

中图法分类号: TP391.7 文献标识码: A 文章编号: 1006-8961(2023)09-2733-16

论文引用格式: Lin F C, Xie H T, Liu C B and Zhang Y D. 2023. Mitosis detection by appearance and motion pattern perception. Journal of Image and Graphics, 28(09):2733-2748(林凡超, 谢洪涛, 刘传彬, 张勇东. 2023. 外观和运动模式感知的有丝分裂细胞检测. 中国图象图形学报, 28(09):2733-2748)[DOI:10.11834/jig.220901]

外观和运动模式感知的有丝分裂细胞检测

林凡超, 谢洪涛, 刘传彬*, 张勇东

中国科学技术大学信息科学技术学院, 合肥 230026

摘要: 目的 在癌症筛查和药物研发等医学研究和诊疗过程中, 显微图像中的有丝分裂细胞检测可以提供重要的生物学判据。然而, 不同培养条件下图像分布差异明显, 且细胞密度逐渐增大导致场景变得复杂, 常规预处理方法难以进行有效的区域筛选; 不同阶段细胞外观相似、运动过程模糊, 现有方法缺乏对区域特征编码的显式监督, 容易因为语义区分能力不足导致错误预测。为此, 本文提出基于外观和运动模式感知的检测框架, 通过两阶段预处理和对细胞状态模式的判别性学习, 实现复杂场景下的精准预测。**方法** 本文方法采用3阶段检测框架: 在预处理阶段结合区域分割网络 and 先验优化算法来充分精简候选区域; 在预训练阶段构造基于图像分类和重构的两种辅助任务, 为候选区域的外观和运动编码提供直接监督, 使编码网络具备对不同细胞状态的语义感知能力; 在全模型训练和预测阶段, 以预处理得到的候选区域序列作为输入, 用预训练的编码网络提取候选区域特征, 最终通过时序网络融合序列上下文信息得到细胞检测结果。**结果** 在C2C12-16数据集上的实验结果表明, 本文方法的平均性能达到: 验证集精准率85.3%, 召回率89.3%, F得分87.2%; 测试集精准率86.4%, 召回率86.1%, F得分86.2%, 时序检测误差 0.221 ± 0.536 帧, 空间检测误差 3.321 ± 2.461 像素, 在检测精度和稳定性上都超过了现有方法。**结论** 本文提出了新的复杂场景下有丝分裂细胞检测框架。所采用的预处理策略可以有效精简候选区域, 显著提高检测效率; 针对编码网络的辅助任务预训练充分提升了模型对候选区域外观和运动特征的学习能力, 最终能够克服电镜图像中复杂场景和细胞模式的干扰, 准确且稳定地对有丝分裂细胞进行时空检测。

关键词: 相衬显微图像; 有丝分裂细胞检测; 多阶段检测; 时空特征编码; 辅助训练

Mitosis detection by appearance and motion pattern perception

Lin Fanchao, Xie Hongtao, Liu Chuanbin*, Zhang Yongdong

Department of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

Abstract: Objective In the processes of medical research and diagnosis, such as cancer screening and drug development, mitosis detection under phase-contrast microscopy image provides a very important biological criterion. Manual counting of mitotic cells takes a lot of time and labor. Thus, automatic mitosis detection is more efficient and economic than the manual process. On the one hand, the distributions of mitosis images are significantly different under various culture conditions. Moreover, the increment of cell density makes screening out the cell regions difficult for conventional preprocessing meth-

收稿日期: 2022-08-31; 修回日期: 2023-03-26; 预印本日期: 2023-04-02

* 通信作者: 刘传彬 liuchb92@ustc.edu.cn

基金项目: 中国科学院稳定支持基础研究领域青年团队计划(YSBR-067); 国家自然科学基金项目(62022076, 62272436); 中国博士后科学基金面上项目(2021M703081); 中央高校基本科研业务费专项资金资助(WK2100000026); 安徽省自然科学基金项目(2208085QF190)

Supported by: CAS Project for Young Scientists in Basic Research (YSBR-067); National Natural Science Foundation of China (62022076, 62272436); China Postdoctoral Science Foundation (2021M703081); Fundamental Research Funds for the Central Universities (WK2100000026); Natural Science Foundation of Anhui Province, China (2208085QF190)

ods. On the other hand, the cells at different stages have similar appearances and blurred motion processes. They also require the model to have a strong ability to discriminate cell types and states. Recent deep-learning-based works use three-dimensional convolutions or temporal networks to obtain context information from the sequence images. However, an explicit supervision process for learning cell states is lacking, making effective pattern information of target regions difficult to achieve. As a result, these methods are not fully capable of distinguishing different cells and background areas from feature encoding, and their performance and generalization ability are limited. Therefore, this study explores a detection framework based on cell appearance and motion pattern perception to solve the above problems. An accurate prediction under complex scenes is also achieved through effective preprocessing and discriminative learning of cell patterns. **Method** The proposed method consists of three stages. The first stage aims to extract regions of interest as candidates. This stage serves as the preprocessing for finding the notable areas and facilitating the later detection. The original electron microscope image is divided into local slices. An instance segmentation network is also trained to segment roughly all the candidate regions that may contain mitosis. Then, a candidate region refinement algorithm is designed based on a concise spatiotemporal hypothesis to refine the candidates and reduce the redundant results. In the second stage, two encoding networks are pretrained to maintain the feature encoding of both appearance and motion information by building proxy learning processes. In particular, an image classification task is conducted for the appearance encoding network training, which learns to predict the cell categories from the spatial context of a single patch. Moreover, an image reconstruction task is conducted for the motion encoding network training, which considers patches from adjacent frames and learns the information of inter-frame changes by recovering the raw patches. These two processes complement each other to help model the cell states from different aspects. Finally, in the third stage, the whole spatiotemporal model is trained end-to-end by classifying the candidate patch sequences. The spatial modules are initialized with the pretrained parameters of encoding networks in the second stage, thereby allowing them to be aware of the cell patterns at the beginning of the training. Given the appropriate spatial context, the temporal modules are optimized to combine the interframe information and make the final prediction. The overall model provides a confidence score for each patch. The position with the highest score is regarded as a mitosis point.

Result We conduct experiments on the public C2C12-16 benchmark. The experimental results demonstrate the superior detection ability of the proposed method. On the C2C12-16 validation set, the mean precision reaches 85.3%, the mean recall reaches 89.3%, and the mean F-score is 87.2%. On the C2C12-16 test set, the mean precision reaches 86.4%, the mean recall reaches 86.1%, and the F-score is 86.2%. The proposed method demonstrates high performance and can generate stable predictions under various conditions. The mean temporal bias of the proposed method in all groups is only 0.221 ± 0.536 frames, and the mean spatial bias is 3.321 ± 2.461 pixels, both of which are much lower than those obtained by the counterpart method. **Conclusion** This study explores a new framework to tackle the hard cases under complex scenes in mitosis detection. The preprocessing strategy effectively extracts candidate regions and substantially improves detection efficiency. The pre-training of the feature encoding network based on proxy tasks fully enhances the model's ability to learn the appearance and motion patterns of the candidate regions. With the preprocessing and pretraining designs, our framework can distinguish the discrepancy of visual patterns between mitosis cells, common cells, and background noises, overcome the interference of complex scenes and cell patterns in the microscope image, and achieve both accurate and stable mitosis detection from spatiotemporal dimensions.

Key words: phase contrast microscopy image; mitosis detection; multistage detection; spatiotemporal feature encoding; proxy training

0 引言

有丝分裂是细胞增殖活动中必不可少的过程,分析活体细胞的有丝分裂情况对于研究细胞群的行为方式等课题具有重要意义。通过对相衬显微图像

序列中的有丝分裂细胞进行检测可以实现细胞跟踪和细胞谱系构建,进而为干细胞培养、组织工程、药物制造和癌症的筛查评估等医学研究任务提供丰富的信息判据(Liu等,2017a)。

然而,复杂的时空场景和背景干扰,使得相衬显微图像序列中的有丝分裂细胞检测任务充满挑战

(Liu等, 2017a; Su等, 2021)。在空间上, 有丝分裂细胞往往具有不规则的外观, 在形态、纹理和方向等方面具有较大差异, 而不处于分裂期的普通细胞的视觉模式也可能与有丝分裂细胞相似, 这大大增加了分裂状态的检测难度。不同培养条件下细胞的空间分布和亮度特性也存在显著区别, 使目标区域的筛选面临困难。在时间上, 由于细胞发育过程的可变形特性, 不同细胞间和同一细胞在分类周期的不同阶段都具有多样性的外观表现, 这对关键时序的检测提出了挑战。此外, 随着细胞培育过程进行, 培养皿中的细胞密度也会急剧增加, 使多个细胞相互靠近, 待检测目标因此更容易受到背景细胞的干扰。

在早期发展过程中, 研究者利用传统机器学习技术, 通过细胞簇分离(Yang等, 2005)、均值偏移算法(Debeir等, 2005)、水平集算法和双边滤波器(Li等, 2008)、卡尔曼滤波器和局部图匹配(Liu等, 2017b)等方法进行细胞跟踪和分裂检测。也有一些方法无需进行跟踪, 而是根据任务设计了特定的预处理和事件检测技术。例如, Li和Kanade(2009)和Liu等人(2013)通过凸函数优化实现图像预处理; Liu等人(2010a)用支持向量机(support vector machine, SVM)分类器实现细胞分裂事件检测; Marcuzzo等人(2009)利用收敛指数滤波器检测细胞核; Quelhas等人(2010)结合图像配准和光流分析来搜索分裂细胞。为充分利用序列图像间的关联信息, 不少研究者用图模型实现时空联合的细胞状态建模和预测。Gallardo等人(2004)训练多个不同的隐马尔可夫模型对图像块进行分类。Liang等人(2007)利用条件随机场模型来克服有丝分裂检测任务中的少样本和数据多样性问题。Huh等人(2011)进一步提出事件检测的条件随机场方法, 利用分裂事件的时序信息强调细胞的动态变化。Liu等人(2010b)定义成像模型来筛选细胞分裂区域, 并学习隐式条件随机场分类器对候选区域进行分类。然而, 传统机器学习方法的特征提取和序列建模能力有限, 在复杂图像场景下往往难以得到鲁棒的检测结果。

随着深度学习的兴起, 基于深度神经网络的方法在许多医学影像分析任务中获得了应用(蒋希等, 2022), 也为有丝分裂细胞的自动检测提供了新的方向。在相衬显微图像序列中, 由于细胞和背景之间、不同类型细胞之间容易产生干扰和混淆, 需要结合图像的模态特性和细胞区域的时空特征来设计

方法框架。Nie等人(2016)首次提出使用三维的卷积神经网络(convolutional neural network, CNN)来提取细胞分裂过程中局部时空区域的特征, 并采用支持向量机分类器对区域是否属于有丝分裂细胞进行分类。Nishimura和Bise(2020)基于前背景亮度差异提取可能包含有丝分裂细胞的候选区域, 再用V-Net网络(Milletari等, 2016)进行端到端的候选区域热力图估计, 直接预测有丝分裂出现的时空位置。这些方法在模型中使用了三维卷积和池化操作, 能够提取时序特征, 但所需的模型参数量和计算量过大, 处理速度较慢。Mao和Yin(2016)使用多个二维卷积网络来分别提取细胞区域的外观和运动特征, 并基于多层分级结构实现序列特征的融合来综合预测细胞类别。Su等人(2017)将卷积网络与长短时记忆(long short term memory, LSTM)网络(Hochreiter和Schmidhuber, 1997)结合, 把卷积网络提取的图像块的特征输入LSTM进行切片间的关系建模和预测。在此基础上, Lu等人(2020)设计了显著性导向的深度网络, 额外进行序列显著性和序列结构学习, 同时实现有丝分裂细胞的识别和定位。Su等人(2022)将分裂过程中关键帧的查询视为马尔可夫过程, 并引入深度强化学习, 通过细胞识别模块的激励来学习关键帧选择过程中的决策准则。

多数现有深度学习方法的基本思路是首先通过基于亮度的预处理进行区域筛选, 再将空间特征编码网络与时序网络级联预测, 融合时空特征进行端到端训练。然而, 该框架仍然存在两个亟待解决的问题。一方面, 基于亮度的图像特征判别处理能力有限, 复杂场景下筛选结果冗余性大, 影响后续预测的效率和精度; 另一方面, 训练过程缺乏对于区域编码的显式监督学习, 从时序网络到编码网络的梯度衰减导致端到端优化难以为底层的特征编码过程提供有效监督, 使编码网络不能充分学习不同类别细胞以及背景区域之间的多样化语义区别。

为了解决上述问题, 本文提出新的3阶段检测模型。首先, 针对复杂场景导致的区域冗余性问题, 本文结合DeepLabV3+深度分割网络和先验优化算法进行序列预处理, 逐步提高筛选精度, 得到候选细胞区域; 然后, 针对多样化目标区域的特征学习问题, 本文提出新的预训练策略, 用CNN编码网络提取候选区域的外观和运动特征, 并通过设计图像分类和图像重构两种辅助任务为编码网络提供直接监

督,强化网络对细胞不同状态模式的感知和区分能力;最终,基于预处理得到的候选区域和预训练得到的编码网络进行CNN—双向LSTM时空模型的端到端训练和预测,实现复杂场景下的精准检测。

本文的主要贡献如下:1)提出了一种新的相衬显微图像序列中有丝分裂细胞检测框架,通过设计多层次的检测过程和多阶段的训练策略来由粗到精对图像序列进行预测,逐步优化检测结果。2)提出了两阶段预处理策略,结合可学习的深度网络和基于先验假设的优化算法,提升预处理过程的区域筛选精度。3)提出了基于辅助任务的编码网络预训练方案,通过分类任务训练外观编码网络的空间特征提取能力,通过重构任务训练运动编码网络的区域变化感知能力,解决了编码网络优化困难的问题,使候选区域的特征编码具备更强的判别性。4)剖析了在复杂图像场景中进行细胞检测的任务难点,进行针对性的方法设计,达到比现有方法更高精度和更加稳定的检测结果。

1 本文方法

1.1 任务概述

有丝分裂细胞检测任务中,输入一段连续的相衬显微图像序列,要求对序列中出现的有丝分裂细胞进行检测,输出其所在的时间和空间坐标。由于该任务需要在时空两个维度进行预测,当前的主流思路是首先用预处理过滤无关的背景区域,提取候选的细胞区域,再通过时空特征的关联分析来优化候选区域,得到最终检测结果。其中,在预处理阶

段,现有方法多采用基于亮度的阈值化方法进行区域选择(Mao和Yin,2016;Nishimura和Bise,2020)。这种方法虽然过程较为简明,但由于缺乏对细胞外观的高层语义理解,得到的结果中包含较多背景区域和不处于分裂期的普通细胞,冗余性较大。而在时空特征关联阶段,典型的做法是将每帧独立的空间特征编码(如CNN网络)和连续多帧的时序状态建模(如LSTM网络)结合进行端到端训练,预测细胞的分类结果并通过标签进行监督(Mao等,2016;Lu等,2020)。该过程通过检测结果误差的梯度反向传播进行联合优化,由于从时序到空间网络的传播过程存在梯度衰减,仅基于端到端的学习优化难以保证特征编码网络具备对不同细胞模式的感知能力。

针对现有方法在预处理和时空特征关联过程中存在的问题,本文提出一种基于外观和运动模式感知的多阶段细胞有丝分裂检测方案,整体框架如图1所示。本文方法主要分为3个部分:1)候选区域预处理。在预处理过程中,针对亮度阈值化方法存在的区域冗余性问题,采用将深度网络和基于先验假设的优化算法相结合的方式区域筛选,充分精简候选区域;2)编码网络预训练。针对候选区域特征编码的学习优化问题,将编码网络分为外观编码和运动编码两部分,分别设计对应的辅助任务为编码过程提供直接监督,提升对于细胞间不同外观和运动模式的语义感知能力;3)空间—时间全模型训练和预测。将预训练得到的空间编码网络与时序网络相结合,建模候选区域的时空特征,得到最终检测结果。

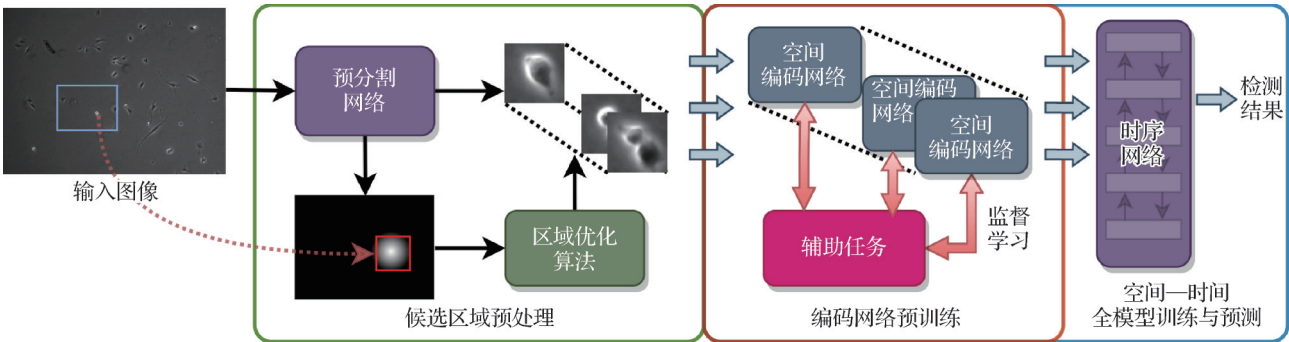


图1 整体框架图

Fig. 1 The overall framework

1.2 候选区域预处理

原始电镜图像序列长、分辨率高,而细胞所占面

积小,因而图像数据中绝大部分是与细胞分裂过程无关的背景区域。而由于有丝分裂过程发展较快,

在可见的细胞中处于有丝分裂阶段的个体数量更为稀少。在细胞群发展的早期阶段,图像中大部分区域不包含任何细胞,而在发展后期,则会出现大量处于分裂间期的细胞。因此,通过预处理来剔除无关信息的干扰,筛选出可能包含有丝分裂行为的时空区域对于实现高效的细胞检测十分重要。

此前的预处理方法需要在图像上进行遍历式的滑动搜索,基于细胞区域所特有的亮度差异来确定候选区域。这种方式下预处理结果受到窗口大小和亮度差异阈值等超参数设置的影响较大,且预处理后得到的候选区域往往仍具有较大的冗余性,精准率较低(Mao和Yin,2016)。为了在预处理阶段得到更加精简的候选区域,提高后续预测的效率。本文采用两阶段的预处理过程:第1部分使用可学习的预分割网络从电镜图像序列中分割出潜在的兴趣区域;第2部分对第1部分提取的兴趣区域进行整合与优化,依据细胞分裂过程在时空关系上的先验来滤除冗余的兴趣区域,得到精简的预分割结果作为有丝分裂细胞候选区域。

1.2.1 兴趣区域预分割

深度网络通过训练和学习具备强大的语义表征能力(曹家乐等,2022),相比传统的亮度阈值类方法可以更好地理解图像区域的复杂视觉模式,对不同形态的分裂期细胞具有更强的识别能力。如图2所示,本文首先使用可学习的深度预分割网络,以热力图估计的形式预测潜在的候选区域。

本文采用图像分割领域经典的DeepLabV3+模型(Chen等,2018)作为预分割网络。该模型具有编码—解码结构和多尺度的空间特征提取模块,可以较好地关注预分割任务所需的局部语义。如图2所示,为降低训练所需的显存同时保留图像数据中的信息,首先将全分辨率(1 392 × 1 040 像素)的电镜图像在横向和纵向上分别4等分,得到16个切片图像后分别处理。考虑到细胞有丝分裂是一个动态过程,本文将切片图像依据时间顺序取对应位置的前后3帧切片,连接成三通道的伪彩色图像作为预分割网络的输入。这样,网络通过连续3帧的切片外观信息,足以对是否存在候选区域做出初步判断。预分割网络输出单通道的分割概率图,通过有丝分裂坐标点转换得到的概率图标签进行监督。设有丝分裂点坐标为 (x_i, y_i) ,则对应的概率图标签为

$$U_{x,y} = \begin{cases} 1 - \frac{1}{r} d(x, y) & d(x, y) \leq r \\ 0 & d(x, y) > r \end{cases} \quad (1)$$

式中, $d(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$ 表示从图中任意位置 (x, y) 到有丝分裂点的距离,这样就以有丝分裂点为中心、 r 为半径构造出了伪分割标签,越靠近分裂中心点标签值越高。通过对伪标签的拟合,预分割网络可以对存在有丝分裂的区域给出较高的预测值,从而达到筛选候选区域的目的。实际训练中,由于有丝分裂细胞较为稀疏,为保证前景和背景像素点具有合理的样本比例,本文仅利用切片中包含有丝分裂细胞的图像数据进行预分割网络训练。测试时,每帧图像输入预分割网络得到候选区域的预分割概率图,再通过候选区域优化算法对密集的概率预测进行精简和优化。

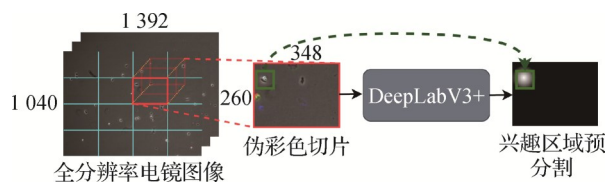


图2 兴趣区域预分割

Fig. 2 Pre-segmentation of regions of interest

1.2.2 候选区域优化

预分割网络的检测结果具有很高的召回率,能够最大程度地保留潜在的有丝分裂细胞样本。但由于没有考虑更加长程的时序信息,容易导致同一细胞区域在不同时刻被重复检出,因此精准率仍然较低。本文基于细胞运动过程的先验假设,即细胞在相邻时序中的位移相对较小,设计了候选区域优化算法,对时空近邻的区域进行整合。伪代码如下:

输入: 预分割概率图列表 M , 优化阈值 R_p, R_r, R_t, R_{xy} 。

输出: 候选区域列表 O 。

初始化: 兴趣区域列表 $G = \emptyset$, 概率列表 $P = \emptyset$;

1) 遍历所有时刻概率图 $M_i \in M$

当 M_i 中存在最大值 $p_i > R_p$, 循环执行:

列表 G 添加 p_i 所在位置 $[t, x, y, p_i]$;

M_i 以 (x, y) 为中心、半径 R_r 内置 0;

初始化: 候选区域列表 $O = \emptyset$, 标记列表 $S =$

$[s_0, s_1, \dots, s_N]$ 置0;

2) 遍历兴趣区域列表 G

初始化: 概率列表 P 为空;

对区域 $i: [t_i, x_i, y_i, p_i]$, 若标记 $s_i = 0$:

遍历时序邻域 $t'_i \in [t_i - R_t, t_i + R_t]$:

遍历 G 中时序为 t'_i 的区域 m :

计算 m 到 i 的空间距离:

$$D = \sqrt{(x_m - x_i)^2 + (y_m - y_i)^2};$$

若 $D < R_{xy}$ 且 $s_i = 0$:

标记 s_m 置为 1;

概率列表 P 中添加 p_i ;

查找列表 P 中最大值对应的时序 n ;

列表 O 添加候选区域 $[t_n, x_n, y_n]$ 。

如上述伪代码所示, 步骤 1) 中设置概率阈值 R_p , 循环提取预分割概率图中高于 R_p 的最大值位置作为兴趣区域中心点, 并将该位置半径 R_r 范围内概率置 0, 保证邻域内的区域提取结果不重复; 步骤 2) 中设置时间阈值 R_t 和空间阈值 R_{xy} , 对时空重叠程度较高的区域位置进行合并, 同一时空范围内只保留概率最大的候选结果, 使预处理结果更加精简, 进一步提高了检测的效率。两阶段预处理前后的性能对

比表明, 本文的预处理方法可以有效提取候选区域并保持较高的召回率。

1.3 编码网络预训练

利用多帧切片的时空信息可以更好地学习和预测细胞的有丝分裂过程, 为此需要将局部的特征编码网络和长程的时序网络相结合。然而, 现有方法的训练过程只通过时序网络的最终预测结果进行监督, 而对时序网络之前的特征编码缺乏直接优化。为了提升特征编码网络对于不同细胞状态的语义建模能力, 本文提出一种基于空间模式感知的编码网络预训练策略, 通过构造辅助任务来直接训练特征编码网络, 从而为后续空间—时间全模型的端到端训练提供充分的视觉模式感知能力。

细胞区域的视觉模式可以分为静态和动态两方面。静态视觉模式是指细胞区域在一帧图像内有特定的形态和空间分布, 动态视觉模式细胞在相邻两帧之间具有特定的运动变化过程。如图 3 所示, 本文设计了外观和运动两种编码网络来对候选区域进行特征编码, 其中外观编码网络学习细胞的静态视觉模式, 并通过图像分类的辅助任务进行显式的监督训练; 运动编码网络学习细胞的动态运动模式, 基于图像重构的辅助任务进行训练优化。

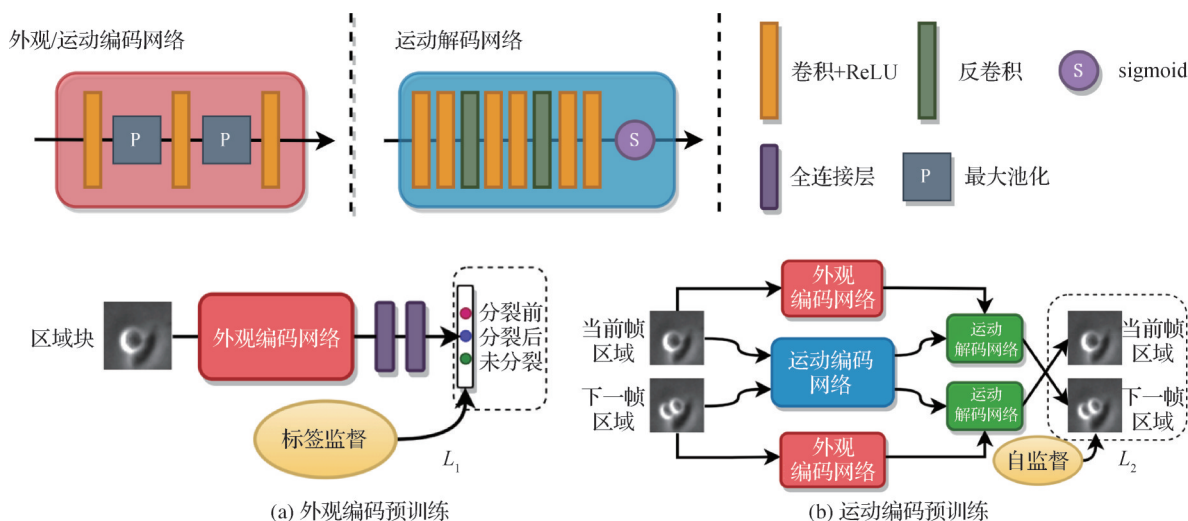


图 3 编码网络预训练

Fig. 3 The pre-training of encoding networks ((a) pre-training of the appearance encoder; (b) pre-training of the motion encoder)

1.3.1 预训练数据构造

辅助任务的学习过程首先需要构造不同类型的样本和对应的标签。为使训练所用的正负样本更加接近实际预测过程, 本文用 1.2.1 节中训练完成的 DeepLabV3+ 模型对训练集数据进行重新预测, 从预

测得到的候选区域中选择样本(取 40×40 切片), 不同类别样本的选择如图 4 所示。

在分裂时刻附近, 细胞的外观会发生显著变化(如形态变圆、亮度提升等), 通过选取分裂细胞周围的时空区域作为正样本, 选取未分裂细胞和背景区

域作为负样本,可以训练编码网络更好地理解分裂细胞所特有的模式信息。然而,细胞从常规状态发展到分裂状态存在较长的过渡过程,在此期间细胞外观缓慢变化,为了避免训练过程使用的标签产生歧义,需要对细胞的不同时期进行明确的类别划分。图4分别统计了训练集中分裂前1帧($t-1$)和分裂后1帧(t)图像相对于其他时刻图像的特征余弦相似度。可见,分裂时刻前后3帧内的图像区域具有较为明显的分裂期特性,特征相似度高,用于构造正样本,其中前3帧标记为“分裂前”类别,后3帧标记为“分裂后”类别;与分裂时刻时序间隔8帧以上的图像区域与有丝分裂区域之间有明确的特征差异,用于构造负样本,标记为“未分裂”类别;其他时刻的图像区域不用于训练。这样,构造出的正负样本对具有较为明确的语义区分,不会产生混淆而误导编码训练过程。同时,相较于直接使用标签点来构造样本,通过上述策略构造的样本不仅更贴近测试时的实际网络输入,减小了训练和测试阶段的差异,也具有更好的空间分布多样性。

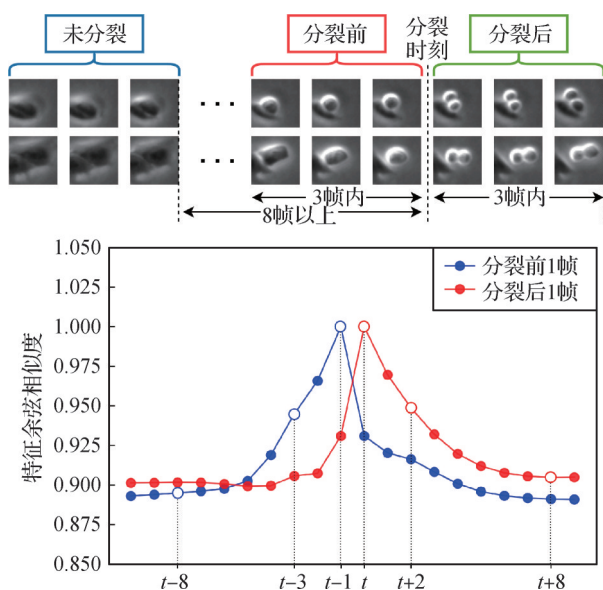


图4 特征相似性与样本选择

Fig. 4 Feature similarity and sample selection

1.3.2 外观编码预训练

预训练的目标是为编码网络提供直接监督,根据任务要求针对性地优化候选区域的特征提取过程。其中,外观编码网络旨在提取候选区域的视觉语义特征,特征信息应当具有足够的判别性,能够区分处于不同阶段的细胞和背景区域。基于上述考

虑,本文设计了分类辅助任务对外观编码网络进行预训练。其基本思想在于:如果编码网络能够正确地分类候选区域,说明其学习到了类别间的判别性信息,具备识别类别特有的外观模式的能力。

如图3(a)所示,外观编码网络 f 由3个卷积核大小分别为7、6和4的卷积模块构成,每个卷积模块由一个不使用边界填充(padding)的卷积层和一个线性整流函数(rectified linear unit, ReLU)构成。前两个卷积模块后分别用一个最大池化层降低特征分辨率。设输入样本为区域块 I ,经过外观编码网络处理后得到外观特征 $f(I)$ 。将外观特征展开成一维特征向量后输入两个全连接层组成的分类器,最终得到三分类的预测结果,分别对应分裂前、分裂后和未分裂区域。预测结果通过均方误差(mean square error, MSE)损失进行监督,具体为

$$L_1 = L_{\text{mse}}(\varphi(f(I)), Y) \quad (2)$$

式中, φ 为分类器函数, Y 为候选区域的类别标签, L_{mse} 表示均方误差损失函数。

在外观编码预训练中,编码网络需要根据候选区域的局部外观特征进行静态的分类判断,这使得训练后的编码网络能够区分类别间的视觉语义差异,具备对于区域外观模式的感知能力。

1.3.3 运动编码预训练

由图4可见,细胞在分裂时刻前后的两帧中变化最为明显,结合相邻两帧的运动特征有助于理解细胞状态的变化过程,改善检测结果(Mao和Yin, 2016)。本文提出基于图像重构的辅助任务来进行运动编码网络的预训练,先用运动编码网络提取相邻两帧图像间的运动特征,再将运动特征与其中一帧的外观特征结合来预测和重构另一帧图像。其基本原理在于设置信息瓶颈来迫使网络学习运动特征,只有当编码网络能有效提取相邻两帧间的运动和变化信息,才能利用其中一帧来重构另一帧。

如图3(b)所示,运动编码预训练过程的输入为时序相邻、空间位置相同的区域块 I_t 和 I_{t+1} 。首先对两个区域图像分别用外观编码网络提取外观特征 $f(I_t)$ 和 $f(I_{t+1})$,然后将两个区域块同时输入运动编码网络 g 提取帧间运动特征 $g(I_t, I_{t+1})$ 。运动编码网络的结构与外观编码网络相同,但将第1个卷积层改为两通道输入,使网络能够同时感知两帧图像。为保证提取到的特征确实包含了运动状态信息,采

用基于图像重构的自监督学习进行运动编码网络的训练。对于图像块 I_t ,将运动编码特征和外观编码特征按通道方向连接后,输入运动解码网络 d 。该解码网络由三阶段卷积构成,每个阶段卷积都使用两个带 ReLU 激活函数的 3×3 卷积层。前两个卷积阶段分别用一个反卷积(即转置卷积)层进行特征上采样,最终输出与输入图像块具有相同尺度的预测图像块 $M_{t+1} = d(g(I_t, I_{t+1}), f(I_t))$,该预测结果用输入图像块 I_{t+1} 进行监督。同样地,将图像块 I_{t+1} 的外观编码特征与运动编码特征输入运动解码网络,输出预测结果为 $M_t = d(g(I_t, I_{t+1}), f(I_{t+1}))$,用输入图像块 I_t 进行监督。整体的损失函数为

$$L_2 = L_{mse}(M_t, I_t) + L_{mse}(M_{t+1}, I_{t+1}) \tag{3}$$

由式(3)可见,运动编码网络的预训练是一个自监督学习过程,不需要额外的数据标注,只利用图像本身的重构就可以实现有效的训练监督。图像重构任务天然构成信息瓶颈,由于两帧图像并不相同,运动编码特征必须能够表征从一帧图像到另一帧图像的动态变化,才能提供恢复另一图像块所需的信息。因此,该训练过程可以使得运动编码网络有效理解细胞在不同阶段下所特有的运动模式。

1.4 空间—时间全模型训练与预测

预训练使空间编码网络能够充分提取候选区域的外观和运动特征,为时序的状态记忆和预测提供所需信息。预训练结束后,将空间编码网络和时序

网络级联进行空间—时间全模型(CNN—双向 LSTM 网络)的端到端训练,通过优化序列图像的时空特征表示来预测有丝分裂细胞。

如图 5 所示,模型输入为 1.2 小节预处理生成的候选区域序列,由 1.3.2 小节中预训练得到的外观编码网络(N1)提取区域外观特征,由 1.3.3 小节中预训练得到的运动编码网络(N2)提取区域运动特征。将两种区域特征按通道方向连接后输入双向的 LSTM(Hochreiter 和 Schmidhuber, 1997)时序网络进行特征的时序状态记忆与更新。时序网络由 12 组时序单元组成,每组时序单元包含一个前向 LSTM 单元和一个反向 LSTM 单元,通过时序信息的传播动态地更新区域状态。每个 LSTM 单元由遗忘门、输入门和输出门组成。遗忘门从区域状态中丢弃无用信息;输入门接收当前时刻的新信息来更新区域状态;输出门输出当前区域状态并进行时序信息传播。在 t 时刻,LSTM 单元的处理过程可表示为

$$\begin{cases} i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\ f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\ o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ h_t = o_t \odot \tanh(c_t) \end{cases} \tag{4}$$

式中, i_t 为输入门结果, f_t 为遗忘门结果, o_t 为输出门结果, c_t 为更新后的单元状态, h_t 为用于传播和输出的最终状态; W 和 b 为可学习的权重和偏置参数, σ

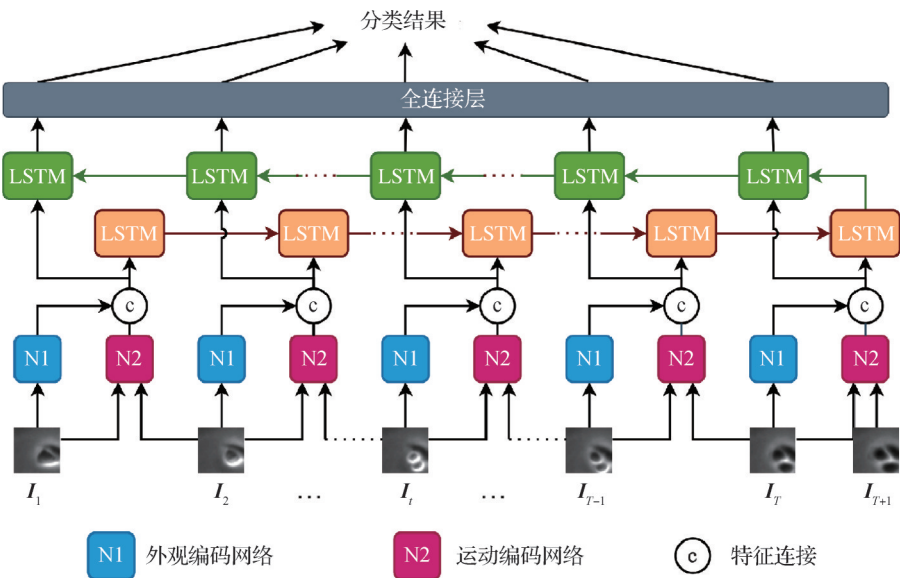


图 5 CNN—双向 LSTM 检测网络
Fig. 5 CNN—bidirectional LSTM detection network

是 sigmoid 激活函数; $\mathbf{x}_t$ 是将 N1 和 N2 网络的输出连接后得到的外观和运动联合特征; $\odot$ 表示对应位置相乘。由于预训练过程对 N1 和 N2 网络进行了直接的监督训练, 在全模型训练初期就能为时序网络提供有效的语义信息, 进而优化时序单元的训练过程。

最后, 将每组时序单元的输出经过两层全连接层映射得到 3 个预测值, 分别对应分裂前、分裂后和未分裂状态, 并由 sigmoid 函数归一化得到对该时刻类别的概率预测 $\hat{y}_t$ 。训练时通过分类标签来进行端到端的监督学习, 用均方误差损失进行监督, 具体为

$$L = \frac{1}{T} \sum_{t=1}^T L_{\text{mse}}(\hat{y}_t, y_t) \quad (5)$$

式中, $\hat{y}_t$ 表示 t 时刻图像的预测结果, y_t 表示 t 时刻的类别标签, T 为输入图像序列的帧数。测试时, 记录属于分裂后区域的预测概率, 若其最大值大于阈值 (设为 0.9), 将其所在时空位置作为检测到的有丝分裂中心点。

2 实验与分析

2.1 实验环境与设置

实验环境条件如下: 在 Ubuntu16.04 系统平台, 使用单个具有 11 GB 显存的 GTX 1080Ti 显卡, 基于 PyTorch1.3.0 深度学习框架进行训练和测试。

训练过程的参数设置如下: 第 1 阶段训练预分割网络时, 批量大小设置为 5, 以 0.001 的学习率迭代训练 500 个周期; 第 2 阶段编码网络预训练时, 对于外观编码网络和运动编码网络, 批量大小分别设置为 4 000 和 200, 两个网络都以 0.000 5 的学习率迭代训练 100 个周期; 第 3 阶段全模型训练时, 批量大小设置为 400, 以 0.000 05 的学习率迭代训练 100 个周期。训练过程使用 Adam 优化器。

2.2 数据集

实验数据集是由 CVPR2019 (IEEE Conference on Computer Vision and Pattern Recognition 2019) 有丝分裂细胞检测竞赛提供的 C2C12-16 数据集 (Su 等, 2021)。该数据集包含 16 个序列的成肌干细胞全载玻片相衬显微图像。每个序列中, 细胞被随机放置于培养皿中并每隔 5 min 进行一次图像采样, 共得到 1 013 帧分辨率为 $1\,392 \times 1\,040$ 像素的灰度图像。所有序列根据培养过程添加的不同药物分为 4 组, 其中第 1 组不添加任何药物, 第 2 组单独添加

成纤维生长因子 2 (fibroblast growth factor-2, FGF2), 第 3 组单独添加骨形成蛋白 2 (bone morphogenetic protein-2, BMP2), 第 4 组同时添加 FGF2 和 BMP2。每组第 1、2 个序列作为训练集, 第 3 个序列作为验证集, 第 4 个序列作为测试集。图 6 对不同培养条件下的图像样本进行了比较, 可见第 2、4 组图像包含更多与细胞分裂事件相关的高亮度区域, 场景更加复杂。

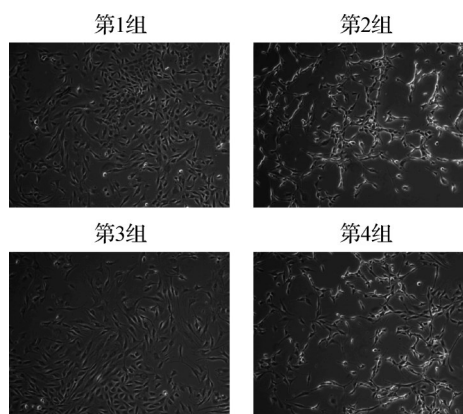


图6 不同培养条件下的图像对比

Fig. 6 Comparison of images under different conditions

2.3 性能评价指标

采用平均精确率、平均召回率和平均 F 得分作为对检测性能的评价指标。根据数据集官方设置, 若检测出的有丝分裂细胞中心点位置与真实标签位置在时序上相差 6 帧以内, 在空间上相差 15 个像素以内, 则认为检测结果正确。此外, 检测结果可根据对错分为真阳性 (true positive, TP)、假阳性 (false positive, FP)、真阴性 (true negative, TN) 和假阴性 (false negative, FN)。

精确率计算为

$$P = \frac{TP}{TP + FP} \quad (6)$$

召回率计算为

$$R = \frac{TP}{TP + FN} \quad (7)$$

F 得分计算为

$$F = \frac{2P \times R}{P + R} \quad (8)$$

因此, F 得分是结合精确率和召回率的综合性指标。

2.4 实验结果

2.4.1 与现有方法的比较

本小节分别在 C2C12-16 数据集的验证集和测

试集上与EDCRF(event detection conditional random field)(Huh 等, 2011)、V-Net(Nishimura 和 Bise, 2020)、MM-HCRF(max-margin hidden conditional random field)(Liu 等, 2012)、3DCNN(Nie 等, 2016)、CNN+LSTM(Su 等, 2017)、SSG-DNN(sequential saliency guided deep neural network)(Lu 等, 2020)、PSSD(progressive sequence saliency discovery network)(Su 等, 2022)和MViTv2(multiscale vision transformers: version 2)(Li 等, 2022)方法进行比较。其中,EDCRF和MM-HCRF属于基于条件随机场的方法;V-Net和3DCNN属于使用三维卷积网络的方法,CNN+LSTM和SSG-DNN属于结合二维卷积网络和时序网络的方法,PSSD则进一步引入了强化学习。此外,MViTv2是使用Transformer架构的最新的通用目标检测方法。本文选取不同类型的方法进行对比,从而实现更加全面的分析。

本文方法与EDCRF、V-Net和MViTv2方法在验证集上的比较结果如表1所示。EDCRF采用尺度不变特征变换(scale-invariant feature transform, SIFT)算子,其在第2组检测中F得分为39.4%,而本文方

法在该组得分达到85.0%,显示了其对于不同培养模式下得到的图像具有更好的泛化性。V-Net使用三维卷积直接提取时空特征,模型参数数量和计算量大,其预测结果召回率较高,而精确率较低。本文方法在4组序列上的平均F得分达到87.2%,比V-Net高出4.7%,具有更好的整体性能。此外,V-Net需要对每组序列训练不同的最优模型进行检测,但其在不同组中的性能表现仍然差异较大,最低精确率(第2组)为64.6%,最低召回率(第4组)为74.6%。相比之下,本文方法仅使用单模型进行全部4组序列的预测,最低精确率(第4组)达到79.4%,最低召回率(第2组)达到85.3%,对不同组数据始终能得到较为稳定的检测结果。为了与最新的通用检测方法进行比较,本文用MViTv2提供的代码和模型在本文任务上进行了重新训练和测试。MViTv2使用高复杂度的Transformer架构,通过级联的自注意力模块实现特征编码,而本文方法只使用简单的卷积编码网络,通过辅助任务的预训练就获得了更加有效的区域特征,平均F得分比MViTv2(79.2%)高8.0%。

表1 C2C12-16验证集检测性能比较
Table 1 Comparison of detection performance on C2C12-16 validation set

方法	F0003 (第1组)			F0007 (第2组)			F00011 (第3组)			F00015 (第4组)		
	精确率	召回率	F得分	精确率	召回率	F得分	精确率	召回率	F得分	精确率	召回率	F得分
EDCRF	69.9	76.5	73.1	34.7	45.4	39.4	84.0	84.5	84.3	53.9	60.4	56.9
V-Net	85.7	89.8	87.7	64.6	84.1	73.1	86.4	94.7	90.4	83.1	74.6	78.6
MViTv2	79.9	86.7	83.1	66.3	81.6	73.2	82.9	88.4	85.5	66.3	86.4	75.0
本文	88.9	86.5	87.7	84.6	85.3	85.0	88.4	95.0	91.6	79.4	90.4	84.5

注:加粗字体表示各列最优结果。

本文方法与MM-HCRF、3DCNN、CNN+LSTM、SSG-DNN、PSSD和MViTv2方法在测试集上的比较结果如表2所示。基于神经网络的框架整体性能优于使用SIFT特征算子的MM-HCRF。其中,本文方法在4组序列的平均F得分(86.2%)显著高于基于三维卷积网络的3DCNN方法。相比于同样使用空间—时间模型架构的CNN+LSTM和SSG-DNN,本文方法由于结合了静态的细胞外观信息和动态的运动变化信息,并针对两种编码网络构造了不同的辅助任务进行预训练,能更好地为时序网络的记忆和信

息融合提供判别性的视觉语义,平均F得分比CNN+LSTM(70.1%)高16.1%,比SSG-DNN(79.6%)高6.6%。尤其在对较为困难的第2、4组序列检测中,本文方法在保持较高召回率的同时,精确率分别达到了83.3%和83.2%,远高于其他方法,证明其对于复杂的图像场景和难以区分的样本仍然具有稳定的检测性能。

近期研究中,PSSD引入了深度强化学习来筛选序列中的关键帧,可以从长序列中关注到有丝分裂附近的时刻。从实验结果上看,PSSD在C2C12-16

表 2 C2C12-16 测试集检测性能比较
Table 2 Comparison of detection performance on C2C12-16 test set

方法	F0004 (第 1 组)			F0008 (第 2 组)			F0012 (第 3 组)			F0016 (第 4 组)		
	精确率	召回率	F 得分	精确率	召回率	F 得分	精确率	召回率	F 得分	精确率	召回率	F 得分
MM-HCRF	67.4	92.3	77.9	22.8	87.5	36.1	64.5	99.6	78.3	20.4	89.4	33.2
3DCNN	87.0	77.5	81.9	25.6	80.2	38.8	89.0	85.2	87.1	23.2	81.7	36.2
CNN+LSTM	63.8	88.9	74.3	51.1	92.8	65.9	66.7	98.0	79.4	48.2	83.3	61.1
SSG-DNN	90.6	82.5	86.4	60.4	84.3	70.4	87.3	94.7	90.9	60.1	85.8	70.7
PSSD	94.4	84.7	89.3	76.1	85.1	80.3	90.3	98.0	94.0	75.9	82.9	79.3
MViTv2	79.4	83.6	81.4	67.5	79.8	73.1	83.2	89.6	86.3	66.3	81.8	73.2
本文	89.1	86.0	87.5	83.3	79.1	81.2	90.1	94.8	92.4	83.2	84.5	83.8

注:加粗字体表示各列最优结果。

测试集的第 1、3 组测试中取得较高精度,而在第 2、4 组测试中结果较差。这可能是由于第 1、3 组的测试数据与训练数据更为接近,更有利于基于强化学习的帧选择策略发挥作用。此外,PSSD 方法的性能受到所使用的显著性帧数量这一超参数影响,不同超参数的选择会影响不同组的检测效果。从整体性能而言,PSSD 在 4 组序列中的 F 得分为 $85.7 \pm 6.2\%$,而本文方法的 F 得分为 $86.2 \pm 4.2\%$,具有更高的平均预测精度,同时预测结果也更加稳定。此外,与最新的通用图像检测方法 MViTv2 相比,本文方法在训练集和测试集上都取得明显更优的性能表现,证明相比于在模型结构上进行调整,本文的预处理和预训练策略可以更好地提升复杂场景下的细胞检测能力。

2.4.2 消融实验

为检验不同模块在检测过程中的作用,本文在 C2C12-16 测试集上进行了消融实验(取 4 组平均结

果)。整体框架分为 3 个阶段:1)候选区域预处理阶段(S1)从图像序列中初步筛选出可能存在有丝分裂细胞的候选区域。2)编码网络预训练阶段(S2)构造辅助任务,训练外观和运动编码网络,学习从候选区域中提取判别性的语义特征。3)空间—时间全模型训练与预测阶段(S3)将预训练的空间编码网络与时序网络结合,通过端到端训练对预处理后的候选区域进行序列预测,得到最终的检测结果。

本文方法在不同阶段的检测性能如表 3 所示。可见,预处理中使用深度网络进行兴趣区域预分割能有效地对潜在的候选区域进行筛选,同时保持很高的召回率(97.3%),充分保留处于分裂期的细胞区域。应用算法 1 进行候选区域优化后,精确率从 12.7% 上升到了 58.5%,说明该算法能够基于简单的时空先验对冗余的候选区域进行有效滤除,使得测试时的候选区域更为精简。全模型预测阶段基于时空特征进行预测,充分考虑了细胞的时序状态变

表 3 不同阶段的检测性能比较
Table 3 Comparison of detection performance at different stages

预处理(S1)		预训练(S2)	全模型训练(S3)	精确率	召回率	F 得分
预分割	区域优化					
√	—	—	—	12.7	97.3	22.4
√	√	—	—	58.5	90.2	70.8
√	√	—	√	76.3	84.5	80.1
√	√	√	√	86.4	86.1	86.2

注:加粗字体表示各列最优结果,“√”表示使用相应模块,“—”表示未使用相应模块。

化,将F得分从70.8%提升到80.1%。而预训练阶段通过对外观编码网络和运动编码网络进行直接的监督学习,大大提升了编码网络对于不同细胞模式的语义感知能力,使S3阶段的全模型训练更加充分,将F得分提高到了86.2%。

表4展示了预处理阶段采用不同策略进行区域筛选的检测性能。现有方法主要根据亮度差异进行遍历搜索。其中,亮度均值法(Huh等,2011; Mao和

Yin, 2016)对图像进行均值滤波后设置阈值提取候选区域,达到1.2%的精确率(Mao和Yin, 2016)。亮度标准差法(Su等, 2017)计算三维滑窗内的亮度标准差并设置阈值提取候选区域。本文对标准差法进行了复现,精准率仅为0.1%。相比于基于亮度的预处理方法,本文将预分割网络和先验优化算法结合的方式可以达到更高的精确率(58.5%),显著减少了候选区域中的冗余结果,提升检测效率。

表4 不同预处理方法的比较
Table 4 Comparison of different pre-processing methods

预处理方法	精确率/%	召回率/%	区域数量
均值滤波法	1.2	—	—
标准差法	0.1	96.8	1 220 124
预分割(本文)	12.7	97.3	9 641
预分割+区域优化(本文)	58.5	90.2	1 973

注:“—”表示无相应数据。

为了进一步细化不同编码网络训练策略对于检测性能的影响,根据是否使用外观编码和运动编码预训练进行了消融实验,结果如表5所示。

表5 不同预训练策略的比较
Table 5 Comparison of different pre-training strategies
/%

预训练策略	精确率	召回率	F得分
无预训练	76.6	85.3	80.7
外观编码	82.1	86.6	84.3
运动编码	83.0	83.1	83.0
外观+运动	86.4	86.1	86.2

注:加粗字体表示各列最优结果。

不使用编码网络预训练时,全模型训练阶段的编码网络参数随机初始化,不具备有效的特征提取能力,难以为时序网络提供有效的细胞状态信息。而模型训练的反向传播过程需要经历全连接层、级联的LSTM时序单元之后才能作用于CNN编码网络,导致编码网络优化缓慢,进而使全模型收敛后的性能受限,最终检测结果的F得分为80.7%。使用基于分类辅助任务的外观编码预训练后,外观编码网络能够为预测图像块类别提供必要的空间模式信息,在精确度和召回率指标上获得了全面提升。使用基于图像重构的运动编码预训练后,网络能够从

区域外观变化中提取细胞的运动模式信息。单独对运动编码网络进行预训练,精确率提升6.4%,但召回率有所下降,这可能是由于网络过于依赖运动特征而忽视外观信息,导致对变化过程不明显的区域产生漏检。结合外观和运动编码可以有效解决对于细胞模式的单一类型感知导致的问题,在保持召回率的前提下大幅度提升精确率,最终将F得分指标提升了5.5%。

图7对使用预训练前后的检测结果进行了对比。其中青色边框表示真阳性(TP)检测结果,红色边框表示假阳性(FP)检测结果,黄色边框表示假阴性(FN)检测结果。由图7(a)可见,外观编码预训练使单帧图像的特征编码具有细胞类别的判别能力,相比于不进行预训练的模型能够更有效地排除假阳性,同时减少漏检的情况。由图7(b)可见,一些不处于分裂期的细胞也具有和分裂前细胞相似的外观特征,运动编码预训练使模型能结合候选区域的运动变化进行判断,识别出不符合有丝分裂变化过程的案例。

在空间—时间全模型训练和预测阶段,本文使用LSTM网络进行时序信息融合。LSTM网络中特征逐帧传播的过程天然具有时序形式,适合于建模有丝分裂细胞的时序状态变化过程。除LSTM外,Transformer模块同样可以实现序列化的信息处理和预测,但其时序信息需要通过在输入端加入时间位

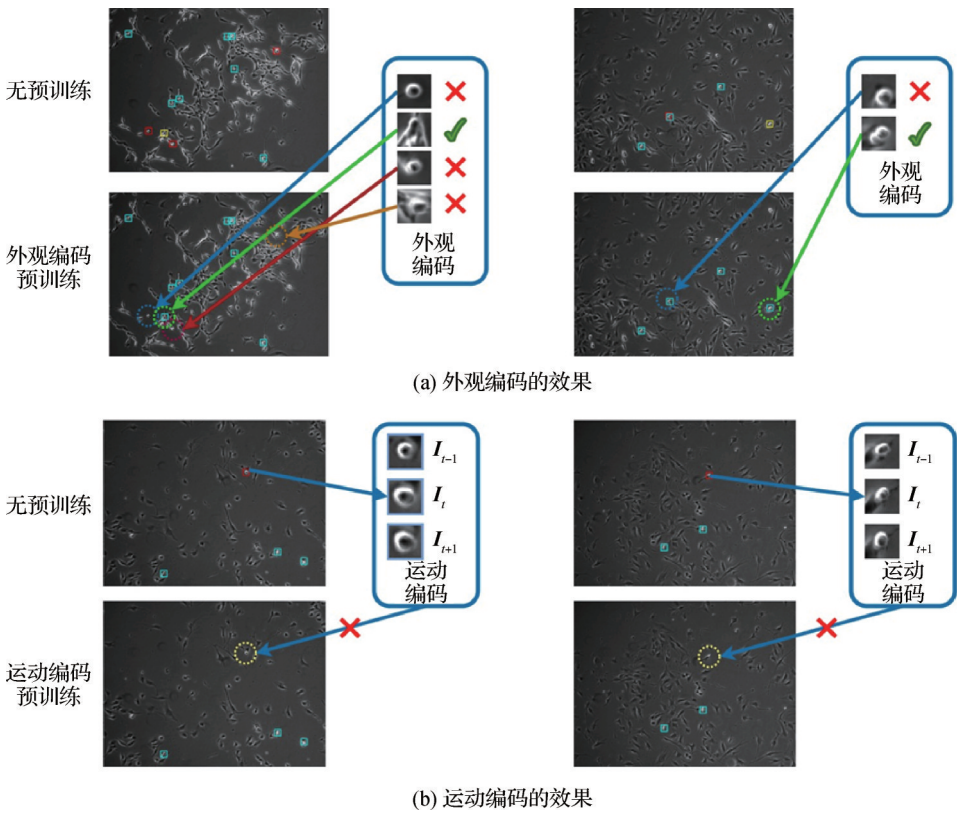


图 7 预训练前后检测结果对比

Fig. 7 Comparison of detection results before and after pre-training
((a) the effect of appearance encoding; (b) the effect of motion encoding)

置编码来实现。同时,相比于 LSTM 的逐帧传播,Transformer 模块需要密集的帧间交互,具有更大的模型容量和更高的计算复杂度。为了更直观地进行比较,将本文方法中的 LSTM 网络替换为 Transformer 编码网络(采用原论文设置)进行实验,表 6 展示了对应方法的平均 F 得分、单序列输入下所需的每秒浮点数计算量(floating-point operation per second, FLOPs)以及模型参数量。将 LSTM 网络替换为 Transformer 进行时序建模增加了模型参数量和计算量,同时检测精度反而降低。这可能有两方面原因:一是 Transformer 网络的模型复杂度较高,需要较大数据量的训练或更加精细的初始化策略来达到较好的性能(Raghu 等, 2021; Matsoukas 等, 2021)。二是 Transformer 通过额外的位置编码对时间维度进行建模,在有丝分裂细胞检测任务中其时序建模效果和 LSTM 网络存在差距。本文最终选择 LSTM 以更加适应于任务本身的数据规模和性能需求。

2. 4. 3 可视化与误差分析

图 8 展示了按时间顺序对不同组的检测结果进行的可视化。随着时间推移,相衬显微图像中的场

表 6 不同时序网络检测性能和代价对比

Table 6 Comparison of detection performance and cost using different temporal networks

时序网络	平均 F 得分/%	FLOPs/G	参数量/M
Transformer	83.1	0.96	37.2
LSTM	86.2	0.63	9.4

景变得更加复杂,发生的有丝分裂情况也更多。由图 8 可见,在不同分组的培养条件和不同的时序阶段下,本文方法都能从复杂场景中排除背景和未分裂细胞的干扰,检测出大部分有丝分裂细胞。

本文方法也存在一些预测失败的案例。如图 9 所示,当候选区域具有典型的外观和运动模式时,本文方法可以做出精准的识别和检测,但当细胞外观语义模糊以及运动特征不明显时,会出现一些误检(FP)和漏检(FN)的情况,这可能是由于训练过程缺乏对于此类样本的学习。

表 7 对检测结果的时序偏差和空间偏差进行了量化统计。在 4 组测试序列中,本文方法的最小时序偏差均值为 0. 203 帧,最大时序偏差均值为 0. 242

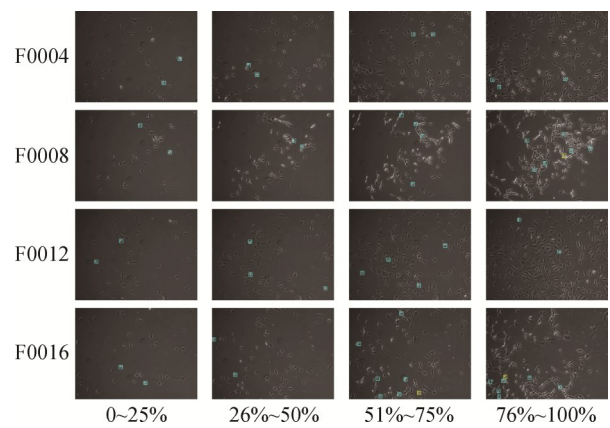


图8 时序检测结果可视化
Fig. 8 Visualization of temporal detection results



图9 误检测案例分析
Fig. 9 Analysis of failure cases

帧;最小空间偏差均值为3.137像素,最大空间偏差均值为3.561像素。可见,本文方法对不同组实验条件下的图像检测能力相差不大,检测结果具有一致性。此外,时序和空间上的最大标准差仅为0.618帧和2.603像素,体现了本文方法的预测性能具有较强的稳定性。

表7 检测结果的时空偏差

Table 7 Spatiotemporal biases of detection results

序列	时序偏差/帧	空间偏差/像素
F0004(第1组)	0.203±0.548	3.149±2.339
F0008(第2组)	0.217±0.618	3.393±2.603
F0012(第3组)	0.242±0.466	3.561±2.603
F0016(第4组)	0.224±0.438	3.137±2.152
平均	0.221±0.536	3.321±2.461

图10对比了本文方法和CNN+LSTM(Su等,2017)以及PSSD(Su等,2022)在不同组序列的时序偏差。其中,柱状图为偏差平均值,折线为偏差的标准差。由图10可见,CNN+LSTM方法的检测结果时序偏差较大(平均1.05帧),且不同组的检测偏差变化范围大,对于第2、4组的困难场景的检测偏差显著高于第1、3组中的偏差。与之相比,本文方法在4组序列中的表现较为平均,时序偏差均始终保持在不到0.25帧的低数值。这主要是由于本文方法

采用了编码网络预训练策略,能够从复杂多样的图像场景中关注到候选区域的外观和运动变化等本质信息,使得时序预测结果更加稳定可靠。相比于PSSD方法,尽管本文没有显式的显著性帧检测过程,但编码网络可以有效识别具有分裂期特征的关键时序,达到更低的检测偏差均值和标准差。

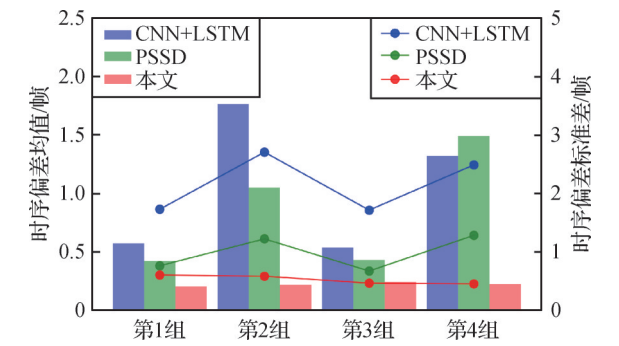


图10 不同方法的时序偏差对比
Fig. 10 Comparison of temporal biases of different methods

3 结 论

本文针对复杂显微图像场景中的有丝分裂细胞检测任务,设计了基于外观和运动模式感知的模型学习框架。该框架由3个阶段组成,在第1阶段,结合深度预分割网络和候选区域优化算法对长序列高分辨率的电镜图像序列进行预处理,筛选可能存在有丝分裂细胞的候选区域;第2阶段构造辅助任务来预训练外观和运动编码网络,对候选区域的编码过程实现直接的监督优化,提供具有判别性的区域内细胞模式信息;第3阶段基于预训练权重进行空间—时序网络的全模型训练和预测,结合时空特征从候选区域序列中检测出有丝分裂细胞。与现有方法相比,本文的预处理策略结合了可学习的有参网络和基于先验假设的无参优化过程,可以更好地在保持目标召回率的同时过滤无关区域,大幅度精简后续预测过程。本文提出的预训练策略可以为编码网络提供检测任务所需的区域状态感知能力,进而使时序网络更加充分地融合时空信息,提升全模型的预测能力。在C2C12-16数据集上的实验结果显示,本文方法达到了比其他方法更强的检测性能和更好的稳定性,且在不同实验条件下的图像检测效果具有一致性。

本文方法的重点在于设计预处理策略来提高检

测效率和设计预训练策略来优化编码网络的训练过程, 而没有在数据分布层面针对任务进行设计。后续研究可以尝试更加多样和均衡的数据增广策略来强化模型对少见样本的检测能力。

参考文献(References)

- Cao J L, Li Y L, Sun H Q, Xie J, Huang K Q and Pang Y W. 2022. A survey on deep learning based visual object detection. *Journal of Image and Graphics*, 27(6): 1697-1722 (曹家乐, 李亚利, 孙汉卿, 谢今, 黄凯奇, 庞彦伟. 2022. 基于深度学习的视觉目标检测技术综述. *中国图象图形学报*, 27(6): 1697-1722) [DOI: 10.11834/jig.220069]
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]
- Debeir O, van Ham P, Kiss R and Decaestecker C. 2005. Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *IEEE Transactions on Medical Imaging*, 24(6): 697-711 [DOI: 10.1109/TMI.2005.846851]
- Gallardo G M, Yang F X, Ianzini F, Mackey M and Sonka M. 2004. Mitotic cell recognition with hidden Markov models//*Proceedings of SPIE 5367, Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display*. San Diego, USA: SPIE: 661-668 [DOI: 10.1117/12.535778]
- Hochreiter S and Schmidhuber J. 1997. Long short-term memory. *Neural Computation*, 9(8): 1735-1780 [DOI: 10.1162/neco.1997.9.8.1735]
- Huh S, Ker DFE, Bise R, Chen M and Kanade T. 2011. Automated mitosis detection of stem cell populations in phase-contrast microscopy images. *IEEE Transactions on Medical Imaging*, 30(3): 586-596 [DOI: 10.1109/TMI.2010.2089384]
- Jiang X, Yuan Y X, Wang Y P, Xiao Z X, Zhu M L, Chen Z H, Liu T M and Shen D G. 2022. Review and prospect of medical imaging artificial intelligence in China in the past 20 years. *Journal of Image and Graphics*, 27(3): 655-671 (蒋希, 袁奕萱, 王雅萍, 肖振祥, 朱美芦, 陈泽华, 陈泽华, 刘天明, 沈定刚. 2022. 中国医学影像人工智能20年回顾和展望. *中国图象图形学报*, 27(3): 655-671) [DOI: 10.11834/jig.211162]
- Li K and Kanade T. 2009. Nonnegative mixed-norm preconditioning for microscopy image segmentation//*Proceedings of the 21st International Conference on Information Processing in Medical Imaging*. Williamsburg, USA: Springer: 362-373 [DOI: 10.1007/978-3-642-02498-6_30]
- Li K, Miller E D, Chen M, Kanade T, Weiss L E and Campbell P G. 2008. Computer vision tracking of stemness//*Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: from Nano to Macro*. Paris, France: IEEE: 847-850 [DOI: 10.1109/ISBI.2008.4541129]
- Li Y H, Wu C Y, Fan H Q, Mangalam K, Xiong B, Malik J and Feichtenhofer C. 2022. MViT2: improved multiscale vision transformers for classification and detection//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 4794-4804 [DOI: 10.1109/cvpr52688.2022.00476]
- Liang L C, Zhou X B, Li F H, Wong S T C, Huckins J and King R W. 2007. Mitosis cell identification with conditional random fields//*Proceedings of 2007 IEEE/NIH Life Science Systems and Applications Workshop*. Bethesda, USA: IEEE: 2007: 9-12 [DOI: 10.1109/lssa.2007.4400872]
- Liu A A, Hao T, Gao Z, Su Y T and Yang Z X. 2013. Nonnegative mixed-norm convex optimization for mitotic cell detection in phase contrast microscopy. *Computational and Mathematical Methods in Medicine*, 2013: #176272 [DOI: 10.1155/2013/176272]
- Liu A A, Li K and Kanade T. 2010a. Spatiotemporal mitosis event detection in time-lapse phase contrast microscopy image sequences//*Proceedings of 2010 IEEE International Conference on Multimedia and Expo*. Singapore, Singapore: IEEE: 161-166 [DOI: 10.1109/ICME.2010.5583299]
- Liu A A, Li K and Kanade T. 2010b. Mitosis sequence detection using hidden conditional random fields//*Proceedings of 2010 IEEE International Symposium on Biomedical Imaging: from Nano to Macro*. Rotterdam, the Netherlands: IEEE: 580-583 [DOI: 10.1109/ISBI.2010.5490279]
- Liu A A, Li K and Kanade T. 2012. A semi-Markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations. *IEEE Transactions on Medical Imaging*, 31(2): 359-369 [DOI: 10.1109/TMI.2011.2169495]
- Liu A A, Lu Y, Chen M and Su Y T. 2017a. Mitosis detection in phase contrast microscopy image sequences of stem cell populations: a critical review. *IEEE Transactions on Big Data*, 3(4): 443-457 [DOI: 10.1109/TBDATA.2017.2721438]
- Liu M, He Y, Wei Y L and Xiang P. 2017b. Plant cell tracking using Kalman filter based local graph matching. *Image and Vision Computing*, 60: 154-161 [DOI: 10.1016/j.imavis.2016.08.005]
- Lu Y, Liu A A, Chen M, Nie W Z and Su Y T. 2020. Sequential saliency guided deep neural network for joint mitosis identification and localization in time-lapse phase contrast microscopy images. *IEEE Journal of Biomedical and Health Informatics*, 24(5): 1367-1378 [DOI: 10.1109/JBHI.2019.2943228]
- Mao Y X and Yin Z Z. 2016. A hierarchical convolutional neural network for mitosis detection in phase-contrast microscopy images//*Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens, Greece: Springer: 685-692 [DOI: 10.1007/978-3-319-46723-8_79]

- Marcuzzo M, Guichard T, Quelhas P, Mendonça A M and Campilho A. 2009. Cell division detection on the arabidopsis thaliana root//Proceedings of the 4th Iberian Conference on Pattern Recognition and Image Analysis. Póvoa de Varzim, Portugal: Springer: 168-175 [DOI: 10.1007/978-3-642-02172-5_23]
- Matsoukas C, Haslum J F, Söderberg M and Smith K. 2021. Is it time to replace CNNs with transformers for medical images? [EB/OL]. [2021-08-20]. <https://arxiv.org/pdf/2108.09038.pdf>
- Milletari F, Navab N and Ahmadi S A. 2016. V-Net: fully convolutional neural networks for volumetric medical-image segmentation//Proceedings of the 4th International Conference on 3D Vision. Stanford, USA: IEEE: 565-571 [DOI: 10.1109/3DV.2016.79]
- Nie W Z, Li W H, Liu A A, Hao T and Su Y T. 2016. 3D convolutional networks-based mitotic event detection in time-lapse phase contrast microscopy image sequences of stem cell populations//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Las Vegas, USA: IEEE: 1359-1366 [DOI: 10.1109/cvprw.2016.171]
- Nishimura K and Bise R. 2020. Spatial-temporal mitosis detection in phase-contrast microscopy via likelihood map estimation by 3DCNN//Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Montreal, Canada: IEEE: 1811-1815 [DOI: 10.1109/EMBC44109.2020.9175676]
- Quelhas P, Mendonça A M and Campilho A. 2010. Optical flow based arabidopsis thaliana root meristem cell division detection//Proceedings of the 7th International Conference on Image Analysis and Recognition. Póvoa de Varzim, Portugal: Springer: 217-226 [DOI: 10.1007/978-3-642-13775-4_22]
- Raghu M, Unterthiner T, Kornblith S, Zhang C Y and Dosovitskiy A. 2021. Do vision transformers see like convolutional neural networks?//Proceedings of the 35th Conference on Neural Information Processing Systems. [s.l.]: NeurIPS: 12116-12128
- Su Y T, Lu Y, Chen M and Liu A A. 2017. Spatiotemporal joint mitosis detection using CNN-LSTM network in time-lapse phase contrast microscopy images. IEEE Access, 5: 18033-18041 [DOI: 10.1109/ACCESS.2017.2745544]
- Su Y T, Lu Y, Chen M and Liu A A. 2022. Deep reinforcement learning-based progressive sequence saliency discovery network for mitosis detection in time-lapse phase-contrast microscopy images. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19(2): 854-865 [DOI: 10.1109/TCBB.2020.3019042]
- Su Y T, Lu Y, Liu J, Chen M and Liu A A. 2021. Spatio-temporal mitosis detection in time-lapse phase-contrast microscopy image sequences: a benchmark. IEEE Transactions on Medical Imaging, 40(5): 1319-1328 [DOI: 10.1109/TMI.2021.3052854]
- Yang F X, Mackey M A, Ianzini F, Gallardo G and Sonka M. 2005. Cell segmentation, tracking, and mitosis detection using temporal context//Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention. Palm Springs, USA: Springer: 302-309 [DOI: 10.1007/11566465_38]

作者简介

- 林凡超,男,博士研究生,主要研究方向为计算机视觉与视频处理。E-mail: lfc1995@mail.ustc.edu.cn
- 刘传彬,通信作者,男,副研究员,主要研究方向为计算机视觉与医学影像分析。E-mail: liuch92@ustc.edu.cn
- 谢洪涛,男,教授,主要研究方向为多媒体内容分析与检索。E-mail: htxie@ustc.edu.cn
- 张勇东,男,教授,主要研究方向为多媒体内容分析与内容安全。E-mail: zhyd73@ustc.edu.cn