetic research. In this paper, we focus on the instance-level chromosome segmentation and classification and proposed KaryoXpert. KaryoXpert leverages the strengths of both morphology algorithms and deep learning models, enabling efficient training, that breaks the limit for instance-level mask annotation requirement. Meanwhile, its strong robustness guarantees a plug-and-play manner to boost performance under multiple application scenarios under domain shift. KaryoXpert was predominantly developed and rigorously assessed using real-world clinical datasets. Specifically, the G-band chromosome samples employed in this study were primarily sourced from peripheral blood and amniotic fluid.

The probable limitation of KaryoXpert stems from its restricted capacity to handle R-band cytogenetic data extracted from bone marrow samples, which tend to have a lower banding resolution with more blurred patterns. Nevertheless, the proposed method introduces a novel paradigm to AI-driven automated karyotyping analysis, offering substantial clinical value and promising future prospects.

Acknowledgment

This work was supported by the Office of Sponsored Research (OSR), King Abdullah University of Science and Technology (KAUST), under grant URF/1/4352-01-01, FCC/1/1976-44-01, FCC/1/1976-45-01, REI/1/5234-01-01, REI/1/5414-01-01, REI/1/5289-01-01, REI/1/5404-01-01.

References

Abid, F., Hamami, L., 2018. A survey of neural network based automated systems for human chromosome classification. Artificial Intelligence Review 49, 41–56.

Al-Kharraz, M.S., Elrefaei, L.A., Fadel, M.A., 2020. Automated system for chromosome karyotyping to recognize the most common numerical abnormalities using deep learning. IEEE Access 8, 157727–157747.

Beucher, S., 1979. Use of watersheds in contour detection, in: Proc. Int. Workshop on Image Processing, Sept. 1979, pp. 17–21.

Biyani, P., Wu, X., Sinha, A., 2005. Joint classification and pairing of human chromosomes. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2, 102–109.

Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 .

Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2019. Yolact: Real-time instance segmentation, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 9157–9166.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C.C., Lin, D., 2019. MMDetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155 .

Devaraj, S., Madian, N., Suresh, S., 2022. Mathematical approach for segmenting chromosome clusters in metaspread images. Experimental Cell Research 418, 113251.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 .

Fan, Y.S., Siu, V.M., Jung, J.H., Xu, J., 2000. Sensitivity of multiple color spectral karyotyping in detecting small interchromosomal rearrangements. Genetic testing 4, 9–14.

Haferlach, C., Hänselmann, S., Walter, W., Volkert, S., Zenger, M., Kern, W., Stengel, A., Lörch, T., Haferlach, T., 2020. Artificial intelligence substantially supports chromosome banding analysis maintaining its strengths in hematologic diagnostics even in the era of newer technologies. Blood 136, 47–48.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, in: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Jindal, S., Gupta, G., Yadav, M., Sharma, M., Vig, L., 2017. Siamese networks for chromosome classification, in: Proceedings of the IEEE international conference on computer vision workshops, pp. 72–81.

Kang, S., Han, J., Chu, Y., Lee, I., Joo, H., Yang, S., 2022. Automated chromosomes counting systems using deep neural network, in: 2022 International Conference on Electronics, Information, and Communication (ICEIC), IEEE. pp. 1–3.

Lerner, B., Guterman, H., Dinstein, I., Romem, Y., 1995. Medial axis transform-based features and a neural network for human chromosome classification. Pattern Recognition 28, 1673–1683.

Lin, C., Zhao, G., Yin, A., Yang, Z., Guo, L., Chen, H., Zhao, L., Li, S., Luo, H., Ma, Z., 2021. A novel chromosome cluster types identification method using resnext wsl model. Medical Image Analysis 69, 101943.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer. pp. 740–755.

Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11976–11986.

Marcel, S., Rodriguez, Y., 2010. Torchvision the machine-vision package of torch, in: Proceedings of the 18th ACM international conference on Multimedia, pp. 1485–1488.

Markou, C., Maramis, C., Delopoulos, A., Daiou, C., Lambropoulos, A., 2012. Automatic chromosome classification using support vector machines. Pattern Recognition: Methods and Applications , 1–24.

Mei, L., Yu, Y., Shen, H., Weng, Y., Liu, Y., Wang, D., Liu, S., Zhou, F., Lei, C., 2022. Adversarial multiscale feature learning framework for overlapping chromosome segmentation. Entropy 24, 522.

Micci, F., Teixeira, M.R., Heim, S., 2001. Complete cytogenetic characterization of the human breast cancer cell line ma11 combining g-banding, comparative genomic hybridization, multicolor fluorescence in situ hybridization, rx fish, and chromosome-specific painting. Cancer genetics and cytogenetics 131, 25–30.

Minaee, S., Fotouhi, M., Khalaj, B.H., 2014. A geometric approach to fully automatic chromosome segmentation, in: 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), IEEE. pp. 1–6.

Ming, D., Tian, J., 2010. Automatic pattern extraction and classification for chromosome images. Journal of Infrared, Millimeter, and Terahertz Waves 31, 866–877.

Peng, J., Lin, C., Zhao, G., Yin, A., Chen, H., Guo, L., Li, S., 2021. Identification of incorrect karyotypes using deep learning, in: International Conference on Artificial Neural Networks, Springer. pp. 453–464.

Qin, Y., Wen, J., Zheng, H., Huang, X., Yang, J., Song, N., Zhu, Y.M., Wu, L., Yang, G.Z., 2019. Varifocal-net: A chromosome classification approach using deep convolutional networks. IEEE transactions on medical imaging 38, 2569–2581.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 28.

Rose, R., Venkatesh, A., Pietilä, S., Jabeen, G., Jagadeesh, S.M., Seshadri, S., 2019. Utility and performance of bacterial artificial chromosomes-on-beads assays in chromosome analysis of clinical prenatal samples, products of conception and blood samples. Journal of Obstetrics and Gynaecology Research 45, 830–840.

Saleh, H.M., Saad, N.H., Isa, N.A.M., 2019. Overlapping chromosome segmentation using u-net: convolutional networks with test time augmentation. Procedia Computer Science 159, 524–533.

Sharma, M., Saha, O., Sriraman, A., Hebbalaguppe, R., Vig, L., Karande, S., 2017. Crowdsourcing for chromosome segmentation and deep classification, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 34–41.

Sharma, M., Vig, L., et al., 2018. Automatic classification of low-resolution chromosomal images, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp. 0–0.

Tan, M., Le, Q., 2021. E**ffi**cientnetv2: Smaller models and faster training, in: International conference on machine learning, PMLR. pp. 10096–10106.

Tseng, J.J., Lu, C.H., Li, J.Z., Lai, H.Y., Chen, M.H., Cheng, F.Y., Kuo, C.E., 2023. An open dataset of annotated metaphase cell images for chromosome identification. Scientific Data 10, 104.

Vajen, B., Hänselmann, S., Lutterloh, F., Käfer, S., Espenkötter, J., Beening, A., Bogin, J., Schlegelberger, B., Göhring, G., 2022. Classification of fluorescent r-band metaphase chromosomes using a convolutional neural network is precise and fast in generating karyograms of hematologic neoplastic cells. Cancer Genetics 260, 23–29.

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an e**ffi**cient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis & Machine Intelligence 13, 583–598.

Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2023. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7464–7475.

Wang, X., Kong, T., Shen, C., Jiang, Y., Li, L., 2020. Solo: Segmenting objects by locations, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16, Springer. pp. 649–665.

Wei, H., Gao, W., Nie, H., Sun, J., Zhu, M., 2022. Classification of giemsa staining chromosome using input-aware deep convolutional neural network with integrated uncertainty estimates. Biomedical Signal Processing and Control 71, 103120.

Wu, Y., Tan, X., Lu, T., 2020. A new multiple-distribution gan model to solve complexity in end-to-end chromosome karyotyping. Complexity 2020, 1–15.

Wu, Y., Yue, Y., Tan, X., Wang, W., Lu, T., 2018. End-to-end chromosome karyotyping with data augmentation using gan, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE. pp. 2456–2460.

Xia, C., Wang, J., Qin, Y., Gu, Y., Chen, B., Yang, J., 2022. An end-to-end combinatorial optimization method for r-band chromosome recognition with grouping guided attention, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 3–13.

Xia, C., Wang, J., Qin, Y., Wen, J., Liu, Z., Song, N., Wu, L., Chen, B., Gu, Y., Yang, J., 2023. Karyonet: Chromosome recognition with end-to-end combinatorial optimization network. IEEE Transactions on Medical Imaging .

Xiao, L., Luo, C., 2021. Deepacc: automate chromosome classification based on metaphase images using deep learning framework fused with priori knowledge, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 607–610.

Xiao, L., Luo, C., Yu, T., Luo, Y., Wang, M., Yu, F., Li, Y., Tian, C., Qiao, J., 2020. Deepacev2: Automated chromosome enumeration in metaphase cell images using deep convolutional neural networks. IEEE Transactions on Medical Imaging 39, 3920–3932.

Xie, N., Li, X., Li, K., Yang, Y., Shen, H.T., 2019. Statistical karyotype analysis using cnn and geometric optimization. IEEE Access 7, 179445–179453.

Yang, W., Stotler, B., Sevilla, D.W., Emmons, F.N., Murty, V.V., Alobeid, B., Bhagat, G., 2010. Fish analysis in addition to g-band karyotyping: utility in evaluation of myelodysplastic syndromes? Leukemia research 34, 420–425.

Zhang, J., Hu, W., Li, S., Wen, Y., Bao, Y., Huang, H., Xu, C., Qian, D., 2021. Chromosome classification and straightening based on an interleaved and multi-task network. IEEE Journal of Biomedical and Health Informatics 25, 3240–3251.

- We propose KaryoXpert as a new paradigm in automated karyotyping, that eliminates the need for manually annotated ground truth instance masks during training.

- Advantages from both morphology algorithm and deep-learning model are merged, which guarantees KaryoXpert has strong robustness and high segmentation accuracy across multiple datasets that bear domain shift.

- Metric learning-based chromosome classification is proposed to enable a more accurate chromosome recognition, effectively reducing the impact of batch effects, domain shift, and inter-class similarity.

- GPU acceleration enables real-time inference with cutting-edge clinical accuracy on real-world clinic cytogenetic samples.

In contrast to conventional karyotyping methods relying heavily on mask annotation, our methodology integrates a novel fusion of deep learning and morphological algorithm. This innovative fusion enables mask-free training, enhancing model robustness, and boost segmentation and classification accuracy. Our experiments on real-world NIPT datasets showcase a remarkable 20% improvement in clinical evaluation compared with existing methods, showing the potential clinical impact of our innovative approach in aiding diagnosis and treatment planning in cytogenetic imaging

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: