E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

**中图法分类号**:TP399　　**文献标识码**:A　　**文章编号**:1006-8961(2023)02-0570-19

**论文引用格式**:Yi X S, Yin A H, Huang J S, Peng J, Chen H B, Guo L, Lin C C, Li S Y and Zhao G S. 2023. Performance evaluation of mainstream chromosome recognition algorithms under deep learning. Journal of Image and Graphics,28(02):0570-0588(易序晟,尹爱华,黄杰晟,彭璟,陈汉彪,郭莉,林成创,李双印,赵淦森. 2023. 深度学习下主流染色体分类算法的性能评估. 中国图象图形学报,28(02):0570-0588)[DOI:10.11834/jig.210669]

# 深度学习下主流染色体分类算法的性能评估

易序晟[1,2,3],尹爱华[4],黄杰晟[1,2,3],彭璟[1,2,3],陈汉彪[4],郭莉[4],
林成创[1,2,3],李双印[1,2,3],赵淦森[1,2,3]*

1. 华南师范大学计算机学院,广州　510631;2. 广州市云计算安全与测评技术重点实验室,广州　510631;
3. 华南师范大学唯链区块链技术与应用联合实验室,广州　510631;4. 广东省妇幼保健院,广州　511400

**摘　要**:**目的**　染色体分类是医学影像处理的具体任务之一,最终结果可为医生提供重要的临床诊断信息,在产前诊断中起着重要作用。深度学习由于强大的特征表达能力在医学影像领域得到了广泛应用,但是基于深度学习的大部分染色体分类算法都是在轻量化私有数据库上得到的不同水准的分类结果,难以客观评估不同算法间的优劣,导致缺乏对算法的临床筛选标准,因此迫切需要在大规模数据库上对不同算法开展基于同样数据级的性能评估,以获取具有客观可对比性的性能数据,这对于科研成果的转化具有重要意义。**方法**　本文基于广东省妇幼保健院提供的染色体数据,构建了包含126 453条染色体的临床数据库,精选6个主流染色体分类模型在该数据库上展开对比实验与性能评估。**结果**　在本文构建的大规模染色体临床数据库上,实验和分析发现,参评模型分类准确率均达到92%以上,其中MixNet模型提升后分类效果最好,为98.92%。即使分类性能落后的模型在本数据集上训练也得到明显改善,准确率从86.7%提升至92.09%,相比早期报告的性能提升了5.39%。**结论**　开展实证研究实验发现,数据库规模大小是影响染色体分类算法能否取得理想分类效果的重要因素之一。对于染色体分类任务而言,残差神经网络是比较合适的网络结构,但结果方面缺乏可解释性等原因,导致与高精度临床应用要求还存在差距。基于深度学习技术的染色体分类研究还需要进一步深入开展。

**关键词**:医学影像处理;深度学习;染色体分类;残差神经网络;分类评估

# Performance evaluation of mainstream chromosome recognition algorithms under deep learning

Yi Xusheng[1,2,3], Yin Aihua[4], Huang Jiesheng[1,2,3], Peng Jing[1,2,3], Chen Hanbiao[4],
Guo Li[4], Lin Chengchuang[1,2,3], Li Shuangyin[1,2,3], Zhao Gansen[1,2,3]*

*1. School of Computer Science and Technology, South China Normal University, Guangzhou　510631, China;*
*2. Key Laboratory on Cloud Security and Assessment Technology of Guangzhou, Guangzhou　510631, China;*
*3. VeChain Blockchain Technology and Application Joint Laboratory, South China Normal University, Guangzhou　510631, China;*
*4. Guangdong Maternity and Child Health Hospital, Guangzhou　511400, China*

**收稿日期**:2021-08-17;**修回日期**:2022-04-07;**预印本日期**:2022-04-14

* **通信作者**:赵淦森　gzhao@m.scnu.edu.cn

**基金项目**:国家重点研发计划资助(2018YFB1404402);国家社会科学基金重大项目(19ZDA041);中国高校产学研创新基金——新一代信息技术创新项目——基于区块链的专利申请前评估的模型定制技术(2020ITA09006)

**Supported by**: National Key R&D Program of China (2018YFB1404402); National Social Science Foundation Major Project (19ZDA041); China University Industry-University-Research Innovation Fund——New Generation Information Technology Innovation Project——Model Customization Technology for Pre-Patent Application Evaluation Based on Blockchain (2020ITA09006)

**Abstract**: **Objective** Deep learning technique-based medicinal image processing is essential for clinical information in related to disease diagnosis, treatment, and surgical planning. Chromosome-relevant segmentation can be as one of the specific tasks for medical-based image processing. It is beneficial to prenatal diagnosis via clinical diagnosis information gathering and analysis. In recent years, an end-to-end training features-based deep learning technique has been developing intensively. Chromosome-relevant segmentation has been facilitating as well. Chromosomes can be one of the key carriers of genetic information. Chromosomes-based genetic information analysis is often employed for human genetic diseases. Chromosome images-related karyotyping analysis is a commonly-used method for diagnosing birth defects and it can be as the "gold standard" for the clinical diagnosis of genetic diseases. Chromosome segmentation is challenged for the manipulation problem in the context of chromosome karyotype analysis. It has a strong reference value for prenatal diagnosis results. However, most of chromosome-related segmentation algorithms have restricted by its heterogeneity, resulting in a lack of a screening standard for algorithms in clinical applications. To carry out more comparative experiments on the large-scale chromosome-constructed database, we develop a multiple of chromosome-essential segmentation models. **Method** Our database is constructed and segmented in terms of the chromosome karyotype (funded by Guangdong Maternity and Child Health Hospital). first, it consists of large-scale chromosome clinical data in relevant to 126 453 chromosome samples. Then, the publicly-available multi-chromosome-essential recognition models are selected. Finally, experiments and performance evaluation of our model is carried out in the clinical chromosome database. **Result** Random sampling-stratified experiment is used to divide the clinical chromosome data set into training data set (80%), validation data set (10%), and test data set (10%) totally. The models-selected are all developed in terms of Pytorch framework. The training process of the model is summarized as mentioned below: First, all models are pre-trained and migrated from the ImageNet classification task. Second, a single-stage cycle learning (1 cycle LR) learning and training method is used to balance the performance of each model in the clinical data set gradually. The batch size of all experiments is set to 32 (batch_ size = 32). The balanced loss function is based on the cross-entropy loss function smoothed by the mark of $\alpha = 0.1$. The learning rate is set to 1E − 4. The hyper parameter for the maximum number of training iterations is set to 500. Moreover, the early stopping strategy will be implemented to terminate the training process if the verification loss is not decreased in five consecutive periods. Finally, our training weights can be optimized and restored for the corresponding model. Large-scale clinical-oriented chromosome data sets are beneficial to evaluating existing chromosome classification methods and improving their performance. The CirNet and MixNet models-based initial performance and classification effect are optimized on the original ResNet networks. It can strengthen the depth and width of the network, increase the number of parameters, and get a better classification level. The guarantee of the amount of data can alleviate the problem of over-fitting. The classification accuracy rate is optimized and outreached to 98.92%, but there is still a gap between the high-precision clinical applications. **Conclution** To develop deep learning technique-based chromosome classification and ensure its high-precision potentially, a refined network structure is required to be designed and tackled the chromosome images-related homogeneity further. The quality and quantity of chromosome data samples should be guaranteed as well.

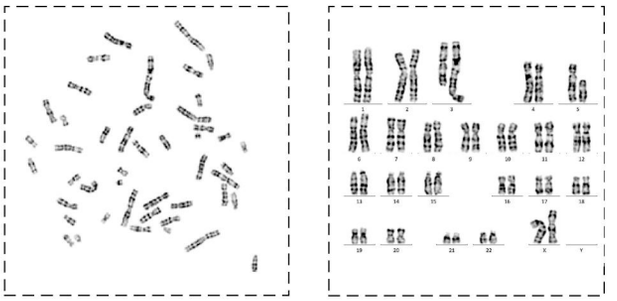
**Key words**: medical image processing; deep learning; chromosome classification; residual neural network; classification evaluation

## 0 引 言

根据2010 年世界卫生组织发布的《出生缺陷报告》,先天性疾病是常见疾病,全世界因先天性异常而死亡的人数约为 260 000 人,占所有新生儿死亡病例大约 7% 。中国的出生缺陷儿绝对数量巨大。据《中国出生缺陷防治报告(2012)》(中华人民共和国卫生部,2012)报告显示,我国出生缺陷发生率在5.6% 左右。先天性疾病和出生缺陷可以造成自然流产和死产,是导致新生儿死亡和残疾的一个重要原因。它们可以危及生命,导致长期残疾,从而对个人、家庭和社会造成不利影响。因此,针对出生缺陷诊断的科学研究具有重要价值(Abid 和 Hamami, 2018)。

染色体包含着人类遗传信息,常用于分析人类遗传病。一般来说,人体细胞中在正常情况下包含46 条染色体,其中由 22 对常染色体(1—22 类)和

1 对性染色体(女性为两条 X 性染色体,男性为一条 X 染色体和一条 Y 染色体)构成(林成创 等,2020)。染色体核型化分析是产前诊断和遗传疾病诊断的重要且常用方法,是临床上遗传疾病诊断的金标准。如图 1 所示,染色体核型化分析是指从细胞分裂中期的显微镜图像中分割得到染色体实例后,凭借 ISCN(international system for human cytogenetic nomenclature)标准(Shaffer 等,2013),最终形成染色体核型图的过程。



(a) 中期细胞图像　　(b) 染色体核型图

图 1　染色体核型分析示例

Fig. 1　Example of chromosome karyotype analysis

((a) metaphase image; (b) chromosome karyotype map)

医学影像是现代医学最重要的临床诊断和鉴别诊断方法,也是计算机辅助诊断的重要依据(田娟秀 等,2018)。染色体分类作为医学影像处理的子任务,是临床遗传疾病诊断的重要方法,通常由熟练的细胞遗传学专家手工完成。受染色体非刚性特性影响,同一类型的染色体常具有完全不同的形状和方向,给染色体分类研究带来巨大挑战。染色体分类是染色体核型化分析中最复杂也是最容易出错的中间环节,对临床医生出具产前诊断结果有着很好的借鉴价值。因此,染色体分类是一类极具挑战性且值得深入研究的热点问题。

在深度学习兴起之前,染色体分类技术主要分为两大技术路线,基于人工神经网络(artificial neural network, ANN)的技术路线(Kubola 和 Wayalun, 2018; Popescu 等, 1999; Moradi 等, 2003; Oskouei 和 Shanbehzadeh, 2010)和基于概率神经网络(probalistic neural network, PNN)的技术路线(Poletti 等, 2011; Rungruangbaiyok 和 Phukpattaranont, 2010; Khan 等, 2012)。第 1 类方法主要采用多层感知机(multilayer perceptron, MLP)提取染色体特征,但有一个复杂的特征筛选过程;第 2 类方法的精度较低,所需的训练时间较短。早期染色体分类方法有着复杂的特征选择过程,手工设计的特征提取器有固有的局限性,无法全面涵盖图像的低级像素特征,更加难以设计出高层抽象的特征,导致染色体的错误分类(金旭 等, 2020)。

随着以深度学习为代表的人工智能技术的快速发展,采取深度学习技术的医学影像分析成为研究热点,吸引众多学者投入染色体分类的研究。Sharma 等人(2017)和 Swati 等人(2017)分别较早地提出基于改进卷积神经网络和基于改进孪生网络的染色体分类方法,在各自选择的数据集中取得了 86.7% 和 84.6% 的分类准确率。Lin 等人(2020b)提出残差神经网络变体模型和自适应接口(image adaptive interface, IAI)的染色体分类方法。其中,自适应接口有效解决了染色体图像尺寸输入不一致问题。在作者人工标注的 2 990 条染色体图像数据集中,测试准确率为 95.98%。Zhang 等人(2021)提出了结合交错和多任务学习网络的染色体分类和校正模型,按 3 个阶段进行特征学习。在作者收集的数据集上进行性能评估,准确率为 98.1%。

染色体分类任务研究取得了一定进展,但基于深度学习的染色体分类研究存在滞后性,且暴露出一些问题。1)难以收集到大规模带标签的染色体数据。数据量和数据质量都在数据获取过程中存在明显的人员和技术阻碍,存在数据壁垒。标准化数据往往采取人工标注方式,存在着标注难度大、标注信息难以获取以及高成本等问题。2)模型性能的好坏与训练数据量密切相关。深度学习需要大量的训练数据,目前染色体分类研究大量依赖高成本标注背景下的染色体小样本数据。虽然目前染色体分类算法已经取得了一些进展,但大部分是在轻量化数据集上得到的不同水准的分类性能,训练数据不够充分,导致现有模型可能处于低层次精度水平。3)模型难以验证,成果转化难。

存在上述问题的根本原因是数据量缺乏。针对上面的问题,本文尝试从以下 3 个方面开展研究工作。1)针对第 1 个问题,构建目前最大规模的临床染色体数据库。其中,分割染色体实例以及开展的数据标注工作全部由专业临床医生团队全程参与并完成。2)针对第 2 个问题,在保证数据量的前提下,提供给现有模型足够的训练数据,尝试验证现有模型性能能否得到整体的提升。3)针对第 3

个问题,在构建好的数据集上展开染色体分类算法之间的性能评估实验,探究不同网络结构之间是否存在性能差异,为临床应用提供对网络模型的初筛建议。

为开展性能评估实验,本文工作从广东省妇幼保健院医学遗传中心获得 126 453 条染色体样本的临床数据集,从近年基于深度学习的染色体分类的相关研究成果中重点挑选出 6 种主流染色体分类算法,所有参评模型进行染色体分类识别性能测试(https://github.com/CloudDataLab/ChromesomeBenchmark)。本文主要贡献如下:1)阐述近 5 年基于深度学习的染色体分类方法。按网络结构的设计思想和时间发展顺序进行介绍,并给出网络间存在的联系,对相关研究人员有较好的借鉴意义。2)精选 6 个主流染色体分类模型在大规模临床数据集展开性能评估实验。本次评估实验是在染色体分类研究领域中第一次进行较大规模实验评估,对染色体分类研究具有重要价值。3)在评估实验中,带有不同参数量的模型借助准确率、精确率、召回率和 F1 值 4 类评价指标来分析模型在大型数据集上的分类潜力,这对于更加准确地认识深度学习在染色体分类中的分类性能具有重要意义。

## 1 基于深度学习的染色体分类算法

染色体具有非刚性的内在性质。在显微镜下染色体形状和方向的多元化对于染色体分类过程带来了巨大挑战(Grisan 等,2009)。随着深度学习技术的发展,将深度学习技术用于染色体分类成为研究热点,但确保数据规模和设计出高性能分类模型存在巨大挑战。

### 1.1 问题定义和描述

染色体分类是将单条染色体图像按照 ISCN 规则确定所有染色体组号和染色体号的过程,如图 2 所示。关于染色体组号,ISCN 规则(McGowan-Jordan 等,2016)将染色体分成 7 组,从 A 到 G 依次编号。其中 A 组包括 1、2、3 号常染色体,B 组包括 4、5 号常染色体,C 组包括 6—12 号常染色体以及 X 性染色体,D 组包括 13、14、15 号常染色体,E 组包括 16、17、18 号常染色体,F 组包括 19、20 号常染色体,G 组包括 21、22 号常染色体和 Y 性染色体。关于染色体号,ISCN 规则中对染色体核型分析的描述如下:在核型图中,常染色体按照长度递减的顺序用数字 1 到 22 编号(唯一例外是 21 号染色体比 22 号短),性染色体用 X 和 Y 表示。

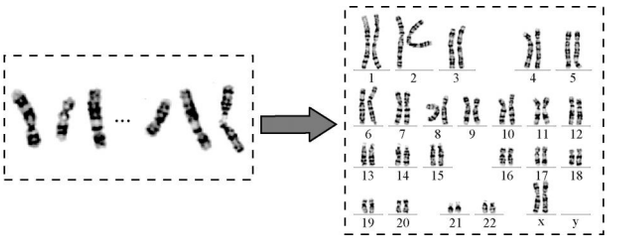



图 2 染色体分类过程

Fig. 2 The process of chromosome classification

染色体分类问题使用形式化公式描述如下:使用 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_N\}$ 表示包含 $N$ 个染色体图像样本的数据集,其中 $x_j$ 表示第 $j$ 个染色体图像样本;使用 $\boldsymbol{Y}=\{y_1,y_2,\cdots,y_N\}$ 表示与 $\boldsymbol{X}$ 数据集对应的染色体真实类别集合,通常称为 ground truth(GT)标签集;其中 $y_j\in\{1,2,\cdots,22,\mathrm{X},\mathrm{Y}\}$ 表示第 $j$ 个染色体图像样本的真实类别;使用 $F(\cdot)$ 表示一个染色体分类算法或模型,使用 $y_j'=F(x_j)$ 表示 $F(\cdot)$ 算法或模型推理第 $j$ 个染色体样本的预测类别;使用 $\boldsymbol{Y}'=\{x_1',x_2',\cdots,x_N'\}$ 表示 $F(\cdot)$ 的分类结果集合,通常称为 Predicted Set 测试数据集。

### 1.2 常用数据库和评价指标

#### 1.2.1 常用数据库

染色体影像依赖显微镜制作。采样过程中,既要保证拍摄的影像完全呈现出细胞中的所有染色体,还需要尽可能提高染色体图像的清晰度,无法同时兼顾成像视野与放大倍数,而且制作染色体图像的过程过于烦琐。现实中收集一个大型且带有标注信息的染色体数据集难度更大,一方面,收集过程费时,还需相关专业领域的知识储备;另一方面,医疗机构出于隐私保护等医学伦理问题的考虑,基本不公开相应染色体数据。即使收集数据途径得到保证,但将数据集进行开放共享需具备一定的前提条件,即科研团队和医疗机构存在着研究合作关系并需满足患者知情以及严格的数据脱敏等要求。打造高质量的大型公开数据集过程相当具有挑战性。历年染色体分类研究数据集如表 1 所示。

最早公开的数据集是 Poletti 等人(2008)提出并由细胞学专家手动分割得到的,名为 BiolmLab 数据集。数据集分为 119 个文件夹(细胞),每个文件夹

表 1　染色体分类常用数据库

Table 1　Common datasets for chromosome classification

| 方法 | 数据库大小/幅 | 是否公开 |
|---|---|---|
| Poletti 等人(2008) | 5 474 | 是 |
| Sharma 等人(2017) | 1 800 | 否 |
| Swati 等人(2017) | 1 740 | 否 |
| Zhang 等人(2018) | 10 304 | 否 |
| Sharma 等人(2018) | 5 256 | 否 |
| Hu 等人(2019) | 4 184 | 是 |
| 吕丹丹(2019) | 18 400 | 否 |
| Qin 等人(2019) | 87 831 | 否 |
| Lin 等人(2020b) | 2 990 | 是 |
| Remya 等人(2020) | 21 423 | 否 |
| Lin 等人(2020a) | 2 990 | 否 |
| 张成成等人(2020) | 1 275 | 否 |
| Zhang 等人(2021) | 32 810 | 否 |

包含 46 条染色体,共 5 474 条染色体,总大小为 10.5 MB(图像格式:BMP,8 位/像素)。每幅图像分辨率大小取染色体的最佳分布位置,以保证标注数据的高质量。BiolmLab 数据集(http://bioimlab.dei. unipd. it/)对染色体分类有着重要的研究价值。部分数据集如图 3 所示。Hu 等人(2019)从 91 张核型图片中分割得到 4 184 条正常染色体图像并制作成公开数据集(https://github. com/Xi-Hu/Chromosome - Images)。其中,图像背景统一扩展为 100(宽) × 220 (高)像素。染色体标记信息体现在文件名中。例如,在“161923. 054. K. chr1. 1invert_pad. gif”文件名中,“161923. 054”是样本编号,K 代表核型图像,“chr1. 1”代表该图像是成对染色体 1 号中的一条(另一条染色编号为 chr1. 2)。Lin 等人(2020b)于 2020 年发布了染色体数据库(https://github. com/ CloudDataLab/CIR-Net),该数据集包括 2 990 幅尺寸约为 224 × 224 像素的单条染色体图像,每个像素占用的有效比特数为 8(8bpp)。

手工标注染色体数据是一个费时费力的阶段(周雪雁 等,2004)。因此,收集大型公开可用的染色体数据集过程具有挑战性。值得注意的是,即使数据集极少数开源,但染色体分类的研究从未间断。

1.2.2　评价指标

染色体分类是一个多分类任务,但实际操作时往往将每类染色体都单独作为一个二分类问题,然后再对所有类别进行加权计算。其中任意一组染色体的二分类混淆矩阵如表 2 所示。对于第 $j$ 组染色体,$TP_j$ 表示实际属于类别 $j$ 的染色体被模型预测为 $j$。$FP_j$ 表示实际不属于类别 $j$ 的染色体被模型预测为 $j$。$FN_j$ 表示实际属于类别 $j$ 的染色体被模型预测为不是类别 $j$。$TN_j$ 表示实际不属于类别 $j$ 的染色体被模型预测为不是类别 $j$。

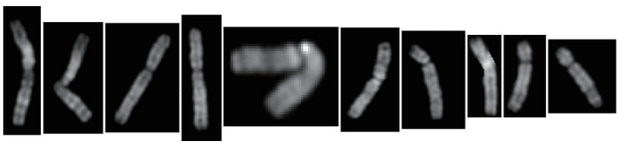


图 3　BiolmLab 数据集部分情况

Fig. 3　Parts of the BiolmLab dataset

表 2　二分类混淆矩阵

Table 2　Confusion matrix of binary classfication

| 真实类别 | 预测结果 | |
|---|---|---|
| | 正例 | 反例 |
| 正例 | 真正例(TP) | 假反例(FN) |
| 反例 | 假正例(FP) | 真反例(TN) |

不同的分类算法需要进行客观评估,尤其是量化的评估指标。对于染色体分类问题,更加需要一类具备客观可量化、可解释以及被广泛接受特性的评价指标。常用染色体分类算法的评价指标有准确率(accuracy, ACC)、精确率 PPV(positive predictive value)、召回率 TPR(true positive rate)和 F1 分数($F_1$ score)。准确率指每一类染色体在所有染色体中正确分类的比例,可以判断总的分类正确率,常作为评估分类效果的最主要和最常用的指标。精确率也称查准率,用于评估所有被预测为正样本中实际为正样本的概率。精确率考察单个类别的准确率,准确率考察整体的准确率。召回率也称查全率,用于评估实际为正样本中被预测为正样本的概率。$F_1$ 分数是精确率和召回率的加权调和平均值。分别表示为

$$ACC = \frac{1}{N}\sum_{j=1}^{N} TP_j \tag{1}$$

$$PPV = \frac{1}{N}\sum_{j=1}^{N} \frac{TP_j}{TP_j + FP_j} \tag{2}$$

$$TPR = \frac{1}{N}\sum_{j=1}^{N} \frac{TP_j}{TP_j + FN_j} \tag{3}$$

$$F_1 = \frac{1}{N}\sum_{j=1}^{N}\frac{2 \times PPV_j \times TPR_j}{PPV_j + TPR_j} \quad (4)$$

式中，$N$ 表示染色体的类别总数。

上述评价指标能够提供分类任务中的评价参考，但在实际使用时需根据任务需求选择不同评价指标进行组合，多方面评价分类算法的准确性和稳定性。

### 1.3 模型设计

染色体分类网络模型主要分为 3 种设计思路，即网络内部堆叠卷积层的设计思路、网络内部精细化模块的优化设计思路（如跳跃连接和相似度匹配的思想）以及拼接图像多尺度特征的多任务模型设计思路。针对以上 3 种设计思路，本文提出了基于深度学习的染色体分类算法的划分框架，如图 4 所示。具体分为基于卷积神经网络（convolutional neural network，CNN）及其变体的分类模型、基于 ResNet 和 Siamese 的分类模型以及自设计的多任务学习分类模型。

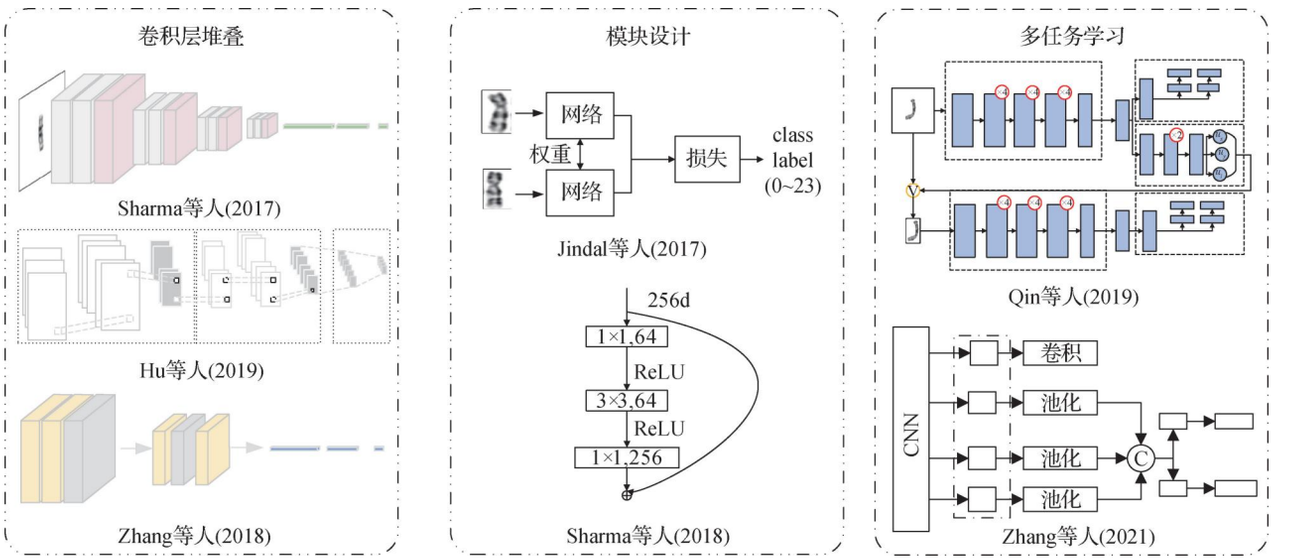



图 4　深度学习下染色体分类研究

Fig. 4　Research on chromosome classification under deep learning

#### 1.3.1 基于 CNN 及其变体的分类模型

卷积神经网络（CNN）作为深度学习的代表性算法之一，是包含卷积计算和一定深度结构的前馈型神经网络，借助堆叠卷积层层数的网络模型能有效将大数据量的图像降维成小数据量，使原图像信号特征增强，并降低噪音。不同于传统的图像分类方法，基于卷积神经网络的图像分类方法具有很好的表征学习能力，其卷积核参数共享和层次化的核心思想使 CNN 能对图像像素进行潜在特征学习，较大幅度地提高图像分类效果。

Sharma 等人（2017）提出结合众包技术、数据预处理和卷积神经网络（CNN）算法进行染色体分类。该研究将染色体灰度图像集的分割任务众包给专业技术人员，构建了染色体数据集。在此数据集上，按照拉直弯曲染色体、染色体图像归一化和使用 Deep CNN 模型 3 个流程将染色体分为 24 类。Deep CNN 模型架构如图 5 所示，该模型串行了 4 个卷积模块，对 120 × 120 像素的染色体图像多重卷积下采样，自学习训练样本的内部特征。池化层也称下采样层，会压缩输入的特征图，一方面减少了特征数，能显著减少训练参数，另一方面保持了特征的某种不变性（旋转、平移和伸缩等）。模型的训练与测试在 1 800 条染色体下进行，经过数据预处理操作后，实验表明，Deep CNN 模型有较好的性能，准确率为 86.7% 。

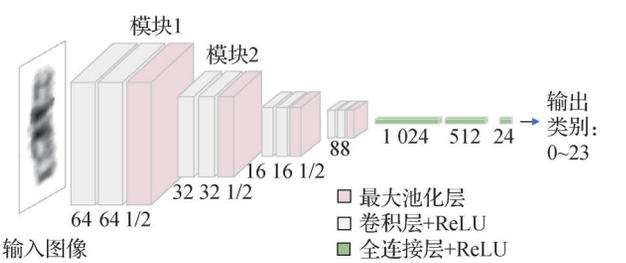



图 5　Deep CNN 架构的染色体分类模型（Sharma 等，2017）

Fig. 5　Chromosome classification model based on Deep CNN architecture（Sharma et al. , 2017）

Remya 等人(2020)在 Kiruthika 等人(2018)方法启发下,提出了 CNN 变体双模型 ChromNet1 和 ChromNet2,如图 6 所示。ChromNet1 是一个包含卷积层和池化层(或称采样层)构成的浅层网络模型。模型选取的 3 × 3 卷积核(小的感受野)不仅能使模型专注图像的细节信息,并且可以减缓后续参数量的剧增。激活函数选取线性整流单元(ReLU)而不是 Sigmoid 函数,能有效应对训练过程中出现的梯度饱和以及梯度消失问题。ChromNet2 网络在 ChromNet1 网络的基础上,增添了多层卷积模块和 dropout 方法(Srivastava 等,2014)。在私有标注的 21 423 条染色体构成的数据集上进行预训练和性能测试,ChromNet2 的分类准确率为 91.3%,ChromNet1 的分类效果在文献中没有提及。

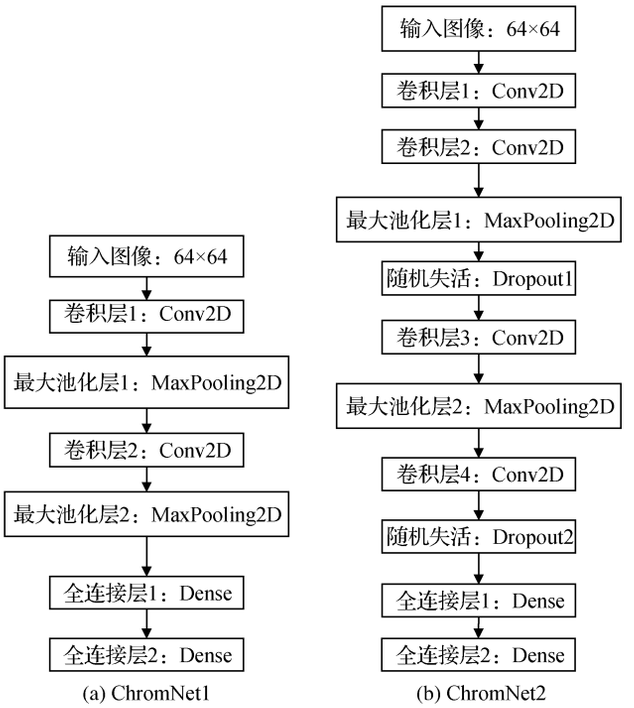



图 6　ChromNet1 和 ChromNet2(Remya 等,2020)

Fig. 6　ChromNet1 and ChromNet2(Remya et al., 2020)

((a) ChromNet1; (b) ChromNet2)

减少训练开销的常用策略是权值共享(weight share),这个策略在卷积神经网络中发挥了重要作用。经典 Le-Net 网络(Lecun 等,1998)一经提出就受到了学者们的广泛关注和深入研究,Zhang 等人(2018)提出 Vanilla CNN 网络就借鉴了这一结构。如图 7, 120 × 40 像素的染色体图像送入 Vanilla CNN 网络,模型拼接多个卷积层和池化层对输入信号进行加工,然后在连接层实现与输出结果之间的映射。每个卷积层都包含多个特征映射,其中每个特征映射是由多个神经元构成的多维平面,并通过卷积核提取特征。池化层(pooling)的别称为“汇合”层,其目的是基于局部相关性原理进行亚采样,从而既能减少数据量也能留住有效信息。Vanilla CNN 网络将原始染色体图像映射成高维特征向量,最后通过一个由 120 个神经元构成的连接层和输出层连接完成分类。其中,随机失活半数神经元(随机失活 dropout = 0.5)可以减少神经元之间复杂的共适应关系,增强网络的泛化能力。该方法在包含 10 304 个染色体图像的私有 G 带染色体数据集上的分类准确率为 92.5%。然而,数据集中样本分布不均衡,仅包含了垂直染色体,所以该模型的分类结果在遇到新的数据集时容易出现染色体的错误分类。

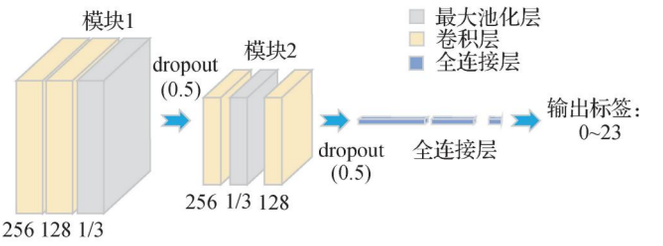



图 7　基于 Le-Net 网络变体的染色体分类模型(Zhang 等,2018)

Fig. 7　Chromosome classification model based on Le-Net network variant(Zhang et al., 2018)

深度神经网络能自然地以端到端的多层次方式将图像低、中、高层次特征(Zeiler 和 Fergus,2014)整合进分类器,并且通过堆积层的数量(深度)来丰富特征的“层次”。VGGNet(Visual Geometry Group network)网络(Simonyan 和 Zisserman,2015)探索了卷积神经网络的深度与其性能之间的关系,成功构筑了深度为 16 ~ 19 层的卷积神经网络,证明了通过小卷积核增加模型深度能在一定程度上影响网络最终的性能,使错误率大幅下降。为解决染色体分类问题,Hu 等人(2019)从 VGGNet 网络中得到启发,提出了如图 8 所示的染色体分类模型,相比基于手工特征的方法有着明显优势。模型设置了 6 个卷积层和 2 个全连接层,其构成的神经网络对 100 × 220 像素的染色体图像进行分类。图中的 Conv1 层专注抽取染色体边缘等低层次特征,而其余卷积层负责连续抽取高层次特征,如染色体条带特征、染色体着丝点特征等。原始图像输入至模型后得到 256 维特征,随后将特征集展平为 1 维表示并连接 FC1

层，随机失活 FC1 层的半数神经元连接至 FC2 层，最终完成染色体分类任务。堆叠多个小卷积核、使用池化操作和 dropout 技巧的加入，保证了模型具有较好的鲁棒性。该团队从 91 张染色体核型图中分割得到 4 184 条染色体作为实验数据集，评估结果显示，分类准确率为 93. 79% 。

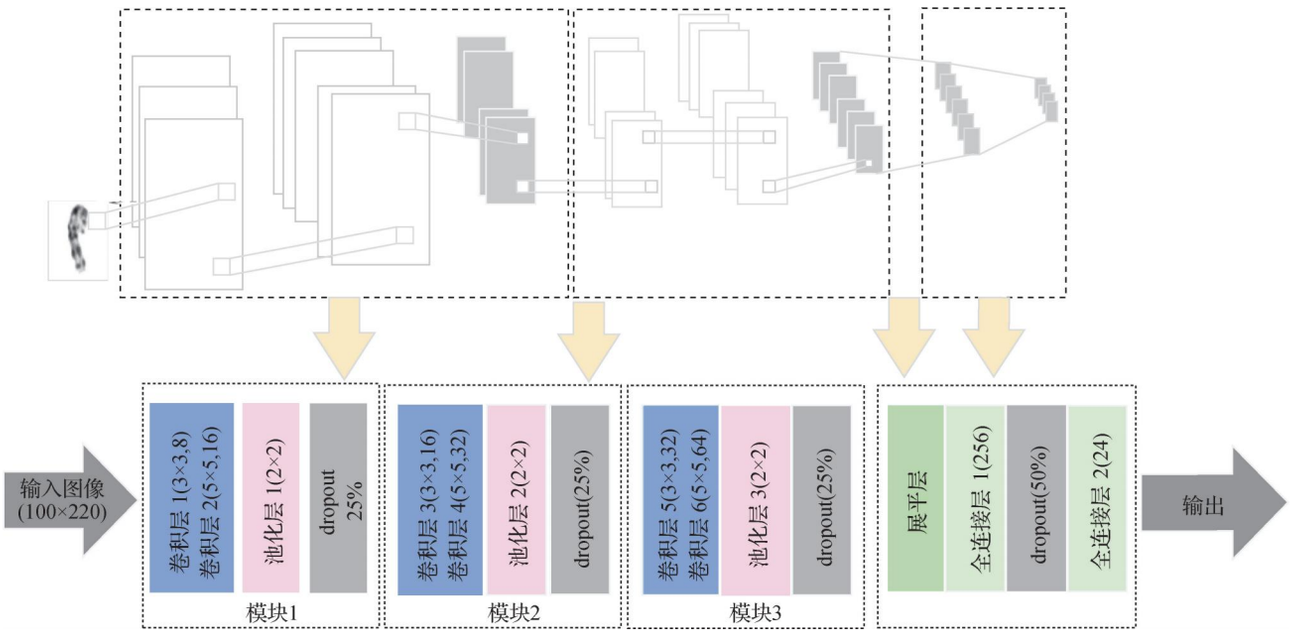



图 8　基于 VGGNet 架构的染色体分类模型(Hu 等，2019)

Fig. 8　The architecture of the proposed VGGNet for chromosome classification model(Hu et al. , 2019)

对于如何充分获取图像的多重特征，目前存在着从模型外部拼接不同尺度图像组合后作为输入端的改进思路。例如，将不同形状图像产生的复合信号输入模型以展开训练。李康等人(2020)认为，染色体形状多变性的客观因素和普通 CNN 模型存在特征提取不充分问题，会带来染色体分类精度不佳的困扰，提出了融合多图像特征的 Multi-input CNN 模型，借助 Mask R-CNN(He 等，2017)实现染色体的自动分割，如图 9 所示。染色体图像经几何优化操作得到优化图像，随后将原始图像和优化图像并行作为 Multi-input CNN 模型的输入信号。分类过程可精简为残差网络提取特征、融合多层次特征图、使用多层感知机(MLP)进行分类预测 3 个步骤。数据集源于医院提供的核型图中分割出的染色体实例，共 5 000 幅 128 × 128 尺寸的染色体图像，每张图片包含 1 条染色体。为了保证 Multi-input

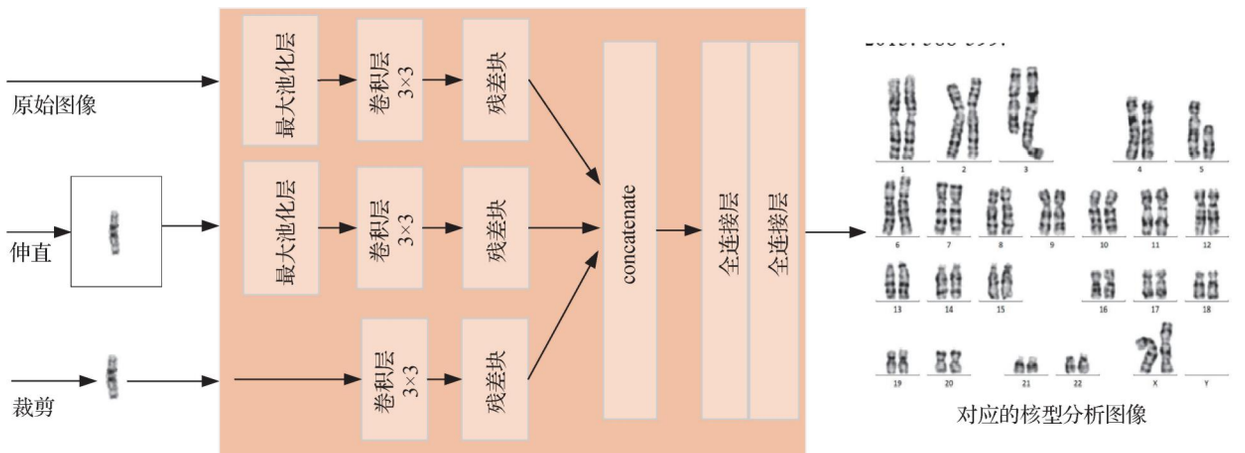



图 9　Multi-input CNN 网络架构(李康 等，2020)

Fig. 9　Multi-input CNN network architecture(Li et al. , 2020)

CNN 模型的效果,先训练3 个单独的 CNN 模块直到收敛,然后拼接 3 个子网络,固定 CNN 的参数且只训练顶层的分类器,最后解锁整个模型,并在一个很低的学习率(0.000 5)下进行微调,评估结果准确率为 95.6% 。

卷积神经网络是一种通过模拟生物大脑皮层结构而特殊设计的含有多隐层的人工神经网络。卷积层、池化层和激活函数是卷积神经网络的核心部分。卷积神经网络通过局部感受野、权重共享和降采样这 3 种策略,降低了网络模型的复杂度,因此增加网络深度进而提升网络的分类性能是积极有效的一种常规技术方案。

1.3.2　基于 ResNet 和 Siamese 的分类模型

为进一步提升网络的分类效果,一些学者从网络内部着手,提出了允许输入信息可以跨越后续网络层并进行传播(即残差网络)和允许两个输入共享编码层(即孪生网络)的基于卷积操作模型。

神经网络一味地增加深度和宽度会面临梯度消失和梯度爆炸导致的网络退化问题,具体表现为网络性能不再随深度增加而提升。为解决此类问题,He 等人(2016)提出残差网络,并在 ImageNet(Deng 等,2009)比赛中取得佳绩。残差网络允许输入信息可以跨越多个隐含层传播,并始终保持信息畅通状态,极大降低了网络复杂度,加速了网络训练过程,同时突破了由网络退化引起的深度限制。

因其通用性与实用性并存,张成成等人(2020)提出了如图 10 所示的染色体分类模型。该模型由 4 个残差子模块组成,每个子模块包含两层 1×1 卷积层(Lin 等,2013)和 3×3 卷积层。将 1 225 例受检人员的分裂中期染色体图像纳入数据集,并分为训练集 735 例、内部验证集 245 例和外部验证集 245 例;另取 50 例染色体图像资料开展人机对抗实验。分类结果显示,模型分类准确率为 91.22% 。同时,该模型每幅图像测试平均耗时仅需 0.06 s,远低于专业医师 62.5 s 的水准。该方法凭借跳跃连接的思想找最优路径,使分类过程耗时短、效率高,但数据量问题仍未得到有效解决。

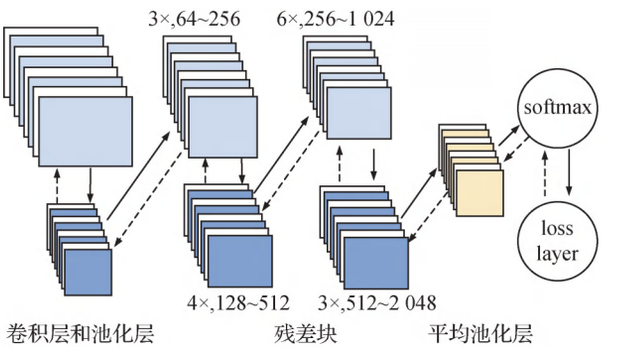



图 10　ResNet 网络的染色体分类模型(张成成 等,2020)

Fig. 10　Chromosome classification model of ResNet network (Zhang et al., 2020)

Sharma 等人(2018)为充分挖掘染色体图像中隐藏的条带信息,设计了融合卷积神经网络 ResNet-50、循环神经网络(recurrent neural network,RNN)和深度注意机制的染色体分类方法 Res-CRANN(residual convolutional recurrent attention neural network),如图 11 所示。具体步骤如下:首先,将标准化后的染色体图像(224×224 像素)经由深度卷积神经网络(主干网络为 ResNet-50)输出 $G \times H \times K$ 的高维特征图;然后,串行 RNN 模型逐步了解染色体条带序列信息;最后,利用残差网络的优点并通过回归层推断染色体的类别标签(0～23)。Res-CRANN 模型主动聚焦染色体深浅交替生成的亮带序列区域范围产生的序列信息,关注常规的深度卷积神经网络忽略的特征,提升了模型对染色体的识别能力,在 BiolmLab 数据集的分类准确率为 91.94% 。

对于如何更进一步提升卷积神经网络性能,目前的主流观点是从网络内部结构着手,通过增加网络深度和宽度来实现,但会出现网络训练参数量大

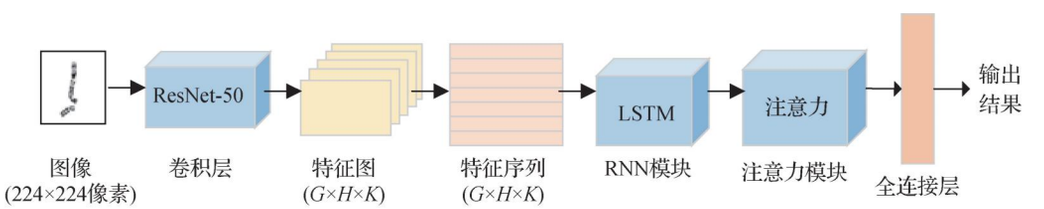



图 11　Res-CRANN 网络架构图(Sharma 等,2018)

Fig. 11　The proposed architecture of Res-CRANN network (Sharma et al., 2018)

幅增加和训练速度降低问题。为解决上述问题，Szegedy 等人(2015)从网中网(Lin 等,2013)中得到启发,提出了 Inception-V1 网络结构。与 Lin 等人(2013)方法类似,Szegedy 等人(2017)使用了类似结构,设计并提出了 Inception-ResNet 网络模型。

Lin 等人(2020b)借鉴了 Inception-ResNet 模型架构,提出一种多注意网络 CirNet 进行分类,如图 12 所示。与传统卷积神经网络不同,该网络以成熟的 Inception-ResNet 架构作为主干网络,附加深度可分离卷积和卷积核分解等优化策略,并增添了图像增强算法(chromosome data augmentation, CDA)以增强图像特征,同时使用图像自适应接口(image adaptive interface, IAI)将染色体图像规格统一化。CirNet 模型多方位了解和推断 2 990 条染色体,准确率为 95.98%,代码和数据集可从 GitHub 获取。

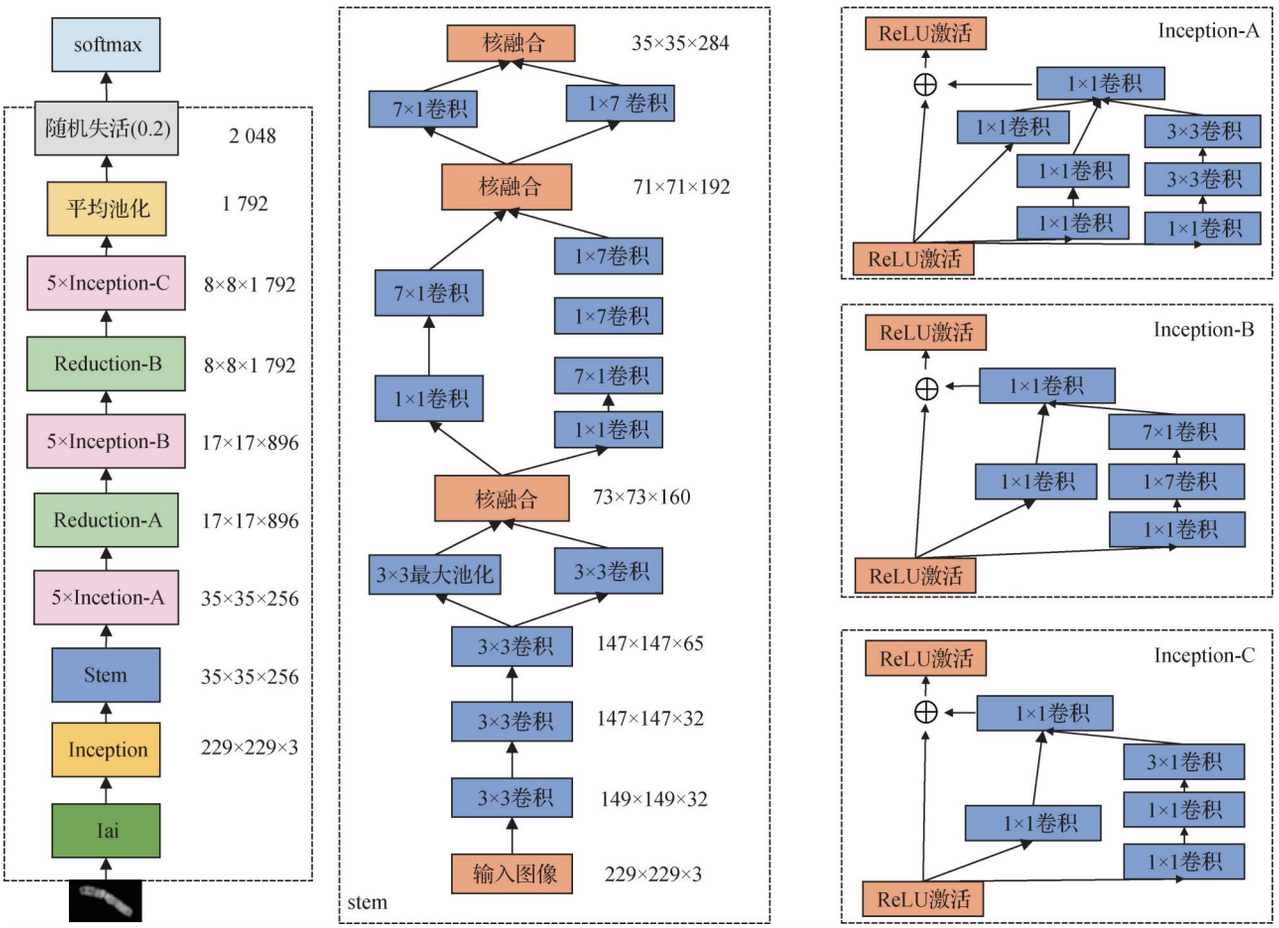



图 12　CirNet 网络的架构图(Lin 等,2020b)

Fig. 12　The architeture of CirNet network(Lin et al., 2020b)

He 等人(2016)首次提出,Xie 等人(2017)开发的批量残差网络 ResNeXt 体系结构是一种精简但高度模块化的网络体系结构,它执行一组变换,每个变换基于低维嵌入,其输出通过求和进行聚合。Lin 等人(2020a)(MixNet)在 ResNeXt 的基础上提出一种新颖的 MixNet 框架,如图 13 所示。模型可以有效地探测染色体图像的类别信息表征,最终在染色体分类任务中获得了满意的结果。MixNet 由双模块构成,染色体编码骨干网络和自适应报头。聚集残差架构构成的染色体编码骨干可以将染色体图像潜在特征通过卷积等方式编码成向量形式,是 MixNet 的染色体特征编码的核心模块;自适应报头通过最大池化层捕捉图像关键特征,同时利用全局平均池化层捕捉一般特征。在两个不同数据集下分别训练和测试,测试结果表明,MixNet 在包含 2 990 条染色体的临床 G 带数据集上的分类准确率为(98.73 ±0.81)%,在包含 5 474 条染色体的公共 Q 带数据集上的分类准确率为(96.50 ± 1.87)%,在两个数据集上准确率均超过 95%,且仅需少数迭代(训练时间不到 1 h)就能使模型结束训练阶段。

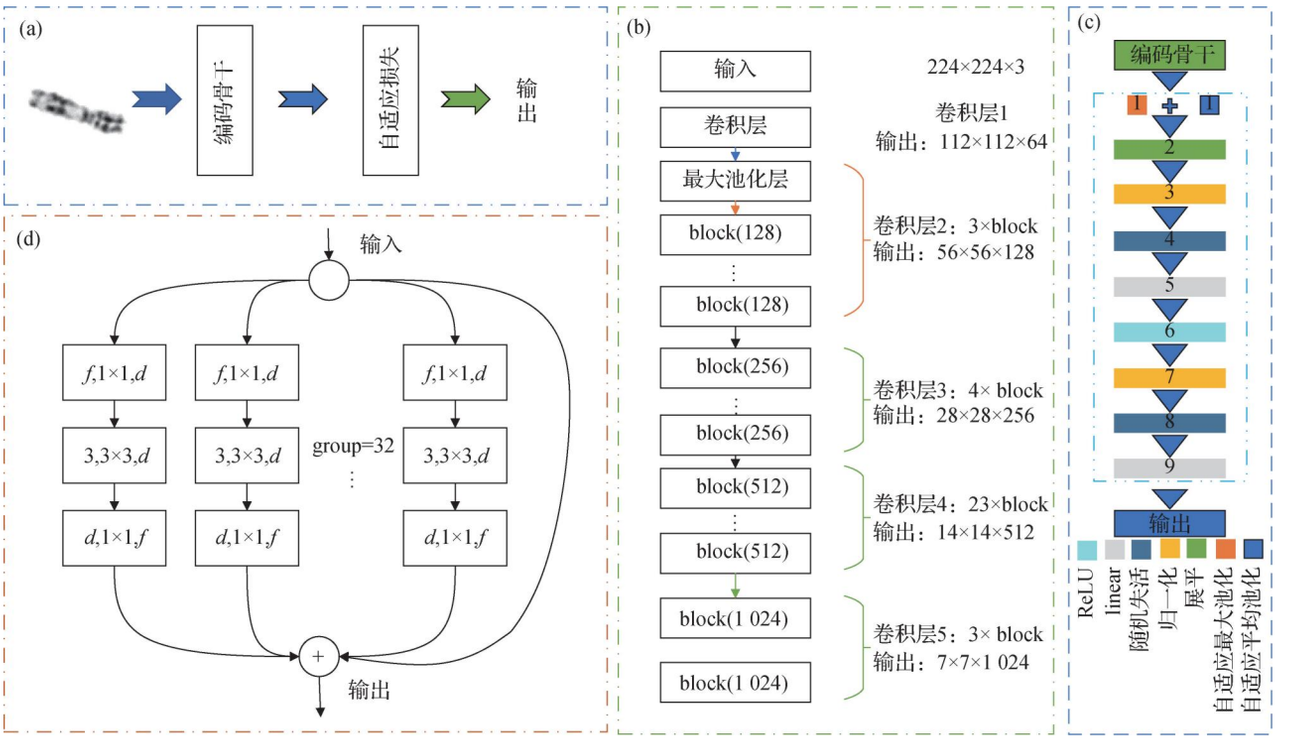



图 13　MixNet 网络架构图(Lin 等,2020a)

Fig. 13　The architeture of MixNet network(Lin et al., 2020a)

上述 4 种染色体分类研究都是在 ResNet 网络的基础上改进的模型结构。现有方法常通过各种图像预处理、扩张模型深度和宽度、引入各种优化技术和多模块串行等方法增强算法的鲁棒性和泛化能力,都取得了超过 90% 的分类效果。孪生网络 Siamese 最早由 Bromley 等人(1993)提出,是用来验证支票字迹与银行预留笔迹是否一致的一种方法。在此基础上,涌现出大量针对孪生网络结构的变体模型(Zagoruyko 和 Komodakis,2015),更有尝试将其应用于计算机视觉领域(Bertinetto 等,2016;Valmadre 等,2017)。Swati 等人(2017)提出将孪生网络用于染色体图像上的分类算法,使用中轴和众包(straightening via medial axis extraction and crowdsouring,SMAC)以及投影向量(straightening via projection vectors,SPV)的矫正方法来拉直染色体。拉直的染色体输入至由 Base-CNN 网络组建的孪生网络进行预训练,如图 14 所示。

孪生网络等同于连体的神经网络,连体通过共享权值实现。在低维空间下,孪生网络可以拉近同类染色体间的空间距离,拉长不同类别染色体之间的空间距离。由于训练数据量缺乏,该模型在染色体分类任务上的准确率为 84.6% 。

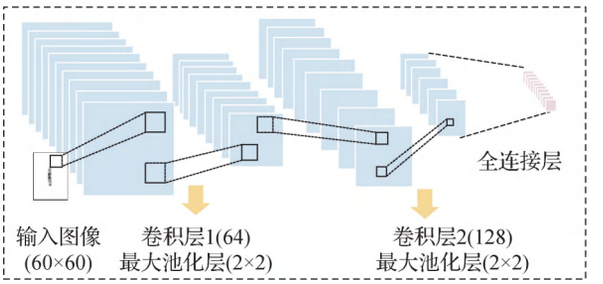



图 14　孪生网络中使用的 Base-CNN 架构(Swati 等,2017)

Fig. 14　The Base-CNN architecture utilized in Siamese network (Swati et al., 2017)

吕丹丹(2019)将孪生网络应用于染色体图像分类中,采用 Multi-Siamese 架构进行迁移学习以及增添空间金字塔池化层和多级卷积,如图 15 所示。

Multi-Siamese 网络模型引入空间金字塔池化层取代网络末端的池化层,解决了对输入数据固定大小的限制。为全面提取染色体特征,模型采用一种多级卷积特征提取结构,该结构能深层次地提取染色体特征,孪生网络模型精度进一步提升。该方法将提取的大小不同的特征图转换成固定大小的特征向量,送入全连接层形成一维向量,最后根据欧氏距离计算特征向量间的相似度。部分参数设定如下:初始学习率为 0.01,最大迭代次数为 $4\times10^4$,动量因子

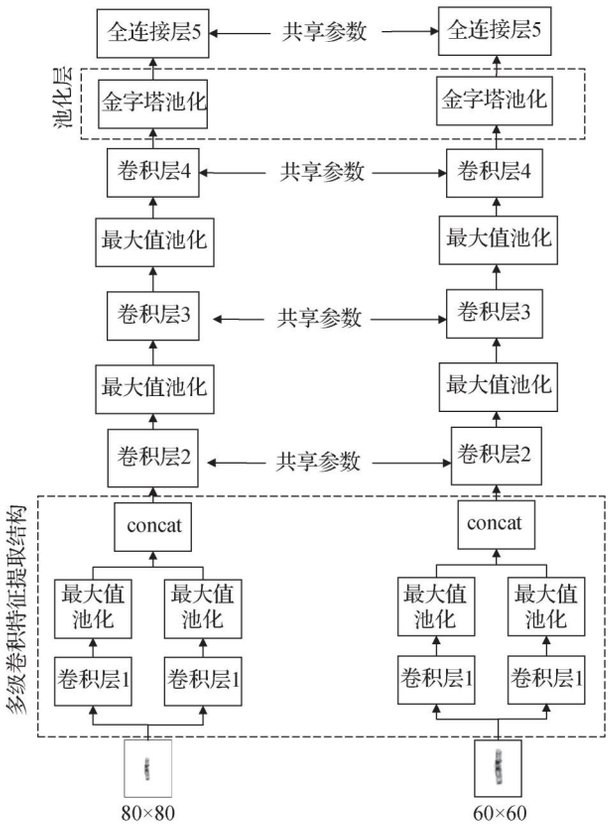



图 15　Multi-Siamese 网络架构(吕丹丹 等,2019)

Fig. 15　Multi-Siamese network architecture(Lyu et al. , 2019)

为 0.9。在 18 400 条染色体数据集的准确率为 85.32% 。

总体来看,基于 ResNet 网络的染色体分类模型的分类性能在 91% 以上,而基于 Siamese 网络的染色体分类模型性能远不如前者。原因可能是残差网络中引入的跳跃连接扩宽了网络模型的深度,使 ResNet 算法能比孪生网络模型学习到更多的图像细节信息,最终带来了较好的性能结果。

### 1.3.3　多任务学习的分类模型

多任务学习是一种推导迁移学习方法,主任务使用相关子任务的训练信号提高主任务的泛化性能,其强大的特征学习能力引起研究者关注,并将其应用于染色体分类任务中。

Qin 等人(2019)提出 Varifocal-Net 网络架构,同时执行染色体分类和染色体极性分类两项任务。Varifocal-Net 网络架构如图 16 所示,包含全局网络(global-scale network,G-Net)和局部网络(local-scale network,L-Net),并分为 3 个阶段逐步开展。首先,引入 G-Net 和 L-Net 网络分别进行全局和局部尺度的图像特征学习,同时加入残差学习和多任务学习策略促进图像的特征提取,改善分类效果。模型后面附加两个多层感知分类器,以融合多尺度

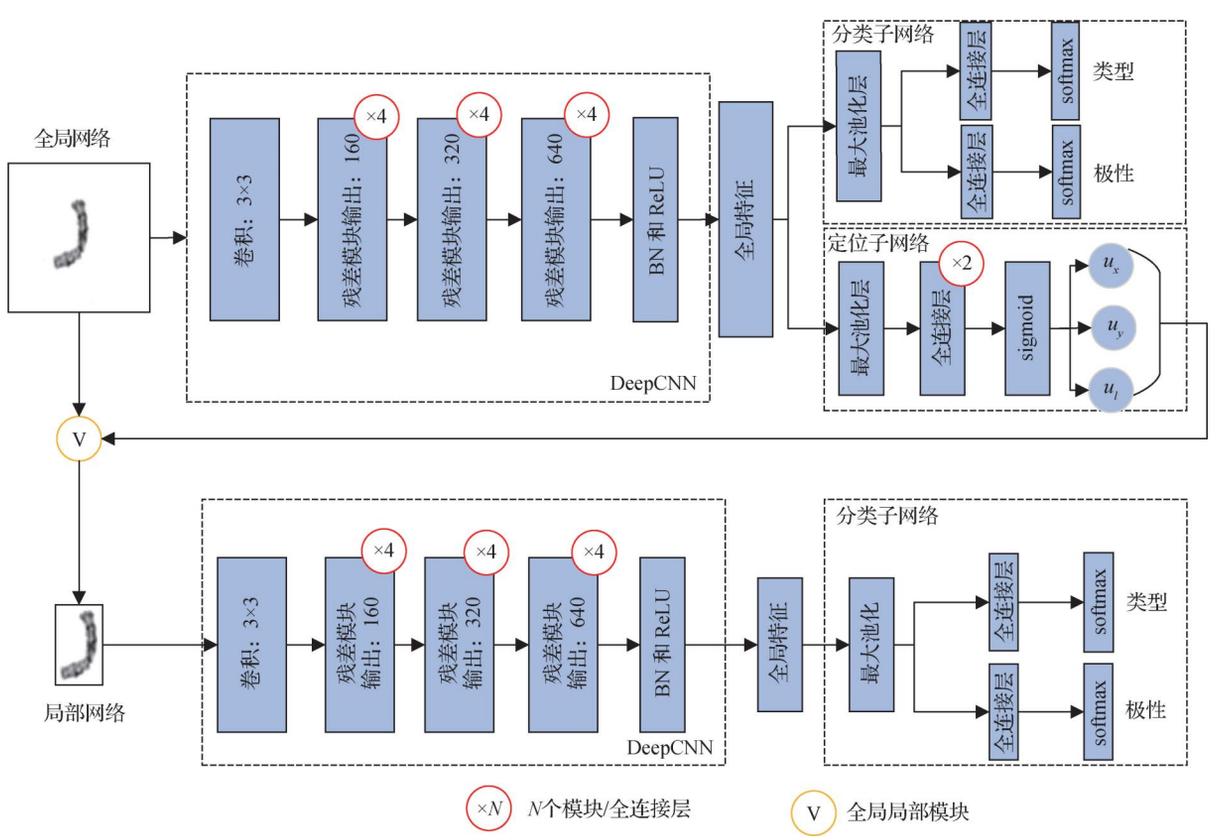



图 16　Varifocal-Net 架构图(Qin 等,2019)

Fig. 16　The architeture of Varifocal-Net(Qin et al. , 2019)

特征信息。最后,模型末端引入核型分析的领域知识和调度策略,得到染色体分类结果。Varifocal-Net模型融入染色体先验知识进行分类,即每一个健康受检者都有46条染色体(男性分为24类,女性分为23类),异常受检者通常情况下每一类染色体数目在1～3之间(极其罕见的除外)。评估结果显示,Varifocal-Net网络在1 909幅核型图中分割得到的87 831条染色体中,分类准确率为99.2%。

Zhang等人(2021)提出一种融入交错和多任务学习网络的染色体分类架构,如图17所示。该模型包含3个阶段。第1阶段,改进的HRNet(high resolution network)网络(Sun等,2019)构成的交错网络机制允许网络选择性地增强有用的特征并抑制无用的特征,甚至不仅限于高分辨率特征,从而实现特征的有效学习。第2阶段,将前一阶段学习到的高分辨率特征信息(提供的关键点信息以定位染色体)送至卷积神经子网络进行染色体联合检测,而融合剩余特征信息喂送入多层感知子网络实现染色体类型和极性的分类。最后一个阶段执行拉直染色体的任务。在收集到的32 810条染色体构成的数据集上进行测试,5折交叉验证结果显示,模型的平均分类准确率为98.1%。

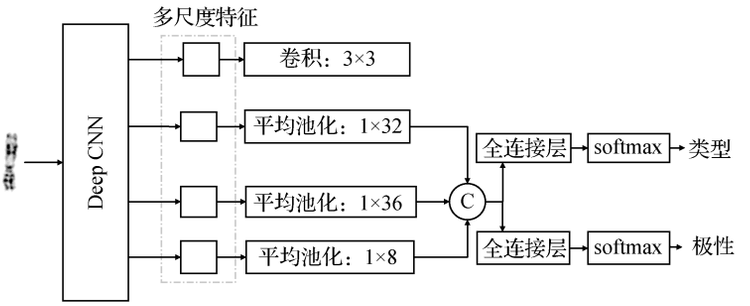



图17　多任务分类架构图(Zhang等,2021)

Fig. 17　Multi-task classification architecture diagram(Zhang et al., 2021)

本文精选并阐述了近5年基于深度学习技术的染色体分类代表性算法,并按照网络设计思路介绍了模型的演变过程,总结了现有染色体分类模型相关细节信息,如表3所示。

**表3　深度学习下的染色体分类代表性算法**

**Table 3　Chromosome classification representative algorithm under deep learning**

| 方法 | 模型 | 分辨率/像素 | 数据量 | ACC/% |
|---|---|---|---|---|
| Sharma等人(2017) | Deep CNN | 120×120 | 1 800 | 86.70 |
| Zhang等人(2018) | Vanilla CNN | 120×40 | 10 304 | 92.50 |
| Hu等人(2019) | VGGNet | 100×220 | 4 184 | 93.79 |
| Remya等人(2020) | ChromNet2 | 64×64 | 21 423 | 91.30 |
| 李康 等人(2020) | Multi-CNN | 128×128 | 5 000 | 95.60 |
| Sharma等人(2018) | ResNet-50 | 224×224 | 5 256 | 91.94 |
| 张成成 等人(2020) | ResNet-101 | 224×224 | 1 275 | 91.22 |
| Lin等人(2020b) | CirNet | 224×224 | 2 990 | 95.98 |
| Lin等人(2020a) | MixNet | 224×224 | 5 474 | 98.73 |
| Swati等人(2017) | Siamese | 60×60 | 1 740 | 84.60 |
| 吕丹丹(2019) | Multi-Siamese | 60×60 | 18 400 | 85.32 |
| Qin等人(2019) | Varifocal-Net | – | 87 831 | 99.20 |
| Zhang等人(2021) | 多任务网络 | – | 32 910 | 98.10 |

注:“–”表示无法获取。

## 2　性能评估

### 2.1　评估使用的数据库

染色体数据库的建立过程如图18所示,主要包括中期细胞图像数据获取、染色体图像筛选和染色体实例分割获取3个阶段。

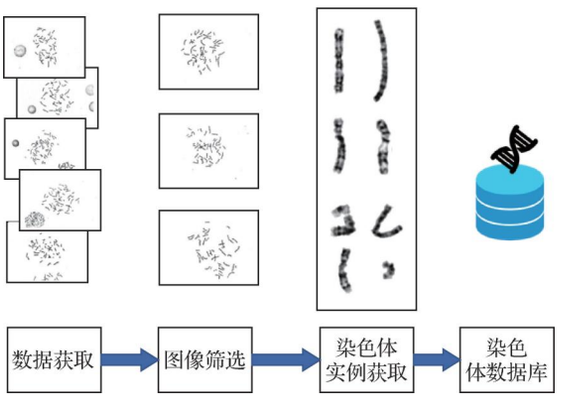



图18　染色体数据库的建立过程

Fig. 18　Chromosome database establishment process

#### 2.1.1　中期细胞图像数据获取

图像数据取自广东省妇幼保健院的408名患者

的 6 587 幅染色细胞显微图像。图像分辨率统一为 1 360 × 1 024 像素,96 dpi。图像文件的命名方式用于区分不同的个人 ID。例如,P21000. 001. A. jpg 至 P21000. 015. A. jpg 表示一个受检者为 P21000 个体的外周血中采集到的 15 幅细胞图像。每位患者至少收集 8 张中期细胞图片,有关隐私信息(姓名、年龄等)均已删除。

2. 1. 2　细胞图像筛选

经显微镜获取的中期细胞图像可能存在质量问题,如细胞杂质、噪声和拍摄模糊等,如图 19 所示。为此,组织 16 位广东省妇幼保健院经验丰富的细胞遗传学专家对低质量细胞图像开展人工剔除。最终在图像选择阶段收集到 2 795 幅高质量的细胞图像。

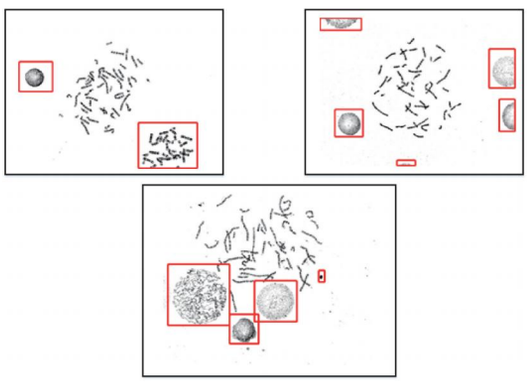


图 19　带杂质细胞图像示例

Fig. 19　Example of an image of a cell with impurities

2. 1. 3　染色体实例获取

经中期细胞图像收集和图像质量评估两个阶段后,细胞图像质量得以保证。对此,核型分析师依据第三方软件对细胞图像进行染色体实例提取,将每幅中期细胞图像分割得到 46 条染色体,最终形成染色体数据库。

在染色体实例提取阶段,首先从核型中分割出单个染色体实例,并将其放置在 300 × 300 像素的空白图像中心作为染色体样本。然后将图像保存到相应的文件夹中,最后由 24 个文件夹和 126 543 个染色体样本组成数据集。本文提到的临床染色体数据库是依托广东省妇女儿童医院医学遗传中心获取的临床染色体显微镜图像数据。染色体核型图像中的染色体实例由细胞遗传学专家按标签信息正确归类并调整为正确的极性方向。

## 2. 2　参评模型及其环境配置

2. 2. 1　参与评估的主流模型

为了测试现有染色体分类算法之间的性能,本文在构建的临床数据集上实现并评估现有算法,从表 3 中精选出 6 种主流算法开展评估实验,如表 4 所示。

**表 4　参与评估的分类算法比较**

**Table 4　Comparison of classification algorithms participating in the evaluation**

| 方法 | 模型 | 准确率 |
|---|---|---|
| Sharma 等人(2017) | Deep CNN | 0. 867 |
| Zhang 等人(2018) | Vanilla CNN | 0. 925 |
| Hu 等人(2019) | CNN | 0. 937 |
| Remya 等人(2020) | ChromNet2 | 0. 913 |
| Lin 等人(2020b) | CirNet | 0. 959 |
| Lin 等人(2020a) | MixNet | 0. 987 |

2. 2. 2　实验设置

实验采用随机分层抽样将 126 453 条染色体组成的临床数据集分为训练数据集(80%)、验证数据集(10%)和测试数据集(10%)(Pedregosa 等,2011)。参与评估的模型在相同的训练和验证数据集上训练,在同一测试数据集上评估。挑选的主流模型均在 Pytorch(Paszke 等,2017)框架上开发。实验环境为 CPU E5-2620 v4 @ 2. 1 kHz、64 GB 的 RAM 内存和 4 个 Nvidia GeForce RTX 2080Ti 的 11019MiB GPU,CentOS 操作系统。

训练过程如下:首先,参评模型从 ImageNet(Deng 等,2009)分类任务中进行预训练和迁移。然后,利用周期性学习率(cycle learning rate, CLR)(Smith,2018)方法,在临床数据集中逐步调整各模型的表现。受限于 GeForce RTX 2080Ti GPU 的运行内存容量,实验的批处理大小设置为 32(batch_size = 32),损失函数通过 $\alpha = 0.1$ 的标记平滑(Szegedy 等,2016)的交叉熵损失函数来调整,学习速率设置为 $10^{-4}$($lr = 10^{-4}$),最大训练迭代次数的超参数设置为 500。当验证损失在 5 个 epoch 没有下降时,提前终止训练过程。最后,训练结束时呈现方法的最佳训练权重恢复到相应模型。

## 2. 3　实验结果与分析

采用准确率、精确率、召回率和F1分数作为定

量评估指标。实验结果如表 5 所示。绝对提升是指评估性能与原有方法性能的绝对差值,相对提升是指绝对提升性能数值占原有方法性能数值的比率。具体提升如图 20 所示。

**表 5　参评模型在构建的数据集上的实验结果**

**Table 5　Experimental results of existing methods on the proposed dataset**

| 方法 | 参数量/MB | 原有准确率/% | 精确率/% | 召回率/% | F1 分数/% | 评估精确率/% | 绝对提升/% | 相对提升/% |
|---|---|---|---|---|---|---|---|---|
| Deep CNN | 2.62 | 86.70 | 91.66 | 90.69 | 91.03 | 92.09 | **+5.39** | **+6.21** |
| Vanilla CNN | 31.82 | 92.50 | 95.40 | 94.74 | 95.02 | 95.48 | +2.98 | +3.22 |
| CNN | 30.58 | 93.79 | 94.33 | 93.59 | 93.89 | 94.70 | +0.91 | +0.97 |
| ChromNet2 | 13.02 | 91.30 | 95.06 | 94.62 | 94.81 | 95.23 | +3.93 | +4.30 |
| CirNet | 207.3 | 95.98 | 98.57 | 98.21 | 98.38 | 98.64 | +2.66 | +2.77 |
| MixNet | 33.11 | **98.73** | **98.91** | **98.71** | **98.81** | **98.92** | +0.19 | +0.19 |

注:加粗字体表示各列最优结果。

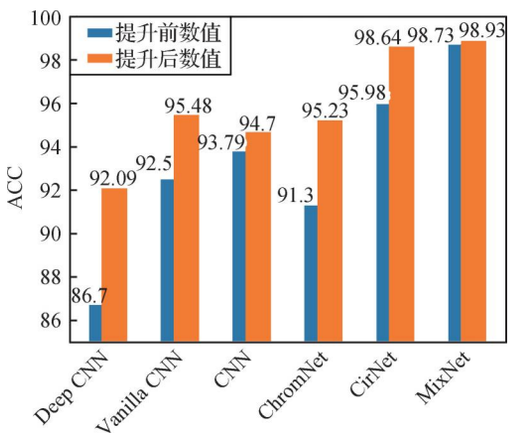



图 20　不同算法的性能提升结果

Fig. 20　Performance improvement results of different algorithms

从表 5 可以看出,1)参数量是深度学习分类模型中的一个重要因素,参数量的改变带来的是网络分类效果的巨大变化。对染色体分类任务来说,参数量为 2.62 MB 的 Deep CNN 轻量化模型准确率为 86.7%,参数量为 207.3 MB 的 CirNet 模型的准确率为 95.98%,表明参数量与模型分类性能存在一定的相关性,但不包含 MixNet 模型。2)评价指标的不同选取方式可以客观评估模型的真实效能。参数量为 33.11 MB 的 MixNet 模型在精确率(98.91%)、准确率(98.92%)、召回率(98.71%)和 F1 值(98.81%)等 4 类分类结果上都取得了最好结果,说明 MixNet 模型分类稳定性较好。模型性能优劣不取决于参数规模的大小,而取决于网络结构设计过程的合理性与否。通过巧妙设计,在降低网络复杂性的同时,可以在同等条件下提升分类性能。3)所有模型均有不同程度上的性能提升。Deep CNN 模型提升幅度最大。在本文选取的数据集上重新评估,Deep CNN 网络的识别效果从 86.7% 提升至 92.09%,有效缩短了与其他模型的差距。Lin 等人(2020a)提出的基于改进残差神经网络的 MixNet 的分类精度为 98.73%,有 0.19% 的小幅度提升。

从表 4、表 5 和图 20 可以看出,深度学习方法非常依赖大数据的学习,训练数据的多少决定了网络模型的效果。大规模临床染色体数据集对于评估现有染色体分类方法的真实性能至关重要。基于本文提出的临床数据集,评估实验能客观了解现有方法的分类水准,并且结果显示所有算法的分类性能都有不同程度上的数值提升,绝对提升幅度为 0.19% ～ 5.39%,相对提升幅度为 0.19% ～ 6.21%,其中数据集起到了关键性作用。结果表明,本文提出的大规模数据集在各个模型上是有效的。

CirNet 和 MixNet 模型无论是初始性能还是提升后的分类效果都明显好于其他算法模型,上述两种模型的原型都是残差神经网络。运用 ResNet 模型开展染色体分类研究的相关算法设计可以取得不错效果,原因可能是残差连接能扩宽网络的深度和宽度,增加训练参数量,提高网络的空间复杂度,增强网络结构本身的表达能力,进而得到最为理想的分类水平。此外,大规模染色体数据量缓解了模型过拟合现象的发生。

现有方法评估后最高分类准确率为 98.92%,但距高精度的临床应用仍有提升空间。因为 1% 的

精度缺失导致的某个误诊可能使一个生命无辜放弃,因此针对网络结构的改进和大型具有代表性的临床染色体数据库的构建,是染色体分类研究在很长一段时间需要解决的重点问题。

## 3 结 论

本文总结了现有染色体分类方法难以在临床实践中广泛应用的两个关键挑战,这也是在大规模数据集下对主流模型进行性能评估的出发点。第 1 个挑战是目前大多数染色体分类方法的实现细节不够具体,私有数据集使不同模型之间无法进行公平比较,现有研究不可直接使用,现有方法很难复制、验证和应用。第 2 个挑战是当前方法是在小规模私有数据集上训练和测试的,性能不够客观。

为此,本文通过构建大规模染色体分类数据集和选取主流模型在该数据集上进行性能评估对比实验,解决了上述两个挑战。首先,详细阐述了近几年采用深度学习技术解决染色体分类任务的相关算法,介绍了模型结构、模型演变以及模型之间的联系。其次,选择 6 种主流染色体分类算法,将 6 种网络的预训练模型在 126 453 条染色体临床数据集上统一进行性能评估和比对实验,从不同维度分析效果提升的原因。

本文主要关注了深度学习中的染色体分类识别算法,并开展评估实验,证明了染色体分类模型的性能优劣与训练数据量有关、不同染色体分类算法之间能进行客观的评估,以及临床应用上存在更为适合染色体分类应用的网络结构。大规模染色体临床数据集对现有模型分类性能具有明显的推动作用,参评模型都有不同程度的性能提升,提升幅度为 0.19%～6.21% 。算法选择上存在对网络结构选择上的侧重,选择 ResNet 变体的网络结构能在染色体分类识别任务中获得更为可观的分类效果。

深度学习算法的训练需要依赖大量数据,染色体图像数据提供了染色体类别特征独特信息的成像特点,需要遗传学专家手工获取和标注,相当费时耗力。未来在基于深度学习的染色体分类研究中,不仅需要打通数据壁垒,而且需要采用数据增广技术扩大数据量,需要持续开展数据自动化标注方面的研究。随着超分辨算法的出现,基于染色体图像的重建是构建高质量临床染色体数据集的重要手段。可解释性对于以神经网络为主的方法相当重要,需要继续开展深度神经网络在染色体分类中的可解释性研究。轻量级深度神经网络如何部署到便携式医疗设备中也是一个值得研究的方向。

深度学习在医学影像中的染色体分类任务中取得了明显的研究进展,但仍需要从方法、应用等方面进一步深入研究。现有研究成果距离临床实际应用的目标还有较大差距,需要继续提高模型精度,完善结果的可解释性。

## 参考文献(References)

Abid F and Hamami L. 2018. A survey of neural network based automated systems for human chromosome classification. Artificial Intelligence Review, 49(1): 41-56 [DOI: 10.1007/s10462-016-9515-5]

Bertinetto L, Valmadre J, Henriques J F, Vedaldi A and Torr P H S. 2016. Fully-convolutional Siamese networks for object tracking//Proceedings of 2016 European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 850-865 [DOI: 10.1007/978-3-319-48881-3_56]

Bromley J, Bentz J W, Bottou L, Guyon I, Lecun Y, Moore C, Säckinger E and Shah R. 1993. Signature verification using a "Siamese" time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 7(4): 669-688 [DOI: 10.1142/S0218001493000339]

Deng J, Dong W, Socher R, Li L J, Li K and Li F F. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 248-255 [DOI: 10.1109/cvpr.2009.5206848]

Grisan E, Poletti E and Ruggeri A. 2009. Automatic segmentation and disentangling of chromosomes in Q-band prometaphase images. IEEE Transactions on Information Technology in Biomedicine, 13(4): 575-581 [DOI: 10.1109/TITB.2009.2014464]

He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hu X, Yi W L, Jiang L, Wu S J, Zhang Y, Du J Q, Ma T Y, Wang T and Wu X M. 2019. Classification of metaphase chromosomes using deep convolutional neural network. Journal of Computational Biology, 26(5): 473-484 [DOI: 10.1089/cmb.2018.0212]

Jin X, Wen K, Lyu G F, Shi J, Chi M X, Wu Z and An H. 2020. Survey on the applications of deep learning to histopathology. Journal of Image and Graphics, 25(10): 1982-1993 (金旭, 文可, 吕国锋, 石军, 迟孟贤, 武铮, 安虹. 2020. 深度学习在组织病理学中的应用综述. 中国图象图形学报, 25(10): 1982-1993) [DOI: 10.11834/jig.200460]

Jindal S, Gupta G, Yadav M, Sharma M and Vig L. 2017. Siamese networks for chromosome classification//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. [s. l.]: IEEE: 72-81

Khan S, DSouza A, Sanches J and Ventura R. 2012. Geometric correction of deformed chromosomes for automatic karyotyping//Proceedings of 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. San Diego, USA: IEEE: 4438-4441 [DOI: 10.1109/EMBC.2012.6346951]

Kiruthika P, Jayanthi K B and Nirmala M. 2018. Classification of metaphase chromosomes using deep learning neural network//Proceedings of the 4th International Conference on Frontiers of Signal Processing. Poitiers, France: IEEE: 110-114 [DOI: 10.1109/ICFSP.2018.8552042]

Kubola K and Wayalun P. 2018. Automatic determination of the G-band chromosomes number based on geometric features//Proceedings of the 15th International Joint Conference on Computer Science and Software Engineering. Nakhonpathom, Thailand: IEEE: 1-5 [DOI: 10.1109/JCSSE.2018.8457330]

Lecun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324 [DOI: 10.1109/5.726791]

Li K, Xie N, Li X and Tan K. 2020. Statistical karyotype analysis using CNN and geometric optimization. Journal of Nanjing University (Natural Science), 56(1): 116-124 (李康, 谢宁, 李旭, 谭凯. 2020. 基于卷积神经网络和几何优化的统计染色体核型分析方法. 南京大学学报 (自然科学), 56(1): 116-124) [DOI: 10.13232/j.cnki.jnju.2020.01.013]

Lin C C, Zhao G A, Yin A H, Guo L, Chen H B and Zhao L. 2020a. MixNet: a better promising approach for chromosome classification based on aggregated residual architecture//Proceedings of 2020 International Conference on Computer Vision, Image and Deep Learning. Chongqing, China: IEEE: 313-318 [DOI: 10.1109/CVIDL51233.2020.00-79]

Lin C C, Zhao G S, Yang Z R, Yin A H, Wang X M, Guo L, Chen H B, Ma Z H, Zhao L, Luo H Y, Wang T X, Ding B C, Pang X W and Chen Q R. 2020b. CIR-Net: automatic classification of human chromosome based on inception-resnet architecture. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19(3): 1285-1293 [DOI: 10.1109/TCBB.2020.3003445]

Lin C C, Zhao G S, Yin A H, Ding B C, Guo L and Chen H B. 2020. AS-PANet: a chromosome instance segmentation method based on improved path aggregation network architecture. Journal of Image and Graphics, 25(10): 2271-2280 (林成创, 赵淦森, 尹爱华, 丁笔超, 郭莉, 陈汉彪. 2020. AS-PANet: 改进路径增强网络的重叠染色体实例分割. 中国图象图形学报, 25(10): 2271-2280) [DOI: 10.11834/JIG.200236]

Lin M, Chen Q and Yan S C. 2013. Network in network [EB/OL]. [2021-07-01]. https://arxiv.org/pdf/1312.4400.pdf

Lyu D D. 2019. Research on automatic segmentation arrangement and abnormal diagnosis of human chromosome images. Hangzhou: Zhejiang Sci-Tech University (吕丹丹. 2019. 人染色体图像自动分割排列与异常诊断研究. 杭州: 浙江理工大学) [DOI: 10.27786/d.cnki.gzjlg.2019.000206]

McGowan-Jordan J, Simons A and Schmid M. 2016. ISCN 2016: An International System for Human Cytogenomic Nomenclature (2016). New York, USA: S. Karger [DOI: 10.1159/isbn.978-3-318-06861-0]

Moradi M, Setarehdan S K and Ghaffari S R. 2003. Automatic landmark detection on chromosomes' images for feature extraction purposes//The 3rd International Symposium on Image and Signal Processing and Analysis. Rome, Italy: IEEE: 567-570 [DOI: 10.1109/ISPA.2003.1296960]

Oskouei B C and Shanbehzadeh J. 2010. Chromosome classification based on wavelet neural network//Proceedings of 2010 International Conference on Digital Image Computing: Techniques and Applications. Sydney, Australia: IEEE: 605-610 [DOI: 10.1109/DICTA.2010.107]

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z M, Desmaison A, Antiga L and Lerer A. 2017. Automatic differentiation in PyTorch//Proceeding of the 31st Conference on Neural Information Processing Systems. Long Beach, USA: [s. n.]

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay É. 2011. Scikit-learn: machine learning in Python. The Journal of Machine Learning Research, 12: 2825-2830

Poletti E, Grisan E and Ruggeri A. 2008. Automatic classification of chromosomes in Q-band images//Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vancouver, Canada: IEEE: 1911-1914 [DOI: 10.1109/IEMBS.2008.4649560]

Poletti E, Ruggeri A and Grisan E. 2011. An improved classification scheme for chromosomes with missing data//Proceedings of 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Boston, USA: IEEE: 5072-5075 [DOI: 10.1109/IEMBS.2011.6091256]

Popescu M, Gader P, Keller J, Klein C, Stanley J and Caldwell C. 1999. Automatic karyotyping of metaphase cells with overlapping chromosomes. Computers in Biology and Medicine, 29(1): 61-82 [DOI: 10.1016/S0010-4825(98)00040-7]

Qin Y L, Wen J, Zheng H, Huang X L, Yang J, Song N, Zhu Y M,

Wu L Q and Yang G Z. 2019. Varifocal-Net: a chromosome classification approach using deep convolutional networks. IEEE Transactions on Medical Imaging, 38(11): 2569-2581 [DOI: 10.1109/TMI.2019.2905841]

Remya R S, Hariharan S, Vinod V, Fernandez D J W, Ajmal N M and Gopakumar C. 2020. A comprehensive study on convolutional neural networks for chromosome classification//Proceedings of 2020 Advanced Computing and Communication Technologies for High Performance Applications. Cochin, India: IEEE: 287-292 [DOI: 10.1109/ACCTHPA49271.2020.9213238]

Rungruangbaiyok S and Phukpattaranont P. 2010. Chromosome image classification using a two-step probabilistic neural network. Songklanakarin Journal of Science and Technology, 32(3): 255-262

Shaffer L G, McGowan-Jordan J and Schmid M. 2013. ISCN 2013: an international system for human cytogenetic nomenclature (2013). Karger Medical and Scientific Publishers

Sharma M, Saha O, Sriraman A, Hebbalaguppe R, Vig L and Karande S. 2017. Crowdsourcing for chromosome segmentation and deep classification//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE: 786-793 [DOI: 10.1109/CVPRW.2017.109]

Sharma M, Swati and Vig L. 2018. Automatic chromosome classification using deep attention based sequence learning of chromosome bands//Proceedings of 2018 International Joint Conference on Neural Networks. Rio de Janeiro, Brazil: IEEE: 1-8 [DOI: 10.1109/IJCNN.2018.8489321]

Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2021-07-01]. https://arxiv.org/pdf/1409.1556.pdf

Smith L N. 2018. A disciplined approach to neural network hyper-parameters: part 1——learning rate, batch size, momentum, and weight decay [EB/OL]. [2021-07-01]. https://arxiv.org/pdf/1803.09820.pdf

Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R. 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1): 1929-1958

Sun K, Xiao B, Liu D and Wang J D. 2019. Deep high-resolution representation learning for human pose estimation//Proceedings of 2919 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5686-5696 [DOI: 10.1109/CVPR.2019.00584]

Swati, Gupta G, Yadav M, Sharma M and Vig L. 2017. Siamese networks for chromosome classification//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. Venice, Italy: IEEE: 72-81 [DOI: 10.1109/ICCVW.2017.17]

Szegedy C, Ioffe S, Vanhoucke V and Alemi A A. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI Press: 4278-4284

Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 1-9 [DOI: 10.1109/cvpr.2015.7298594]

Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z. 2016. Rethinking the inception architecture for computer vision//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2818-2826 [DOI: 10.1109/CVPR.2016.308]

Tian J X, Liu G C, Gu S S, Ju Z J, Liu J G and Gu D D. 2018. Deep learning in medical image analysis and its challenges. Acta Automatica Sinica, 44(3): 401-424 (田娟秀, 刘国才, 谷珊珊, 鞠忠建, 刘劲光, 顾冬冬. 2018. 医学图像分析深度学习方法研究与挑战. 自动化学报, 44(3): 401-424) [DOI: 10.16383/j.aas.2018.c170153]

Valmadre J, Bertinetto L, Henriques J, Vedaldi A and Torr P H S. 2017. End-to-end representation learning for correlation filter based tracking//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5000-5008 [DOI: 10.1109/CVPR.2017.531]

Xie S N, Girshick R, Dollár P, Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5987-5995 [DOI: 10.1109/CVPR.2017.634]

Zagoruyko S and Komodakis N. 2015. Learning to compare image patches via convolutional neural networks//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 4353-4361 [DOI: 10.1109/CVPR.2015.7299064]

Zeiler M D and Fergus R. 2014. Visualizing and understanding convolutional networks//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 818-833 [DOI: 10.1007/978-3-319-10590-1_53]

Zhang C C, Song J P, Xu S Q, Li H, Xu R H, Wang X Y, Zhang L, You Q J, Zhang K and Lin H T. 2020. Application research on classification of metaphase chromosomes based on deep convolutional neural networks. Chinese Journal of New Clinical Medicine, 13(2): 123-126 (张成成, 宋婕萍, 徐淑琴, 李卉, 徐闰红, 王小艳, 张立, 游齐靖, 张凯, 林浩添. 2020. 基于深度卷积神经网络对中期染色体分类的应用研究. 中国临床新医学, 13(2): 123-126) [DOI: 10.3969/j.issn.1674-3806.2020.02.04]

Zhang J P, Hu W J, Li S Y, Wen Y F, Bao Y, Huang H F, Xu C M and Qian D H. 2021. Chromosome classification and straightening based on an interleaved and multi-task network. IEEE Journal of Biomedical and Health Informatics, 25(8): 3240-3251 [DOI: 10.1109/JBHI.2021.3062234]

Zhang W B, Song S F, Bai T M, Zhao Y X, Ma F, Su J L and Yu L M. 2018. Chromosome classification with convolutional neural network based deep learning//Proceedings of the 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. Beijing, China: IEEE: 1-5 [DOI: 10.1109/CISP-BMEI.2018.8633228]

Zhou X Y, Guan W J, Ma Y H and Liu G L. 2004. Chromosome banding technology and its significance in biological research. Shanghai Journal of Animal Husbandry and Veterinary Medicine, (5): 2-3 (周雪雁, 关伟军, 马月辉, 刘桂林. 2004. 染色体显带技术及其在生物学研究中的意义. 上海畜牧兽医通讯, (5): 2-3) [DOI: 10.3969/j.issn.1000-7725.2004.05.001]

## 作者简介

易序晟,男,硕士研究生,主要研究方向为计算机视觉、医疗人工智能。E-mail:xushengyi@m.scnu.edu.cn

赵淦森,通信作者,男,教授,博士生导师,主要研究方向为云计算、信息安全、医疗人工智能与区块链。E-mail:gzhao@m.scnu.edu.cn

尹爱华,女,主任医师,教授,博士生导师,主要研究方向为产前诊断。E-mail:yinaiwa@vip.126.com

黄杰晟,男,博士研究生,主要研究方向为医疗人工智能和机器学习。E-mail:13856487@qq.com

彭璟,男,硕士研究生,主要研究方向为医疗人工智能和计算机视觉。E-mail:pengjing@m.scnu.edu.cn

陈汉彪,男,主管技师,主要研究方向为细胞遗传学。E-mail:chenhanbiao2000@126.com

郭莉,女,主任医师,主要研究方向为胎儿细胞遗传学诊断。E-mail:guoli3861@163.com

林成创,男,博士,主要研究方向为计算机视觉和医疗人工智能。E-mail:chengchuang.lin@m.scnu.edu.cn

李双印,男,博士,主要研究方向为医疗人工智能和自然语言处理。E-mail:shuangyinli@scnu.edu.cn