

基于深度学习的活体细胞有丝分裂检测方法

柯宝生^{1,2,3}, 李颖^{1,2,3}, 任振波^{1,2,3}, 邸江磊^{1,2,3*}, 赵建林^{1,2,3**}¹西北工业大学物理科学与技术学院, 陕西 西安 710129;²陕西省光信息技术重点实验室, 陕西 西安 710129;³超常条件材料物理与化学教育部重点实验室, 陕西 西安 710129

摘要 活体细胞有丝分裂过程的发生具有时间和空间上的随机性, 自动识别并准确定位活体细胞的有丝分裂对科研人员而言充满挑战。提出一种基于深度学习的自动识别并定位活体细胞有丝分裂的检测方法。通过改进 YOLOv3 主干网络并引入注意力机制, 构建名为 DetectNet 的深度神经网络。在明场显微成像条件下, 获取多尺寸活体细胞图像并构建数据集对网络进行训练, 并对 DetectNet 与多个目标检测算法进行对比, 验证其有效性。实验结果表明, 针对活体细胞的明场显微图像, DetectNet 能够高效地从不同尺寸大视场图像中直接识别并定位有丝分裂细胞, 同时具有较高的检测精度和较快的检测速度, 因而在生物和医学领域具有非常大的潜在应用价值。

关键词 成像系统; 活体细胞; 有丝分裂; 深度学习; 目标检测算法; 明场显微成像

中图分类号 O436.1

文献标志码 A

doi: 10.3788/AOS202141.1511001

Deep Learning-Based Detection Method for Mitosis in Living Cells

Ke Baosheng^{1,2,3}, Li Ying^{1,2,3}, Ren Zhenbo^{1,2,3}, Di Jianglei^{1,2,3*}, Zhao Jianlin^{1,2,3**}¹ School of Physical Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi 710129, China;² Shaanxi Key Laboratory of Optical Information Technology, Xi'an, Shaanxi 710129, China;³ Key Laboratory of Material Physics and Chemistry Under Extraordinary Conditions, Ministry of Education, Xi'an, Shaanxi 710129, China

Abstract Owing to the spatiotemporal randomness of mitosis, the automatic identification and accurate location of mitosis in living cells are challenging tasks for researchers. Herein, a deep learning-based detection method was proposed to automatically identify and locate mitosis in living cells. Here, we built a deep neural network called DetectNet by improving the backbone network of YOLOv3 and introducing an attention mechanism. Under the condition of bright-field microscopic imaging, multiscale images of living cells were acquired and then a dataset was constructed to train the network. The trained network DetectNet was compared with multiple object detection algorithms, and its effectiveness was verified. Experimental results show that aiming at the bright-field microscopic images, DetectNet can directly identify and locate mitosis from the multiscale live cell images with a large field, achieving a higher detection accuracy and faster detection speed compared with other multiple object detection algorithms. Thus, DetectNet shows a great potential application value in the fields of biology and medicine.

Key words imaging systems; living cell; mitosis; deep learning; object detection algorithm; bright field microscopic imaging

OCIS codes 110.4850; 110.3010; 200.4260

收稿日期: 2020-12-09; 修回日期: 2021-02-23; 录用日期: 2021-03-05

基金项目: 国家自然科学基金(61927810, 62075183, 61905197)、中央高校基本科研业务费(310201911qd002)

通信作者: *jiangleidi@nwpu.edu.cn; **jlzhao@nwpu.edu.cn

1 引言

细胞的有丝分裂是人体组织生长过程中的一种重要行为。研究细胞有丝分裂过程与行为特征对再生医学、病理分析、癌症的分级与诊断等意义重大,因而受到广泛关注。为研究这一行为,通常需要对细胞形态的动态变化过程进行长期监测。由于细胞在空间分布上的分散性和有丝分裂活动在时间、空间上发生的随机性,对细胞有丝分裂活动的准确监测仍然是一项富有挑战性的工作。尽管当前可以利用显微镜图像采集系统实现细胞样品的长时间自动拍摄,但无法预知视场内的大量细胞中哪个会发生分裂以及何时会发生分裂。因此,为了研究细胞分裂行为,仍需对所拍摄的海量图像进行必要的筛选。

对显微图像进行人工分析非常耗时且缺乏客观性,从而导致对细胞分裂状态的自动监测需求非常迫切。针对这一问题,传统方法主要集中于使用阈值筛选、形态学方法、手工设计特征和机器学习等来实现细胞有丝分裂的自动检测。例如,Kaman 等^[1]在研究乳腺癌时使用阈值筛选方法,从病理图像中分割出有丝分裂细胞;ten Kate 等^[2]使用区域增长算法进行有丝分裂细胞计数;Li 等^[3]和 Liu 等^[4]分别使用 3D Haar 近似特征和隐藏条件随机场进行有丝分裂序列检测;在由相衬显微术获得的病理图像基础上,Huh 等^[5]采用形态学特征和灵活的计数模型方法;Lu 等^[6]则使用了贝叶斯模型和局域阈值方法;陈闯等^[7]在相衬显微术获得的干细胞群数据集上,基于时域演化联合特征和支持向量机(SVM)分类器,检测细胞有丝分裂事件。这些方法取得了较好的效果,但是依然需要专家的先验知识进行针对性设计,并且大量图像需要预处理和繁复的迭代运算,不利于高通量图像的处理和应用。

随着深度卷积神经网络在组织病理学图像处理领域的成功应用,部分研究人员开始探索使用卷积神经网络实现细胞有丝分裂状态的识别。例如,Xing 等^[8]和 Ferrari 等^[9]分别尝试使用卷积神经网络进行细胞有丝分裂检测和计数;Öztürk 等^[10]基于语义分割的思路,使用自定义的卷积神经网络进行细胞有丝分裂识别并获得较好效果;Jiménez 等^[11]则引入 AlexNet 和 U-Net 用于细胞分类和有丝分裂检测,取得了较高的检测准确率。但是,上述方法不同程度地存在输入图像尺寸固定、大量图像需预处理、不能充分利用视场内的全部图像信息等不足,不能适用于不同成像设备获取的多尺寸图像,

因而难以进行针对活体细胞分裂过程的长时间准确高效监测。

与此同时,基于深度学习的目标检测技术发展迅速,可以针对不同场景图像直接定位并识别出感兴趣的目标。目标检测技术可分为两阶段目标检测和单阶段目标检测。两阶段目标检测算法是基于候选建议区域的算法,首先使用主干网络提取目标特征,再将提取到的特征送入分类器进行分类和使用边界框回归网络进行目标定位,从而确定图像中感兴趣目标的类别和位置信息。此算法的目标识别精度高,但是速度慢,难以满足实时性需求,代表性算法有 R-CNN 系列算法,典型网络模型包括 R-CNN^[12]、Fast-rcnn^[13]、Faster-rcnn^[14]、Mask-rcnn^[15]等。单阶段目标检测算法是基于回归的算法,使用主干网络提取特征,将特征提取、目标分类和位置回归作为整体来处理,直接预测图像中感兴趣目标的类别和位置信息。此算法目标识别精度略低,但是速度快,代表性算法包括 YOLO 系列算法和 SSD 算法等,典型网络模型为 YOLOv1^[16]、YOLOv2^[17]、YOLOv3^[18]、SSD300^[19]、SSD512、RetinaNet^[20]等。其中,YOLOv3 模型在速度和精度上取得了较好的平衡,是当前性能优异的主流目标检测器,不过其面向不同应用场景时,仍然存在较大的改进空间。此外,对主干网络的研究也不断推陈出新,Wang 等^[21]提出跨阶段局部网络(CSPNet),以降低计算量并增强主干网络的学习能力;Hu 等^[22-24]利用注意力机制以学习输入特征图中感兴趣目标的潜在权重分布规律,抑制背景信息和噪声带来的影响,提高模型的性能。但是,上述研究主要聚焦于通用目标的快速分类识别,无法直接应用于多尺寸大视场图像中细胞有丝分裂过程的自动监测。

针对上述问题,本文使用倒置明场显微镜获取活体细胞多尺度大视场图像,基于 YOLOv3 网络模型,使用 CSP 模块对其主干网络进行改进;并借鉴 Attention U-Net^[25]中的注意力机制,建立用于细胞有丝分裂监测的 DetectNet 深度神经网络,使用获得的显微图像作为网络输入,实现细胞有丝分裂和非有丝分裂的识别和定位;对所建立的 DetectNet 与多个目标检测算法在有丝分裂识别准确率和识别速度方面进行对比,以验证 DetectNet 的有效性。此外,针对不同尺寸的大视场显微图像,DetectNet 可以直接识别并定位发生有丝分裂的活体细胞,无需针对性设计图片和进行预处理,可实现 99.5% 的高准确率和

41.44 frame/s 的实时检测帧速,能够满足长时间准确快速监测活体细胞有丝分裂活动的要求。

2 方 法

2.1 实验设置

2.1.1 实验装置

图 1 为一倒置明场生物显微镜(型号为 XDS-3)的光路结构^[26]。照明光源为显微镜内置的卤钨灯,卤钨灯发射的光经过一中心波长为 550 nm、带宽为 45 nm 的中性滤光片滤波。显微镜本身由集光镜、聚光镜孔径光阑、聚光镜、物镜、反射镜与成像透镜组成,并将物体成像至显微镜的相机端口处,最终由一单色 CCD 相机采集。CCD 相机型号为 SVCamsvs16000,4872×3248 像素,像素尺寸为 7.4 μm。大部分生物细胞都呈透明状,在自然状态下对可见光的吸收率较低,光波透过细胞时波长和振幅并不发生明显改变。因此,当使用显微镜对所测量的生物细胞进行聚焦成像时往往难以直接观察细胞,从而无法获得它们的内部结构和形态特征。而细胞内部的折射率差异所引起的相位变化在离焦位置会导致细胞的强度分布发生变化,从而大大提高透明细胞的可分辨性。由于显微镜 4f 系统中物像是严格的共轭关系,通过沿光轴方向移动物平面或者 CCD 平面,均可以获得细胞离焦平面上的强度信息。这两种方式在本质上等价,但考虑到横向与轴向放大率之间的关系,物平面和像平面的移动距离与物镜放大倍率 M 呈平方关系,所以本文通过沿光轴方向设置位移台(TSDH-601C)并将 CCD 相机移至显微镜的过焦平面 C 处,来获取细胞的离焦强度图像。

2.1.2 实验样品

实验样品为小鼠活体成骨细胞,使用含 100 mL/L 小牛血清的 MEM 培养液,置于 33.5 °C 温度下含有体积分数为 5% CO₂ 的培养箱中培养 24 h。为了避免细胞重叠,尽可能确保单个细胞相对均匀的背景,使细胞仅铺满培养皿面积的 30% 到 40%。实验时,将细胞培养皿从培养箱中取出放置

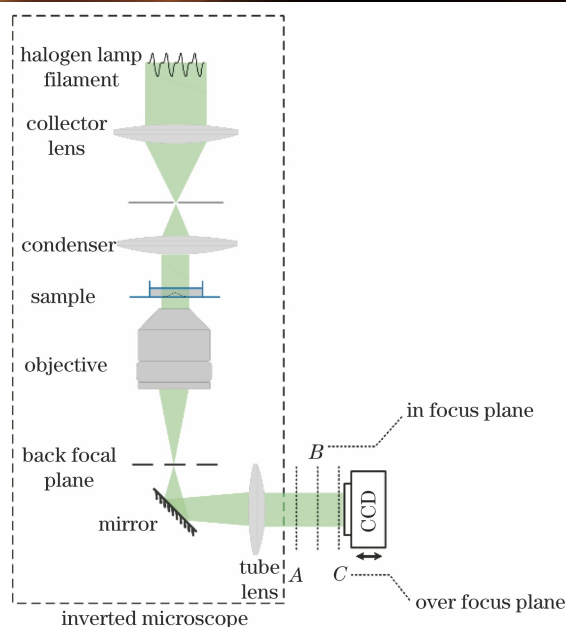


图 1 倒置明场生物显微镜的光路结构

Fig. 1 Optical path structure of an inverted bright field biological microscope

在显微镜载物台上,保持室温 25 °C,采用低功率卤素灯照明,不会对细胞产生热影响。

2.1.3 数据集

对由显微镜所拍摄的细胞过焦图像进行裁剪操作,获得视场内含有细胞的不同尺寸大视场图像,并且对图像中的细胞类别和位置进行标注。对处于有丝分裂间歇期的外形不规则的细胞不进行标注,而对外形发生圆缩并可能发生有丝分裂的细胞进行标注。将细胞分为两类,在一次实验观测过程中发生有丝分裂的细胞标注为“with_division”,不发生有丝分裂的细胞标注为“without_division”。经过多次实验,共采集到 628 幅图像,最终制作得到 over-focus cells(OFC)数据集,包含训练集 482 幅图像,验证集 53 幅图像,测试集 93 幅图像。其中训练集和验证集图像来源于多次实验观测的结果,并按照 9:1 进行随机划分,测试集图像来源于单独的一次实验观测结果,用于测试模型的目标检测能力。图 2 给出了数据集中的部分图像。

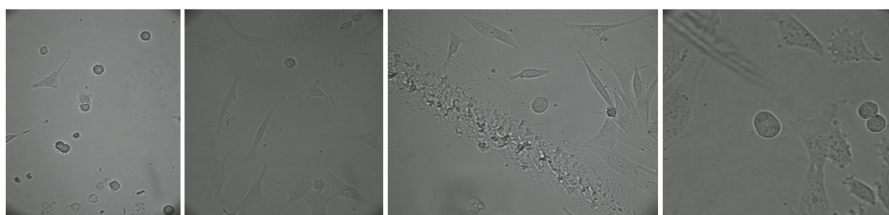


图 2 OFC 数据集中部分尺寸不同的图像

Fig. 2 Some images with different size from OFC dataset

实验观测结果表明,细胞发生有丝分裂在时间上具有一定随机性,进入圆缩状态的细胞最后既可能发生有丝分裂,也可能不发生有丝分裂。仅仅依靠形态变化,判断细胞最终会发生有丝分裂并不完全可靠。图 3(a)最左侧为实验拍摄的细胞图像,虚线框内的细胞进入圆缩状态,右侧为该细胞的局部放大图。沿虚线箭头指示方向,随着时间推移,细胞逐渐圆缩但最终并未发生有丝分裂。图 3(b)最左侧为实验拍摄的另外一个细胞图

像,此时虚线框内的细胞处于圆缩状态,右侧为该细胞的局部放大图,沿虚线箭头指示方向,细胞圆缩并最终发生有丝分裂。细胞内部的物质分布状态和外部形态共同提供了重要的特征信息,可用于目标检测算法检测细胞的有丝分裂。然而,细胞内部的物质形态各异,大小不一,随时间动态变化。因此,如何更好地获取目标特征信息以提高目标检测算法的特征提取能力,是一个比较有挑战性的任务。

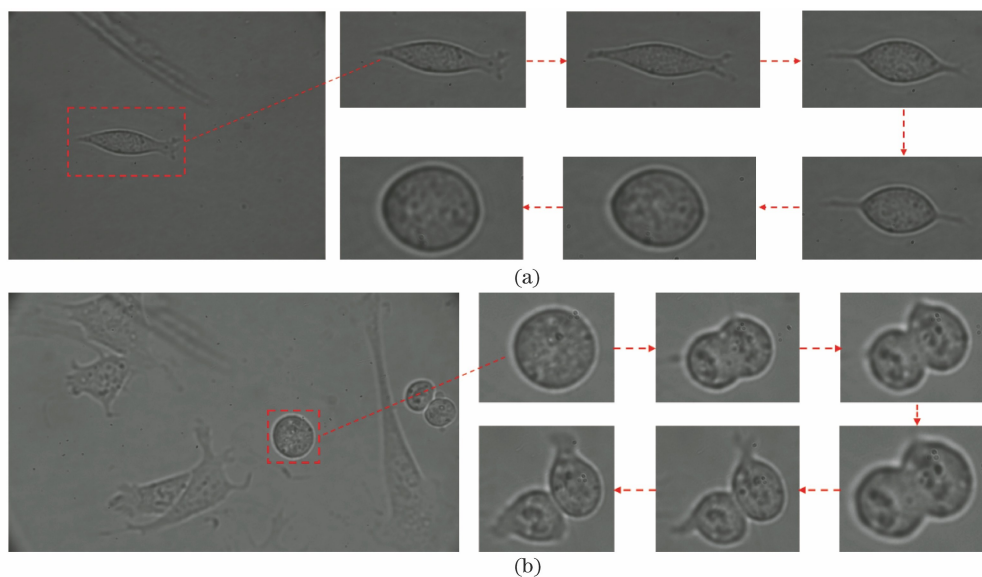


图 3 不同类别的细胞。(a)“without_division”类;(b)“with_division”类

Fig. 3 Cells of different classifications. (a) Class 'without_division'; (b) class 'with_division'

2.2 DetectNet

YOLOv3 是一种端到端的目标检测器,相比其他各类目标检测算法,表现优异,在检测精度和检测速度上都具有明显优势,但是对于某些特定应用场景,如特征不明显的小目标,仍然缺乏较好的检测能力,容易出现重复检测、错误检测和遗漏检测等现象。本文借鉴 CSPNet 在构建网络基础

模块时的特征融合策略,对 YOLOv3 的特征提取主干网络 Darknet53 进行改进,同时在不同层级特征图之间引入注意力机制,从而提高网络对细胞图像的特征提取能力。改进后的网络命名为 DetectNet,网络结构如图 4 所示,基础模块包括卷积模块、残差模块、CSP 模块、注意力模块和 yolo 层等。

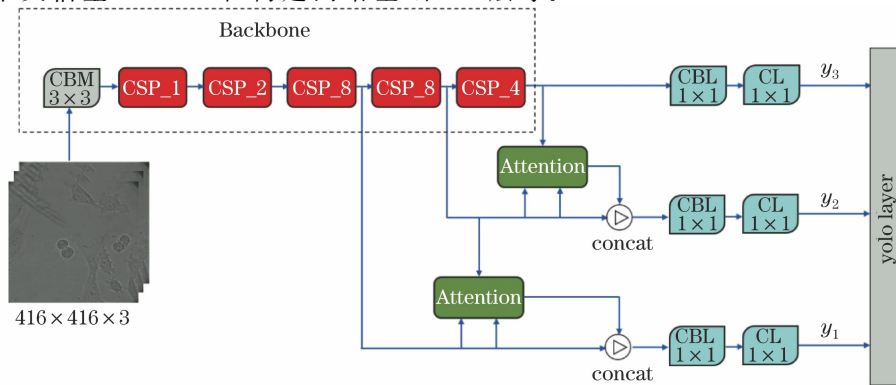


图 4 DetectNet 的网络结构

Fig. 4 Network architecture of DetectNet

2.2.1 卷积模块

所构建的卷积模块由二维卷积层 conv、批量归一化层(BN)和非线性激活层串联而成。卷积层用于提取特征;批量归一化层用于将数据归一化到均值为 0,方差为 1,使网络更容易学习到数据本身特有的分布特性,从而加快网络的收敛速度,并在一定程度上控制梯度爆炸,防止梯度消失和过拟合问题;非线性激活函数用于增强网络的非线性表征能力,建立输入与输出之间的非线性映射关系。LeakyReLU 激活函数是线性修正单元(ReLU)的改进版本,解决了 ReLU 激活函数在反向传播过程中梯度为零时导致神经元不学习的问题。Mish 激活函数是一种自正则的非单调激活函数,允许更好

的信息深入神经网络,从而得到更好的准确性和泛化。在特征提取主干网络中使用 Mish 激活函数,构建其他卷积模块时使用 LeakyReLU 激活函数。LeakyReLU 激活函数可表示为

$$y = \max(0, x) + \lambda \min(0, x), \quad (1)$$

式中: x 和 y 分别为输入和输出; λ 为常数,控制负斜率的大小,一般取值 0.01。Mish 激活函数^[27]表示为

$$y = x \tanh[\ln(1 + e^x)]. \quad (2)$$

同时,卷积模块包括 CBM 模块、CBL 模块、CL 模块和 CBL_{*n*} 模块,其中 n 表示 CBL_{*n*} 模块由 n 个 CBL 模块串联构成,结构如图 5 所示。

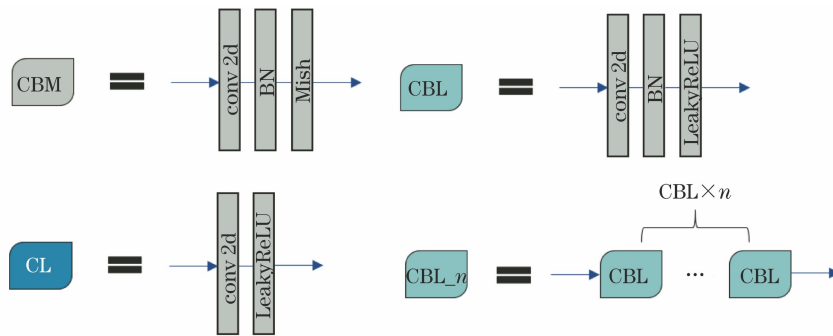


图 5 卷积模块

Fig. 5 Convolutional module

2.2.2 残差模块

残差模块^[28]由 2 个卷积核大小分别为 1×1 和 3×3 的卷积模块 CBM 和 1 个捷径连接分支构成,如图 6 所示。 1×1 卷积层对输入进行降维,减少参

数量和计算量; 3×3 卷积层提取特征并恢复维度;捷径连接实现输入与输出之间的恒等映射,构建残差结构,解决网络层数增加时可能出现的网络性能退化问题。

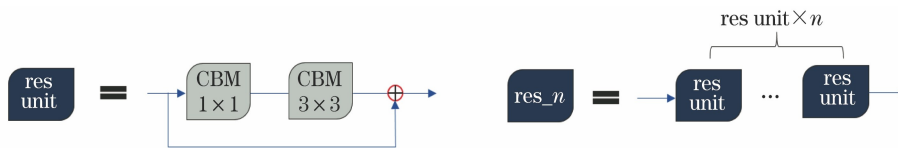


图 6 残差模块

Fig. 6 Residual block

2.2.3 CSP 模块

YOLOv3 是在 YOLOv2 的基础上对主干网络进行改进的结果,使用步长为 2 的 3×3 的卷积层替代最大池化层,实现了全卷积网络,并最大限度地保持网络信息传递的完整性;并且使用 1×1 和 3×3 小卷积核构成的残差模块加深网络层数,降低了计算复杂度。借鉴 CSPNet 的特征融合策略和处理思想,对 YOLOv3 原有的主干网络 Darknet53 中的残差模块进行改进,构建成图 7 所示的 CSP_{*n*} 模块。利用 CSP_{*n*} 模块对特征图进行处理,首先通过步长

为 2 的 3×3 的卷积模块 CBM,之后再经过两个并行支路。其中一路经过 1×1 的卷积模块 CBM 实现特征图降维,再通过 res_{*n*} 模块加深网络层数,随后经过 1×1 的卷积模块 CBM 调整通道数;另一路直接经过 1×1 的卷积模块 CBM 调整通道数。对两路输出的特征图进行拼接操作(concat),最后通过一个 1×1 的卷积模块 CBM 调整输出通道数。

2.2.4 注意力模块

注意力模块在两个不同尺寸的特征图之间构建联系,结构如图 8 所示。输入的低分辨率特征图 x_1

经过 2 倍上采样 (Upsample $2\times$) 后,再经过 1×1 卷积模块调整通道数;输入的高分辨率特征图 x_2 经过 1×1 卷积模块调整通道数。两者按元素位进行加操作,经过激活函数激活后输出;通过 Flatten 层,输出结果打平成一维形式,再经过 Sigmoid 层,

各元素位的值被映射到 0 和 1 之间;再通过 Reshape 层重整为与 x_2 相同形状的权重系数图,最后与 x_2 一起按元素位进行乘操作,输出注意力模块调制后的特征图。

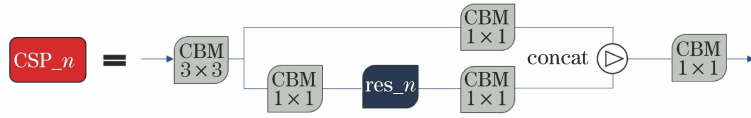


图 7 CSP 模块

Fig. 7 CSP module

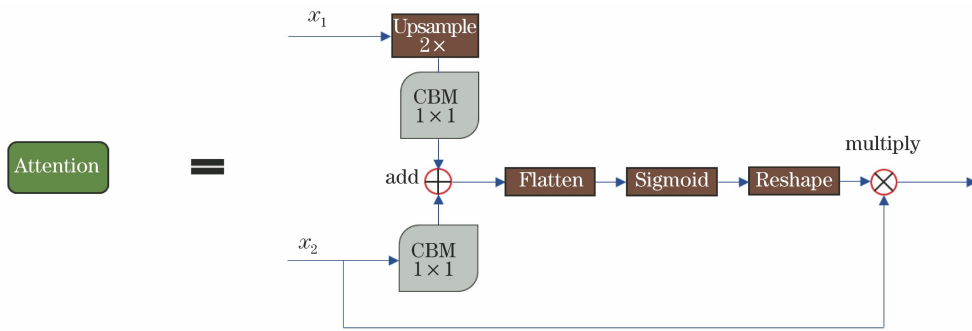


图 8 注意力模块

Fig. 8 Attention module

2.2.5 yolo 层

在 yolo 层的算法实现流程与 YOLOv3 一致。yolo 层在 13×13 、 26×26 和 52×52 三个尺寸上得到最终的特征图 y_3 、 y_2 和 y_1 ,分别独立预测出目标的归一化坐标、置信度和分类结果,并与注释文件中的真值进行比较。网络训练时,计算泛化交并比 (GIoU) 损失、置信度损失和分类损失,反向传播并梯度更新,实现有监督学习。网络推测时,不计算损失,而是在不同尺寸最终预测特征图的细分网格上,根据事先使用 K-means 聚类算法得到的 3 类 3 个尺寸共 9 种锚框,生成预测结果。使用非极大值抑制 (NMS) 算法对预测结果进行处理,并将处理结果映射到初始输入图像上,最终实现准确定位并识别目标。

2.3 目标检测算法评价标准

准确率 ($P_{\text{precision}}$) 和召回率 (R_{recall}) 描述一个目标检测算法的性能,表达式分别为

$$P_{\text{precision}} = \frac{C_{\text{TP}}}{C_{\text{TP}} + C_{\text{FP}}}, \quad (3)$$

$$R_{\text{recall}} = \frac{C_{\text{TP}}}{C_{\text{TP}} + C_{\text{FN}}}, \quad (4)$$

式中: C_{TP} 表示正样本被正确识别为正样本的数量; C_{FP} 表示负样本被错误识别为正样本的数量; C_{FN}

表示正样本被错误识别为负样本的数量。正、负样本为网络预测的边界框,用于框选数据集图像区域,依据 YOLOv3 的原理判定正负,正样本对应预测数据集图像中的目标,负样本主要对应数据集图像中的背景。性能优异的目标检测算法在召回率增加的同时,准确率也保持在较高的水平。而准确率-召回率 (P-R) 曲线则表征了目标检测算法在两个指标上的特性,曲线下的面积 S_{AP} 表征了平均准确率 (AP),表达式为

$$S_{\text{AP}} = \int_0^1 S(r) dr, \quad (5)$$

式中: r 为 P-R 曲线上点的横坐标值; $S(r)$ 为 P-R 曲线上点的纵坐标值。

实验中要检测的目标分两类,为此计算了在所有类上平均准确率的平均值 (mAP),表示为

$$M_{\text{mAP}} = \frac{\sum_{n=1}^N S_{\text{AP}}(n)}{N}, \quad (6)$$

式中: N 为要检测的目标类别总数。 S_{AP} 和 M_{mAP} 值在 0 和 1 之间,越接近于 1,说明目标检测算法的性能越好。

3 分析与讨论

采用的 DetectNet 在服务器上完成搭建、训练

与测试。服务器配置为 Ubuntu 18.04 LTS 操作系统, 12 GB NVIDIA GeForce RTX 2080Ti GPU, cuda 10.1, 开发后端为 Pytorch 1.4.0, 编译语言为 Python 3.7。使用 OFC 数据集中的训练集进行训练, 验证集用于训练过程验证模型性能, 测试集用于测试训练好的模型的性能, 测试集不参与训练。训

练集、验证集和测试集中的图像和注释文件无交叉重复情况。采用 Adam 优化器训练 900 个 epoch, 学习率调整策略为阶梯下降学习率调整机制, 学习率和损失曲线如图 9 所示。可以看出, 网络的损失处于稳定的下降状态并最终收敛, 训练到 400 个 epoch 时, 网络参数已经很稳定。

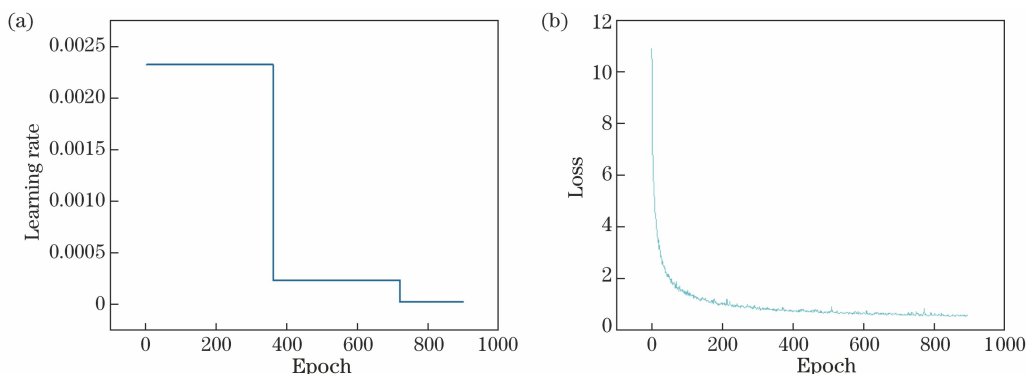


图 9 学习率和损失曲线。(a)学习率;(b)损失

Fig. 9 Learning rate and loss curves. (a) Learning rate; (b) loss

3.1 定性分析

为了定性评估改进后的 DetectNet 的检测效果, 选取测试集中的部分图像, 并与 Faster-rcnn、Faster-rcnn(pretrained)、SSD300 和 YOLOv3 目标

检测算法进行对比, 并给出可视化检测效果, 如图 10 所示。其中, Faster-rcnn(pretrained)为主干网络加载预训练权重后, 在数据集上训练得到的 Faster-rcnn 模型。

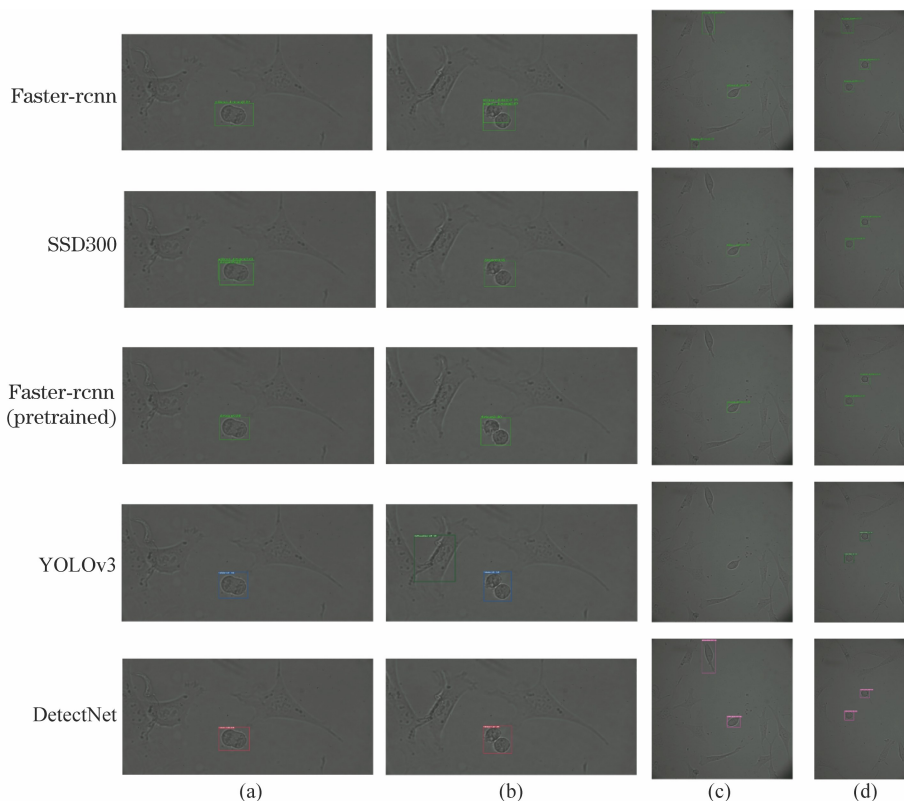


图 10 细胞检测结果的可视化

Fig. 10 Visualization of cell detection results

图 10 为不同目标检测算法对测试集中的 4 幅图像进行处理所获得的测试结果。每一行前 2 幅图像的分辨率为 $1280 \text{ pixel} \times 600 \text{ pixel}$, 第 3 幅图像的分辨率为 $1090 \text{ pixel} \times 1090 \text{ pixel}$, 第 4 幅图像的分辨率为 $890 \text{ pixel} \times 1260 \text{ pixel}$ 。为方便对比, 所有图像都进行了等比例缩小。对于图 10(a)、(b) 中发生有丝分裂的细胞, Faster-rcnn 出现错误分类和错误检测, SSD300 出现重复检测和错误分类, 而 Faster-rcnn(pretrained)、YOLOv3 和 DetectNet 均能正确检测出目标, 并且 DetectNet 检测结果的置信度与 Faster-rcnn(pretrained) 检测结果的置信度相近, 明显高于 YOLOv3 检测结果的置信度。图 10(c) 中有 2 个待检测目标, 类别均为“without_division”, DetectNet 能够全部正确检测出来, 而其他 4 个模型的检测结果则出现不同程度的漏检和错检。图 10

(d) 中有 2 个待检测目标, 类别均为“without_division”, Faster-rcnn 出现错误检测, 其他方法都能正确检测出目标, DetectNet 检测结果的置信度明显高于 YOLOv3。总体而言, 在 4 幅测试图像上, 所提 DetectNet 目标检测模型对 2 个类别的细胞都取得了较好的检测结果, 并且检测精度更高, 完全适合于明场显微成像条件下的细胞有丝分裂和非有丝分裂的检测。

3.2 定量分析

在测试集的 93 幅图像上, 采用 DetectNet 和 YOLOv3 对 2 个类别细胞在不同 IoU 阈值条件下进行检测, 得到的 P-R 曲线和相应的 AP 值如图 11 所示, IoU 表示网络预测框与真实框之间的交并比。需要指出的是, 在对比两个模型的检测结果时, 保留了置信度分数(score)大于 0.3 的结果。

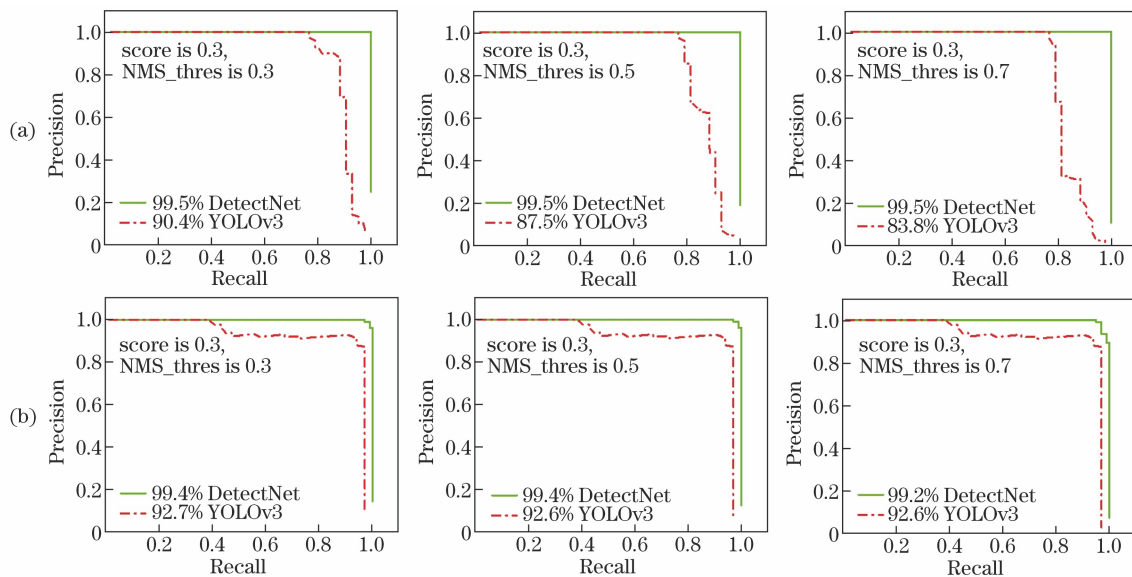


图 11 不同 IoU 阈值下细胞检测结果的 P-R 曲线。(a)“with_division”类;(b)“without_division”类

Fig. 11 P-R curves of cell detection results under different IoU thresholds. (a) Class ‘with_division’;

(b) class ‘without_division’

IoU 阈值设置越高, 对目标检测算法的检测结果要求越严格, 性能较差的目标检测算法越容易出现漏检现象, 导致单个类别检测结果的 P-R 曲线下的面积越小, 相应的 AP 值越低。在图 11(a) 中, 检测结果对应的细胞类别均为“with_division”, 由于在训练集和测试集中, 该类样本的数量都较少, IoU 阈值设置较高对 YOLOv3 模型的影响很明显。从图 11(a) 中 AP 值可以看出, IoU 阈值设置为 0.3 时, YOLOv3 的 AP 为 90.4%, IoU 阈值设置为 0.7 时, YOLOv3 的 AP 已经下降为 83.8%。对比图 11(a) 中 3 个图的 P-R 曲线下的面积可以看出, 随着 IoU 阈值的提高, YOLOv3 的 P-R 曲线下的面积逐

渐减小, 而 DetectNet 的 AP 值在 3 个图中都明显高于 YOLOv3, 并且对 IoU 阈值变化不敏感。在图 11(b) 中, 检测结果对应的细胞类别均为“without_division”。由于在训练集和测试集中, 该类样本的数量较多, 两个模型均表现出对 IoU 阈值变化不敏感, 但 DetectNet 的 AP 值在 3 个图中均明显高于 YOLOv3。IoU 阈值设置越低, 对目标检测算法的检测结果要求越宽松, 越容易出现虚警现象, 因此选择的 IoU 阈值为 0.5, 并在此基础上与其他目标检测算法进行对比。

表 1 给出了 IoU 阈值为 0.5 时, 不同算法在测试集上的 mAP 值、平均处理时间和平均帧率。其

中 processing time 是相应模型对测试集 93 幅图像进行 10 次测试得到的平均总耗时。FPS 为 12 是连贯图像的最低标准,为 20 是 RGB 游戏运行的最低标准,为 30 是视频实时流畅的最低标准。从表 1 可以看出, DetectNet 在 mAP 指标上达到 99.5%,比 YOLOv3 高 9.5 个百分点,比加载预训练权重的 Faster-rcnn(pretrained)高 7.9 个百分点,说明 DetectNet 能够在检测细胞有丝分裂和非有丝分裂方面具备高精度的优势。对比 Faster-rcnn(pretrained)和 Faster-rcnn 可以看出,由于其使用 ResNet50 作主干网络,相比较 SSD300 的 16 层主干网络而言,层数较多,而数据集规模较小,在不加载预训练权重的情况下,网络的初始化较差,导致最终很难训练出一个性能较好的模型。所以

表 1 不同模型在 IoU 阈值为 0.5 时的性能

Table 1 Performance of different models with IoU threshold of 0.5

Method	Backbone	Dataset	mAP	Processing time /s	FPS
Faster-rcnn	ResNet50	OFC	0.336	5	18.6
SSD300	VGG16	OFC	0.754	6	15.5
Faster-rcnn(pretrained)	ResNet50	OFC	0.916	5	18.6
YOLOv3	Darknet53	OFC	0.9	3.433	27.09
DetectNet	Custom Backbone	OFC	0.995	2.244	41.44

4 结 论

在 YOLOv3 模型基础上,采用 CSP 模块对主干网络进行了改进,在不同尺寸特征图之间引入了注意力机制,最终成功建立 DetectNet,有效提高了模型对细胞类目标的检测能力和在测试集上的推理速度,模型在测试集上也保持了较好的鲁棒性和泛化能力。由对比实验可知,在明场显微成像条件下,所提 DetectNet 模型能够对活体细胞有丝分裂和非有丝分裂实现实时高精度准确检测,相比较传统方法、机器学习和浅层卷积神经网络,训练好的模型进行推测时无需先验信息、专家知识和复杂预处理,能够满足实时检测的要求,并适用于多尺寸大视场活体细胞检测和实时反馈的场景,在再生医学、病理分析、癌症的分级与诊断等医学领域具有非常大的潜在应用价值。

参 考 文 献

[1] Kaman E J, Smeulders A W M, Verbeek P W, et al. Image processing for mitoses in sections of breast cancer: a feasibility study[J]. *Cytometry*, 1984, 5(3): 244-249.

[2] ten Kate T K, Beliën J A, Smeulders A W, et al. Method for counting mitoses by image processing in

Faster-rcnn 的 mAP 只有 33.6%,比 SSD300 低 41.8 个百分点。DetectNet 是在数据集上直接从零训练得到的模型,不加载预训练权重或迁移学习,并且网络层数更多,mAP 可以达到 99.5%,说明模型本身具有较好的鲁棒性和泛化能力。然而,对于 DetectNet,虽然在构建网络时不同尺寸特征图之间构建了注意力模块,但相比较 YOLOv3 的原始网络结构而言,相当于取代了原先在不同尺寸特征图之间的特征融合模块,增加的计算负担并不大,而主干网络使用 CSP 模块则大幅度降低了计算负担,因此推理速度更快。从表 1 可以看出,所提模型的速度达到 41.44 frame/s,比 YOLOv3 更快,已经满足实时检测的标准,适合实时检测的应用场景。

Feulgen stained breast cancer sections [J]. *Cytometry*, 1993, 14(3): 241-250.

- [3] Li K, Miller E D, Chen M, et al. Computer vision tracking of stemness [C] // 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, May 14-17, 2008, Paris, France. New York: IEEE Press, 2008: 847-850.
- [4] Liu A, Li K, Kanade T. Mitosis sequence detection using hidden conditional random fields [C] // 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, April 14-17, 2010, Rotterdam, Netherlands. New York: IEEE Press, 2010: 580-583.
- [5] Huh S, Ker D F E, Bise R, et al. Automated mitosis detection of stem cell populations in phase-contrast microscopy images [J]. *IEEE Transactions on Medical Imaging*, 2011, 30(3): 586-596.
- [6] Lu C, Mandal M. Toward automatic mitotic cell detection and segmentation in multispectral histopathological images [J]. *IEEE Journal of Biomedical and Health Informatics*, 2014, 18(2): 594-605.
- [7] Chen C, Jia W W, Wang Y. Recognition and detection of mitosis event based on feature of evolution in time domain [J]. *Laser & Optoelectronics Progress*, 2019, 56(24): 241007.

- 陈闯, 贾文武, 王亚. 基于时域演化特征的细胞有丝分裂事件识别与检测[J]. 激光与光电子学进展, 2019, 56(24): 241007.
- [8] Xing F Y, Xie Y P, Yang L. An automatic learning-based framework for robust nucleus segmentation [J]. IEEE Transactions on Medical Imaging, 2016, 35(2): 550-566.
- [9] Ferrari A, Lombardi S, Signoroni A. Bacterial colony counting with convolutional neural networks in digital microbiology imaging [J]. Pattern Recognition, 2017, 61: 629-640.
- [10] Öztürk Ş, Akdemir B. A convolutional neural network model for semantic segmentation of mitotic events in microscopy images[J]. Neural Computing and Applications, 2019, 31(8): 3719-3728.
- [11] Jiménez G, Racoceanu D. Deep learning for semantic segmentation vs. classification in computational pathology: application to mitosis analysis in breast cancer grading [J]. Frontiers in Bioengineering and Biotechnology, 2019, 7: 145.
- [12] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [13] Girshick R. Fast R-CNN [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [14] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [15] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [16] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [17] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [18] Redmon J, Farhadi A. Yolov3: an incremental improvement [EB/OL]. (2018-04-08) [2020-11-20]. <https://arxiv.org/abs/1804.02767>.
- [19] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [M] // Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [20] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [21] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1571-1580.
- [22] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [23] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module [M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [24] Zhao B, Wang C P, Fu Q, et al. Multi-scale infrared pedestrian detection based on deep attention mechanism [J]. Acta Optica Sinica, 2020, 40(5): 0504001.
赵斌, 王春平, 付强, 等. 基于深度注意力机制的多尺度红外行人检测 [J]. 光学学报, 2020, 40(5): 0504001.
- [25] Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: learning where to look for the pancreas [EB/OL]. (2018-04-11) [2020-11-20]. <https://arxiv.org/abs/1804.03999>.
- [26] Li Y, Di J L, Ma C J, et al. Quantitative phase microscopy for cellular dynamics based on transport of intensity equation [J]. Optics Express, 2018, 26(1): 586-593.
- [27] Mishra D. Mish: a self regularized non-monotonic neural activation function [EB/OL]. (2019-08-23) [2020-11-20]. <https://arxiv.org/abs/1908.08681>.
- [28] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.