

基于密度峰值聚类的自适应细胞定位

谢经辉¹, 杨昊¹, 周航¹, 葛澍²

(1. 成都信息工程大学计算机学院, 四川成都 610225; 2. 四川省科学技术研究成果档案馆, 四川成都 610046)

摘要: 自动准确地定位生物医学成像图像中的大量细胞对生物医学研究具有重要意义。现有图像处理方法定位分布密集和黏连细胞的精度不高,且参数设置对数据十分敏感。为此,提出一种基于密度峰值聚类、自适应优化参数的细胞定位方法。首先,建立深度学习模型进行细胞分割以提升聚类效果;其次,分析图像前景区域局部密度与较高密度点的最小距离趋势性变化,以自动优化密度阈值和距离阈值参数来选取聚类中心,实现对密集、黏连细胞的自动定位。与常用的5种算法比较发现,所提方法在荧光微光学切片断层扫描成像小鼠数据集中的检出率、准确率分别为0.89和0.81,在更复杂的数据集中相较于比较方法更优,为细胞定位工作提供了一种新的高准确度自动化方法,在生物医学图像处理的计算机应用领域具有良好前景。

关键词: 密度峰值聚类; 细胞定位; 自适应; 机器学习

DOI: 10.11907/rjtk.232307

中图分类号: TP391

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)001-0129-07



Adaptive Cell Localization Based on Density Peak Clustering

XIE Jinghui¹, YANG Hao¹, ZHOU Hang¹, GE Shu²

(1. School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China;

2. Sichuan Provincial Archives of Scientific and Technological Research Achievements, Chengdu 610046, China)

Abstract: Automatically and accurately locating a large number of cells in biomedical imaging images is of great significance for biomedical research. The existing image processing methods have low accuracy in locating densely distributed and adherent cells, and parameter settings are highly sensitive to data. Therefore, a cell localization method based on density peak clustering and adaptive optimization parameters is proposed. Firstly, establish a deep learning model for cell segmentation to improve clustering performance; Next, analyze the trend change of the minimum distance between the local density of the foreground area and higher density points in the image, and automatically optimize the density threshold and distance threshold parameters to select clustering centers, achieving automatic localization of dense and adherent cells. Compared with five commonly used algorithms, it was found that the proposed method has a detection rate and accuracy of 0.89 and 0.81, respectively, in the fluorescence micro optical slice tomography imaging mouse dataset. It is superior to the comparison method in more complex datasets and provides a new high-precision automated method for cell localization work. It has good prospects in the computer application field of biomedical image processing.

Key Words: density-peak clustering; cellular localization; self-adaption; machine learning

0 引言

细胞定位研究对理解细胞生物学过程和疾病机制至

关重要,细胞组织异常也与多种疾病相关,例如癌症、神经退行性疾病和先天性异常等。很多疾病诊断与分级均依赖于识别生物医学图像中组织细胞的数量与形态,细胞定位则是神经元形态重建、细胞计数、空间分布定量分析的

收稿日期: 2023-12-28

扫描二维码阅读全文:



基金项目: 四川省科技计划项目(2022YFS0542, 2023JDZH0034); 四川省自然科学基金项目(2022NSFSC0964)

作者简介: 谢经辉(1997-),男,成都信息工程大学计算机学院硕士研究生,研究方向为高性能计算和图形图像处理;杨昊(1968-),男,博士,CCF会员,成都信息工程大学计算机学院副教授、硕士生导师,研究方向为高性能计算、智慧气象;周航(1991-),男,博士,成都信息工程大学计算机学院讲师,研究方向为机器学习、高性能计算;葛澍(1981-),男,四川省科学技术研究成果档案馆科技高级评估师,研究方向为科技管理。本文通讯作者:杨昊。

基本步骤,因此细胞自动分割与定位是计算机辅助病理诊断过程中十分重要的一步^[1,2]。

细胞精确定位不仅能辅助病理学家作出更精确的诊断,还能降低时间与人力成本,然而细胞定位任务领域仍存在许多难题。例如,细胞图像常受到背景干扰、光照变化和噪声等因素影响;细胞类型、细胞形态、细胞密度等差异较大,需要大量低效繁琐的人工操作^[3]。因此,为了提升量化神经元的效率和效果,分析具体数据时经常需要高度定制化的算法。

随着标记和成像技术的快速发展,研究者对3D神经元显微镜图像的自动神经元定位方法的需求不断增长,但神经元荧光显微镜图像中黏连细胞的自动定位面临着以下两个主要挑战:①密集分布的神经元图像中黏连细胞十分常见,自动方法容易错误地将多个黏连细胞识别为单个细胞;②大脑区域、细胞类型和荧光标记过程的变异性都会影响细胞图像不同区域的信号强度分布,即使在同一个神经元内也可能存在信号强度低于其他部分的情况^[5-8]。为此,本文提出一种基于密度峰值聚类、自适应优化参数的细胞定位方法解决上述问题。

1 相关研究

目前,不少学者已提出了多种黏连细胞自动定位方法。例如,Vincent等^[9]在图形图像处理领域提出分水岭模型。肖月等^[10]在细胞分割应用中提出基于梯度修正的改进分水岭模型。徐鸿哲等^[11]提出一种基于分水岭线的CV模型树冠分割优化算法,将图像的亮度或灰度值视为地势高度,在图像局部极小值处添加水滴,让水滴逐渐扩散并汇合形成水流从而形成最终的分割结果,此时便设定一个合适的阈值将图像分割为背景和前景(细胞)两部分。Winter等^[12]提出一种创新的像素复制方法检测单个椭圆物体的接触或重叠情况。Zhou等^[13]提出GTree方法从细胞中心开始构建细胞图像的多个细胞形态,通过构建以根为细胞局部坐标的图结构来修正神经元重建结果的拓扑结构。虽然这些方法有其自身优势,在神经科学研究中起着关键作用,但仅适用于某些特定场景,在复杂背景或噪声环境中的性能较差。例如当图像对比度不均匀时,基于阈值分割的算法经常会过度分割细胞;黏连细胞形态可能存在较大差异以致于算法无法识别不同形态的细胞,或在应用过程中可能面临依赖初始定位、优化算法收敛性、参数选择等问题。

针对细胞图像信号强度不一这一问题,可引入基于深度学习的方法进行解决。传统方法通常依靠手动提取特征或制定规则来分割细胞,这对大规模图像数据处理而言耗时较大且操作十分困难。近年来,深度学习方法在细胞分割领域取得了显著进展,能有效克服这些缺陷^[14,15]。例如,基于注意力机制的卷积神经网络的空间注意力机制自

适应融合了全局语义信息与局部语义信息,通道注意力机制可获取更有效的特征图^[16];在细胞定位任务中,采用Conv-LSTM模块的深度学习方法可在模型中融入细胞在时间上的变化信息,以提升细胞定位精度和鲁棒性^[17];基于多模态深度学习的癌症分类模型从两种模态数据整合、特征集构建、特征选择和多模型集成4个方面提升癌症分类模型的预测性能^[18]。Cellpose方法使用卷积神经网络学习细胞特征表示,并通过分割算法将细胞从图像中分离出来,训练过程包括对大量已标记的细胞图像进行监督学习,从而使网络学习到细胞形状、边界和其他特征^[19]。

虽然上述机器学习方法能更好地适应复杂场景、提取高级特征,在更大范围和更高准确性下进行细胞分割,并在细胞分割方面相较于传统方法自动化、适应性、鲁棒性、可扩展性、精度更高,但在黏连细胞的定位任务中机器学习方法的缺陷会被进一步放大。例如,机器学习模型对光照变化、细胞自身运动等噪声和干扰较为敏感,对于未见过的数据泛化能力有限,这就要求机器学习算法拥有更庞大的数据来训练模型,因此对计算资源要求较高,并且模型的改进和可解释性也有待提高。

研究表明,结合机器学习与传统图像处理方法进行预处理和后处理,以实现较高精度的细胞定位是个可行的解决方案。为此,本文提出一种基于密度峰值聚类算法(Density peak clustering, DPC)的自适应密度峰值聚类方法(Adaptive density peak clustering, ADPC)^[20-22]。该方法通过VoxResNet深度学习方法预分割处理数据且参数设置无需人工设置,还能根据数据特点自动调整参数^[23]。

2 本文方法

2.1 方法概述

基于密度峰值聚类的自适应细胞定位方法巧妙结合了深度学习在细胞分割领域的应用与改进的自适应密度峰值聚类算法,旨在克服传统方法在细胞分割方面存在的问题,解决参数依赖、人工耗时严重及对黏连细胞定位精度不高的问题。密度峰值聚类算法的核心思想是将高密度区域视为聚类中心,将低密度区域视为噪声或边界,通过计算每个样本点的局部密度和该点可达距离找到数据集中的密度峰值点,从而实现聚类。选择合适的参数对获得良好的聚类结果至关重要,但在实践中不同数据集需要重新设置参数,从而增加了算法复杂性,因此本文对密度峰值聚类算法的最小距离参数和最小密度参数进行了自适应改进,整体流程图如图1所示。

(1)对数据进行深度学习预分割处理,代替传统图像预处理方法并提取细胞区域。通过VoxResNet进行细胞图像预分割处理以处理不同类型、形状和大小的细胞,尤其适用于处理3D细胞图像中的复杂信息,相较于传统基于规则或手工设计特征的方法能更好地捕捉细微的细胞形

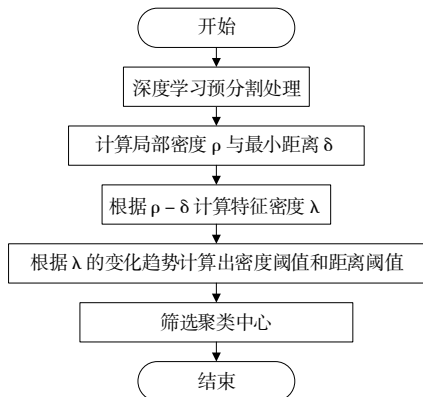


Fig. 1 Process of adaptive density peak clustering algorithm

图 1 自适应密度峰值聚类算法流程

态和结构变化, 从而提升模型分割准确性。

(2) 计算局部密度与更高密度点的最小距离。计算每个样本点与其他样本点的距离, 统计在其指定半径范围内的样本点数量(包括自身), 并将该数量作为局部密度度量。同时, 计算每个样本点与密度更高的邻近样本点距离。局部密度和与更高密度点的最小距离是密度峰值聚类算法核心思想中的关键参数, 能从多个维度分析细胞区域中的图像特征, 对精确识别黏连细胞具有重要意义。

(3) 使用改进的方法自适应确定细胞中心。计算最小密度与最小距离阈值可得每个样本点的得分, 得分越高证明该点越有可能是一个密度峰值, 从而确定哪些点是聚类中心。阈值参数影响着细胞定位结果的数量和准确度, 往往需要人工根据结果不断调整阈值参数以达到最优效果。本文方法提出的改进自适应方案能自动根据局部密度和与较高密度点的最小距离趋势性变化自动调整阈值参数, 既减少了人工成本, 又保证了模型精确度和召回率。

2.2 深度学习预分割

基于 VoxResNet 网络对原始数据进行细胞预分割流程如图 2 所示。VoxResNet 网络架构(见图 2(a))由堆叠的残差模块(即 VoxRes 模块)组成, 总共有 25 个体积卷积层和 4 个反卷积层, 如图 2(b)所示。其中, VoxRes 模块由两个卷积层和批量归一化层组成; 第二个卷积层输出与模块输入相加实现残差连接, 因此信息可直接在前向和后向传递; Conv1c、Conv4、Conv7 等 3 个卷积层步长为 2, 将输入体积的分辨率大小减小了 8 倍, 使网络具有较大的感受野, 因此可获得更多上下文信息, 从而提升了模型的判别能力; 中间插入批归一化层的作用是减少内部协变量漂移, 加速模型训练过程并提升性能。

2.3 数据集准备

本文数据集包括原始细胞图像和标记的二值图像, 标记的二值图像被用作神经网络训练的分割标准, 标注阶段由人工完成^[24-26]。本文从 3 种荧光光学切片断层扫描数据集中选择 128 张二维原始图像作为输入数据集, 每张原始图像大小为 240×240 px, 标注者使用 Amira 软件的细丝编辑器和分割编辑器定位神经元并标注中心区域, 标注

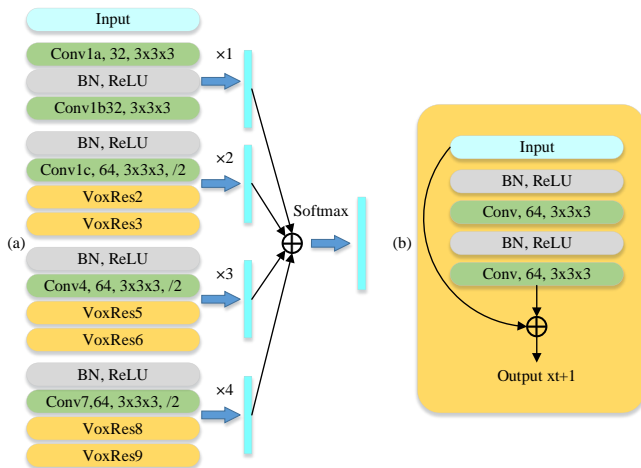


Fig. 2 Pre-process of VoxResNet network for segmenting

raw data cells

图 2 VoxResNet 网络分割原始数据细胞的预流程

后的二值图像的大小也为 240×240 px, 完成 128 张标注工作共需要 3~4 个工时。在训练阶段, 随机选择原始和对应标记的二值图像, 通过旋转、裁剪等方式增强图像来训练 VoxResNet 网络。图 3 展示了部分数据集在 Amira 软件中的三维视图(见图 3 左侧)、原始图像(见图 3 右上角)和标注后的图像(见图 3 右下角)。

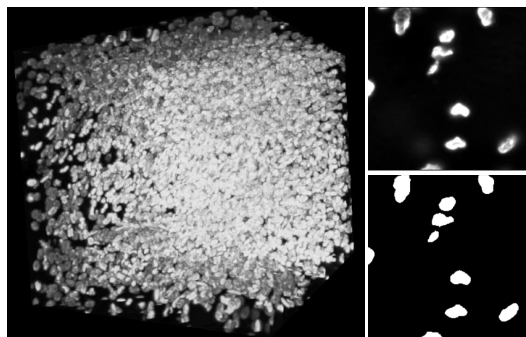


Fig. 3 Data set preparation

图 3 数据集准备

图 4 展示了 VoxResNet 深度学习方法与传统图像处理方法的预分割效果, 从左至右依次为原始图像、传统图像预分割、深度学习预分割。

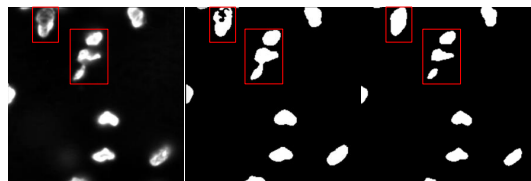


Fig. 4 Traditional image pre-segmentation methods and deep learning pre-segmentation performance

图 4 传统图像预分割方法与深度学习预分割效果

由图 4 可见, 深度学习方法的分割效果具有两个优势: ①能更好地补充因信号强度造成的细胞明暗不一问题; ②能在一定程度上简单分割黏连细胞, 有助于提升细胞定位效果。

2.4 实验设置

本文使用Adam优化器,设置初始学习率为0.0005,应用学习率步长调度器控制学习率下降,每100个epoch后学习率降低到先前的90%,当学习率低于0.0001时学习率停止下降,共训练20000个epoch,每个epoch包含100步。训练平台为Window 10操作系统、Intel 10600kf、NVIDIA RTX 3070、PyTorch深度学习框架。

2.5 局部密度与最小距离计算

在密度峰值算法中局部密度与最小距离是两个关键参数,用于识别聚类中心和划分簇,使密度峰值算法能找到最高局部密度,并将其他簇距离较远的数据点作为聚类中心,因此计算局部密度与最小距离至关重要且是算法最基础的步骤。局部密度衡量了图像中每个像素点在其周围邻域内的密集程度,通过计算每个像素点与其邻近像素点之间的距离和像素值得到该像素点的局部密度。通常情况下,将较高局部密度的像素点认为聚类中心或核心点,因为该点周围有相对较多的邻近点。具体而言,本文使用高斯核函数计算每个数据点的局部密度,通过式(1)计算局部密度 ρ 。

$$\rho_i = \sum_{j=1}^N K(\text{dist}(i, j)) \quad (1)$$

式中: N 表示数据集中的数据点数; $\text{dist}(i, j)$ 表示数据点 i 与 j 之间的欧氏距离; K 表示高斯核函数。

最小距离指一个像素点到更高局部密度的像素点之间的最短距离,也称为更高密度的最小距离。对于给定的图像和其中一个像素点 i ,计算最小距离首先要找到该图像中所有局部密度相较于像素点 i 更高的像素点 j ,其次计算像素点 i 与像素点 j 之间的距离并选择最短距离,具体计算公式如式(2)所示。

$$\delta_i = \frac{\min_{j: \rho_j > \rho_i} \|P_i - P_j\|_2}{\max_{j: \rho_j < \rho_i} \|P_i - P_j\|_2} \quad (2)$$

式中: P_i 表示像素点 i 的坐标; $\min_{j: \rho_j > \rho_i} \|P_i - P_j\|_2$ 表示当前区域中相较于像素点 i 的局部密度更高的点与像素点 i 的最近距离。

2.6 细胞中心自动定位

本文通过获取局部密度 ρ 和最小距离 δ ,使用改进的密度峰值算法定位细胞中心。聚类中心的特点是密度高于其邻居、距离相对较大且充当 ρ - δ 空间中的孤立点。综合考虑局部密度和最小距离对实现准确聚类结果具有重要作用。目前,确定细胞中心的方法由3部分组成:①构造二维 ρ - δ 空间,进行二维高斯卷积计算出特征密度 λ ;②使用自适应算法计算最小密度与最小距离阈值参数;③筛选出可能的聚类中心点,并移除冗余聚类中心。

2.6.1 特征密度计算

特征密度 λ 是综合考虑局部密度 ρ 与最小距离 δ 的参数,用来衡量像素点 i 在 ρ - δ 空间中的孤立程度。本文特征

密度 λ 由最小密度和最小距离构建的特征空间与二维高斯窗口进行卷积所得到的每个点的像素值。具体为,首先离散化特征空间,将局部密度与最小距离映射到一个二维空间中;其次统计落入每个网格的特征点数量;最后使用二维高斯窗口对生成的特征图像进行卷积得到一个滤波图像,其中就包含了特征点密度信息。具体的数学表达式如式(3)所示。

$$\lambda(i, j) = \sum_{k, l} (\mathbf{M}(i+k-5, j+l-5) * \mathbf{G}(k, l)) \quad (3)$$

式中: $\lambda(i, j)$ 表示特征密度的二维矩阵; \mathbf{M} 表示 ρ - δ 空间密集程度的二维矩阵; \mathbf{G} 表示 11×11 的二维高斯核。

最后将 $\lambda(i, j)$ 的值存储到特征密度序列,通过滤波图像中包含特征点区域的像素值 λ_i 来估计每个像素点的特征密度。

2.6.2 最小密度与最小距离阈值参数

最小密度阈值参数是筛选候选聚类中心的一个重要参数。首先过滤掉低密度数据点(噪音点),筛选出具有明显的高密度数据点,并将其作为密度峰;其次将特征密度 λ 从小到大排序,计算其上升趋势 $Trend$ 。

$$Trend[i] = \frac{\lambda_{i+1} - \lambda_i}{2\lambda_i - \lambda_{i-1}} \quad (4)$$

式中: λ_i 为当前特征密度 λ 值; λ_{i+1} 、 λ_{i-1} 分别代表前一时刻和后一时刻的特征密度 λ 值。

在获取上升趋势后,取上升趋势 $Trend$ 最大的点特征密度作为筛选聚类中心最小密度参数 λ_{min} 。图5为特征密度 λ 与上升趋势随序列顺序变化情况,第9个点具有最高的上升趋势,因此选取第9个点对应的局部密度作为最小密度阈值,即将前9个点选为候选聚类中心。

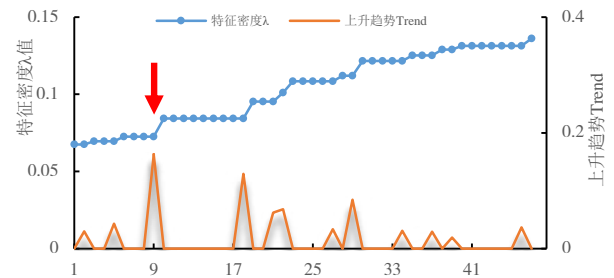


Fig. 5 Characteristic density and upward trend

图5 特征密度与上升趋势

最小距离阈值参数用于定义数据点间认定为邻居的距离阈值,也是筛选聚类中心的一个重要参数。通常经过实验验证或凭借经验获得数据集的细胞半径估计值。由于密度峰值算法性能很大程度上取决于数据集特点和分布情况,对数据集的高度依赖性也无形中增加了人工成本。为此,本文提出一种自动确定最小距离参数方法,如式(5)所示。

$$\delta_{min} = \frac{\theta}{\max_{i, j} \|P_i - P_j\|_2} \quad (5)$$

式中: p 表示该连通域中的点; θ 表示平均细胞距离, 是上升趋势最大的点前所有像素点的几何距离平均值, 即所有符合密度峰阈值点间的平均距离, 从而推算出符合当前连通域的最小距离 θ , 以达到自适应选取参数的效果。

2.4.3 确定细胞中心

在计算出最小密度 λ_{min} 与最小距离阈值 δ_{min} 后使用式(6)获得聚类中心点。

$$p_i | \lambda_i < \lambda_{min} \ \& \ \delta_i > \delta_{min} \quad (6)$$

式中: λ_i 表示像素点 p_i 的特征密度; δ_i 表示像素点 p_i 与较高密度点之间的最小距离。

如果满足式(6)则将该点视为候选聚类中心, 然后将这些候选聚类中心按照局部密度 ρ 、 δ 从大到小的顺序排列; 其次检查任意两个候选聚类中心之间的距离是否小于给定阈值 θ , 如果小于则将他们之间局部密度较小的点视为冗余点并删除; 最后将聚类中心视为细胞中心。

3 实验结果与分析

3.1 数据集与评价指标

本文选取了 3 种数据集: ①Data1 数据集来自荧光光学切片断层扫描 (fMOST) 成像小鼠数据集^[24]; ②Data2 数据集来自双光子荧光光学切片断层扫描系统 (2p-fMOST) 对小鼠大脑成像数据^[25]; ③Data3 数据集来自 Allen 脑图 (<https://portal.brain-map.org>) 开源荧光光学切片断层扫描 (fMOST) 成像小鼠大脑数据集^[26]。3 个数据集的原始体素大小分别为 $0.5 \times 0.5 \times 2 \ \mu\text{m}^3$, $0.5 \times 0.5 \times 2 \ \mu\text{m}^3$ 和 $0.3 \times 0.3 \times 1 \ \mu\text{m}^3$, 数据标注工作由领域专家和标注团队达成共识

后以多人修订的方式使用 Amira 软件完成。本文实验在每种数据集中分别取 5 组数据进行测试, 并与 DPC 聚类算法^[20]、Farsight 聚类算法^[27]、NeuroGPS 算法^[28]、Hu Method^[24] 的深度学习方法、Cellpose^[19] 深度学习方法进行比较。

本文采用召回率和准确率在不同数据集中衡量各算法性能。其中, 召回率 Recall 衡量模型对细胞的正确识别程度; 准确率 Accuracy 衡量模型定位的准确性。具体计算公式如式(7)、式(8)所示。

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

$$Accuracy = \frac{TP}{TP + FP} \quad (8)$$

式中: TP 指正确定位的细胞数量; FP 指错误定位的细胞数量; FN 指未定位出的细胞数量。

如果像素点 $i(x_i, y_i, z_i)$ 与标准细胞中心 (x_o, y_o, z_o) 满足式(9), 则判断像素点 i 为正确定位, 否则判断为错误定位。

$$(x_o - x_i)^2 + (y_o - y_i)^2 + (z_o - z_i)^2 < 3^3 \quad (9)$$

3.2 比较实验

图 6 展示了多种算法在 3 种数据集下细胞定位的效果图, Data1/2/3-Sub 表示 Data1/2/3 中矩形框部分的局部放大图。由此可见, ADPC 算法在 Data1、Data2 数据集上的细胞定位效果与金标准高度一致。

在细胞黏连程度更深的 Data3 数据集上, 原始 DPC 方法虽然能检测出细胞, 但会产生大量冗余结果, 原因为原始 DPC 方法对参数的依赖性较高, 无法充分把握数据的复杂特征, 需要人工干预来优化结果; Farsight 方法整体上呈现出良好的效果但性能较差, 处理黏连细胞时具有一定的

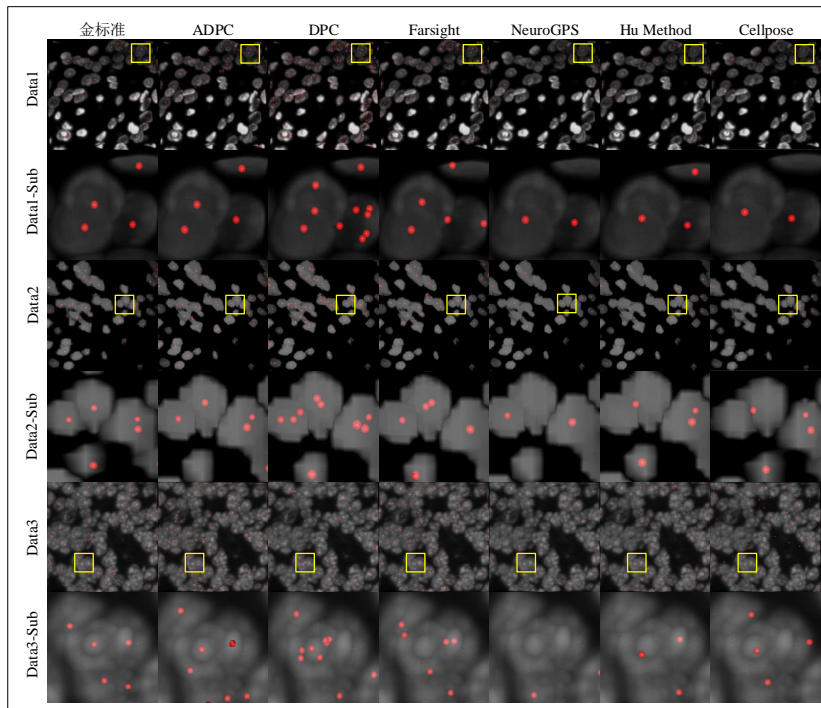


Fig. 6 Positioning effect of each method

图 6 各方法定位效果

局限性;ADPC算法表现相较于其他方法更出色;NeuroGPS检出的细胞较少,较难识别黏连细胞。

Hu Method 和 Cellpose 两种基于深度学习的细胞定位方法定位相对准确,细胞区域遗漏情况较少,但也存在识别黏连细胞困难的问题,在多个细胞重叠的情况下无法准确识别每个细胞。由 Data3-Sub 效果图可见,Cellpose 方法相较于 Hu Method 更好,原因为 Cellpose 模型具备开源、训练数据量大的优势,证明了深度学习细胞定位方法中数据资源对细胞定位能力和泛化能力具有积极作用。

图 7 为不同算法定位效果评价值的定量分析。对于 Data1、Data2 数据集而言,ADPC、DPC 算法、Farsight 算法均呈现出较高的召回率,ADPC 算法相较于原始 DPC 更稳定,精确率评价指标也明显优于原始 DPC 方法,原因为原始 DPC 方法产生了大量冗余结果,导致精确度较低。NeuroGPS 算法和 Hu Method 算法召回率较低但精确率很高,表示他们检测出的细胞数量相对较少。

Cellpose 方法由于仅支持导出二维图像的细胞定位数据,因此仅对该方法进行定性比较。在细胞定位任务中,在追求高准确率的同时高召回率也十分重要,高准确率保证了分割质量,而高召回率则保证了模型能有效捕捉感兴趣的细胞。在细胞黏连程度更深的 Data3 数据集中,ADPC 算法的召回率、精确率评价价值均为第一,原因为 ADPC 算法能根据数据特点自适应调整参数和处理策略,彰显了 ADPC 算法在此情境下的优势。

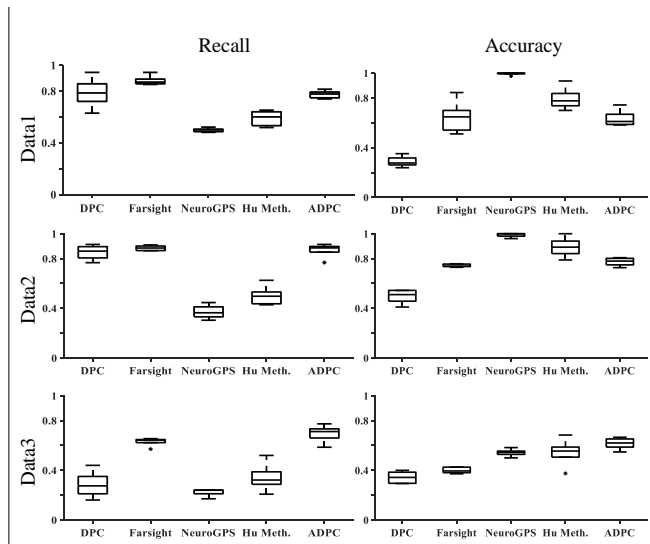


Fig. 7 Evaluation values of various methods

图7 各方法评价价值

4 实验讨论

本文所提方法的基本思路是改进现有基于密度峰值聚类的细胞定位方法,通过考虑局部密度 ρ 和最小距离 δ 之间的关系,提出一种根据数据特点自适应改进的方法,主要解决了不同数据集在原有细胞定位过程中效果差、过

度依赖参数、需要人工干预及传统方法细胞预处理效果不佳的问题。实验表明,所提方法在多种复杂度高、明暗程度不一致、细胞黏附程度高的 2p-fMOST 和 fMOST 小鼠大脑数据集上相较于原有方法和现有多种方法效果更优。

原始 DPC 方法在数据集上表现不佳的主要原因为无法适应数据集之间差异过大的特点;NeuroGPS 方法自身优势是能够识别含有轴突的图像;Hu Method 和 Cellpose 两种基于深度学习的细胞定位方法通常具备普适性、自动化的优势,但对非常小或非常大的细胞定位效果较差,对一些复杂的细胞形状和结构也可能出现误定位。

尽管机器学习在细胞定位问题中仍然具有巨大的潜力,但仍需要庞大的数据支撑和不断改良,对计算资源的要求也较高。同时,深度学习模型通常被视为黑盒模型,其内部决策过程难以解释,对于细胞定位任务而言可能是一个挑战,因为解释模型如何定位细胞对于细胞学研究和医学诊断等领域非常重要。ADPC 方法基于深度学习方法在细胞分割领域较为成熟的优势,并利用改进的密度峰值聚类算法通过数据特点自适应推导参数,模型参数量较小,尤其适用于黏附细胞程度较深的图像。

5 结语

自动、精准定位神经元中心是对细胞结构进行定量分析的基本任务,也对理解脑功能和疾病原理具有重要意义。为此,本文对基于密度峰值聚类的细胞定位方法进行了自适应改进,首先通过局部密度与较高密度点的最小距离趋势性变化,自动优化密度阈值与距离阈值参数,其次使用 VoxResNet 进行细胞预分割处理。

实验表明,本文方法充分发挥了传统图像处理方法和深度学习在细胞分割领域的优点,既能准确定位细胞,又保持了相对较低的计算资源和数据规模,具有较好的聚类效果、强大的鲁棒性及自适应优势,特别在黏连细胞程度较深的数据集上相较于比较方法优势更明显。

未来,将进一步发掘图形图像处理算法和深度学习方法在细胞定位领域的巨大潜力和应用前景,以期为生命科学提供便利。

参考文献:

- [1] CHIANG A S, LIN C Y, CHUANG C C, et al. Three-dimensional reconstruction of brain-wide wiring networks in drosophila at single-cell resolution[J]. Current Biology, 2011, 21(1): 1-11.
- [2] PENG J, LONG B, YUAN J, et al. A quantitative analysis of the distribution of CRH neurons in whole mouse brain[J]. Frontiers in Neuroanatomy, 2017, 11: 63.
- [3] ATTILI S M, SILVA M F M, NGUYEN T V, et al. Cell numbers, distribution, shape, and regional variation throughout the murine hippocampal formation from the adult brain Allen reference Atlas[J]. Brain Structure and Function, 2019, 224(8): 2883-2897.
- [4] PENG H, RUAN Z, ATASOY D, et al. Automatic reconstruction of 3D

- neuron structures using a graph-augmented deformable model[J]. *Bioinformatics*, 2010, 26(12): 38–46.
- [5] ZHONG Q, LI A, JIN R, et al. High-definition imaging using line-illumination modulation microscopy [J]. *Nature Methods*, 2021, 18(3): 309–315.
- [6] GAO X F. Optimization of data acquisition and control software for high definition fluorescence microscopic optical slicing tomography system[D]. Wuhan: Huazhong University of Science and Technology, 2021.
高雪飞. 高清荧光显微光学切片断层成像系统数据采集控制软件优化[D]. 武汉: 华中科技大学, 2021.
- [7] ROYSAM B, SHAIN W, ASCOLI G A. The central role of neuroinformatics in the national academy of engineering's grandest challenge: reverse engineer the brain[J]. *Neuroinformatics*, 2009, 7(1): 1–5.
- [8] LI S, QUAN T, ZHOU H, et al. Identifying weak signals in inhomogeneous neuronal images for large-scale tracing of sparsely distributed neurites[J]. *Neuroinformatics*, 2019, 17: 497–514.
- [9] VINCENT L, SOILLE P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1991, 13(6): 583–598.
- [10] XIAO Y, LI P W. Application of improved watershed model based on gradient correction in cell segmentation [J]. *Biomedical Engineering Research*, 2020, 39(4): 330–336.
肖月, 李朋伟. 基于梯度修正的改进分水岭模型在细胞分割中的应用 [J]. *生物医学工程研究*, 2020, 39(4): 330–336.
- [11] XU H Z, ZHANG J J, LIU D, et al. Optimization algorithm for crown segmentation of CV model based on watershed [J]. *Foreign Electronic Measurement Technology*, 2023, 42(5): 56–61.
徐鸿哲, 张建杰, 刘丹, 等. 基于分水岭线的 CV 模型树冠分割优化算法 [J]. *国外电子测量技术*, 2023, 42(5): 56–61.
- [12] WINTER M, MANKOWSKI W, WAIT E, et al. Separating touching cells using pixel replicated elliptical shape models [J]. *IEEE Transactions on Medical Imaging*, 2019, 38(4): 883–893.
- [13] ZHOU H, LI S, LI A, et al. GTree: an open-source tool for dense reconstruction of brain-wide neuronal population [J]. *Neuroinformatics*, 2021, 19(2): 305–317.
- [14] WANG J, SUN Z Y, GUO P, et al. Improved leukocyte detection algorithm of YOLOv5 [J]. *Computer Engineering and Applications*, 2022, 58(4): 134–142.
王静, 孙紫雲, 郭苹, 等. 改进 YOLOv5 的白细胞检测算法 [J]. *计算机工程与应用*, 2022, 58(4): 134–142.
- [15] WANG T. Overlapping cell segmentation with depth information [C]// 8th International Conference on Advanced Machine Learning and Technologies and Applications, 2022: 278–287.
- [16] YAO Z Y. Research on the application of deep learning in cell image segmentation [D]. Beijing: Beijing University of Posts and Telecommunications, 2021.
药泽一. 深度学习在细胞图像分割中的应用研究 [D]. 北京: 北京邮电大学, 2021.
- [17] VAN H G, MOSQUERA C, NÁPOLES G. A review on the long short-term memory model [J]. *Artificial Intelligence Review*, 2020, 53(8): 5929–5955.
- [18] WANG Y. Research on cancer classification model based on deep learning algorithm [D]. Wuhan: Huazhong University of Science and Technology, 2021.
王毅. 基于深度学习算法的癌症分类模型研究 [D]. 武汉: 华中科技大学, 2021.
- [19] STRINGER C, WANG T, MICHAELSON M, et al. Cellpose: a generalist algorithm for cellular segmentation [J]. *Nature Methods*, 2021, 18(1): 100–106.
- [20] CHENG S H, QUAN T W, LIU X M, et al. Large-scale localization of touching somas from 3D images using density-peak clustering [J]. *BMC Bioinformatics*, 2016, 17(1): 375.
- [21] ZHU S. Research on image segmentation methods based on density peak clustering [D]. Chongqing: Chongqing University of Posts and Telecommunication, 2019.
朱蒹. 基于密度峰值聚类的图像分割方法研究 [D]. 重庆: 重庆邮电大学, 2019.
- [22] MA S H, YOU H R, TANG L, et al. An adaptive density peak clustering algorithm [J]. *Journal of Northeastern University*, 2022, 43(6): 761–768.
马淑华, 尤海荣, 唐亮, 等. 一种自适应的密度峰值聚类算法 [J]. *东北大学学报*, 2022, 43(6): 761–768.
- [23] CHEN H, DOU Q, YU L, et al. VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images [J]. *Neuroimage*, 2018, 170: 446–455.
- [24] HU T, XU X, CHEN S, et al. Accurate neuronal soma segmentation using 3D multi-task learning U-Shaped fully convolutional neural networks [J]. *Frontiers in Neuroanatomy*, 2020, 14: 592806.
- [25] ZHENG T, YANG Z, LI A, et al. Visualization of brain circuits using two-photon fluorescence micro-optical sectioning tomography [J]. *Opt Express*, 2013, 21(8): 9839–9850.
- [26] PENG H, XIE P, LIU L, et al. Morphological diversity of single neurons in molecularly defined cell types [J]. *Nature*, 2021, 598: 174–181.
- [27] AL-KOFAHI Y, LASSOUED W, LEE W, et al. Improved automatic detection and segmentation of cell nuclei in histopathology images [J]. *IEEE Transactions on Biomedical Engineering*, 2010, 57(4): 841–852.
- [28] QUAN T, ZHENG T, YANG Z, et al. NeuroGPS: automated localization of neurons for brain circuits using L1 minimization model [J]. *Science Report*, 2013, 3: 1414.

(责任编辑: 刘嘉文)