An Algorithm to Recognize Yeast Budding Events Based on Convolutional Neural Network*

XIAO Qin[1], ZHANG Yuxin[2], WANG Yingying[2], LIU Ke[2], ZHU Zhen[1,2*]
(1.School of Microelectronics, Southeast University, Nanjing Jiangsu 210096, China;
2.School of Electronic Science and Engineering, Southeast University, Nanjing Jiangsu 210096, China)

Abstract: In yeast aging studies with budding yeast (Saccharomyces cerevisiae), microfluidic chips are used for yeast cell immobilization and cell culture on single cell scale. Based on time-lapse imaging, replicative lifespan (RLS) can be determined through image analysis. However, manual image analysis is time-consuming and laborious. Here, an algorithm based on convolutional neural network is proposed to extract the budding events of single yeast cell, which is tailored for microscopic images of cells on microfluidic chips with "leaky bowl" traps. The effectiveness of the algorithm is proved by high accuracy of 95% and F1 score of 90% on average. Therefore, the proposed algorithm provides a novel method for single-cell analysis, and it could be used for similar work in the field of biomedicine image analysis.

Key words: Saccharomyces cerevisiae; replicative aging; replicative lifespan; neural network
EEACC: 6135 **doi**: 10.3969/j.issn.1005-9490.2023.04.035

基于卷积神经网络的酵母出芽特征识别方法*

肖 秦[1], 张雨昕[2], 王颖瀛[2], 刘 可[2], 朱 真[1,2*]
(1.东南大学微电子学院, 江苏 南京 210096; 2.东南大学电子科学与工程学院, 江苏 南京 210096)

摘 要: 在酿酒酵母(Saccharomyces cerevisiae)的细胞衰老研究中, 微流控芯片可以实现单个酵母细胞的捕获和培养, 同时用光学显微镜进行时序图像拍摄, 最后分析图像得到细胞的复制寿命。针对基于"漏碗"式捕获结构的微流控芯片长时序酵母细胞实验图像, 为提高处理效率, 提出了一种基于卷积神经网络的高效算法, 用于提取酵母细胞在衰老过程中的出芽特征。经验证, 提出的算法模型准确率达到 95%, F1 score 指标达到平均 90%以上, 具有较高的可靠性。为深度学习方法在生物医学图像处理领域的应用提供了新的思路。

关键词: 酿酒酵母; 复制衰老; 复制寿命; 神经网络
中图分类号: Q-334; TP751 **文献标识码**: A **文章编号**: 1005-9490(2023)04-1096-08

随着世界人口老龄化趋势的日益严重, 寿命研究在现代科学中的重要性也越来越大。人类平均寿命的不断增长, 导致与衰老相关的各种疾病更加广泛地出现在大众视野中, 如心血管疾病[1], 糖尿病[2]等, 这些疾病均被证实与人类的年龄相关。然而, 人类等哺乳动物的生命系统过于复杂, 研究人员难以直接对人类个体进行寿命相关的实验研究。因此, 研究人员常常使用酵母[3]、线虫[4]、果蝇[5]、小鼠[6]等模式细胞或模式动物进行衰老寿命研究。其中, 酿酒酵母(Saccharomyces cerevisiae), 也叫出芽酵母, 凭借其繁殖快、寿命短、全基因测序且易于实验室维护的特点[7], 常常被用做寿命研究的模型。

传统的酵母细胞复制寿命测定实验中, 研究人员通常采用显微解剖法[8], 而这种方法实验周期长, 效率低下。近年来, 随着微纳制造技术的发展, 越来越多的微流控芯片被设计出来用于酵母细胞的寿命研究[9-11]。这些芯片有着高通量的优点, 随之也带来了一个巨大的挑战, 也就是大量图像数据的自动化处理。

基于微流控芯片的酵母细胞复制寿命测量主要依赖于光学显微时序成像法。针对实验中获取的图像数据, 现有的生物领域的图像处理软件, 如 ImageJ[12]、YeastQuant[13]等常使用拟合、形态学分析等传统计算机视觉方法对细胞轮廓、面积等进行自动化分析, 这些方法大多需要人为设计算法步骤。而

项目来源: 国家重点研发计划(2021YFF0701002); 国家自然科学基金项目(61774036); 江苏省研究生科研与实践创新计划项目(SJCX20_0034); 中央高校基本科研业务费专项资金(3206002107D)
收稿日期: 2022-05-10 **修改日期**: 2022-05-18

近年来,随着深度学习的发展,更多的神经网络应用到了细胞分析领域,如使用 GAN 提高显微镜拍摄下的细胞图像分辨率[14];使用语义分割对酵母细胞边缘进行提取分割[15]。这些深度学习算法不需要人为设计过多的特征提取步骤,极大提高了效率。无论是传统计算机视觉方法,还是深度学习的方法,现有研究大多针对酵母细胞的轮廓与面积,这些算法或软件无法在时序上对细胞进行跟踪式的出芽特征分析,研究人员需要一种算法,将微流控技术与生物学领域结合,从而实现对酵母细胞出芽特征的快速以及准确分析。本文基于此类问题,提出了一种基于卷积神经网络的酵母出芽特征提取算法。

1　酵母细胞的出芽特征提取

1.1　酵母细胞的复制寿命与微流控芯片

酵母细胞的分裂繁殖是一个不对称的繁殖过程,母细胞复制后会产生一个尺寸更加小的细胞,也就是会产生一个新的芽,复制寿命[16](Replicative Lifespan,RLS)定义为单个酵母细胞在停止分裂前所产生的芽的个数,如图 1 所示。复制寿命常常作为指标用来研究基因和分子机制,为真核生物的衰老机制提供参考。

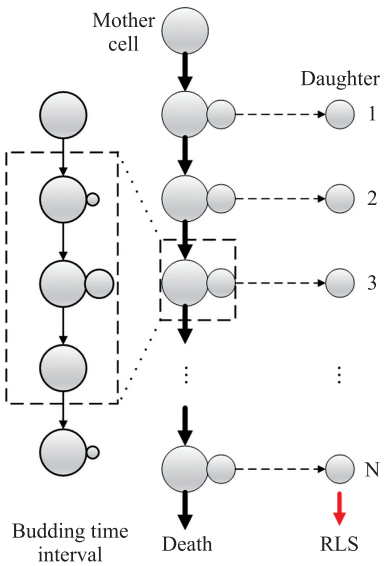
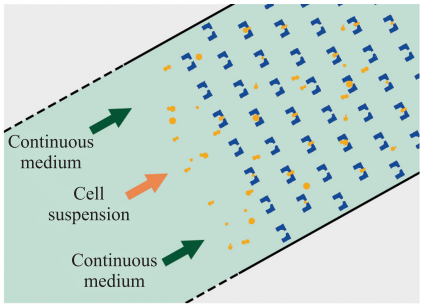


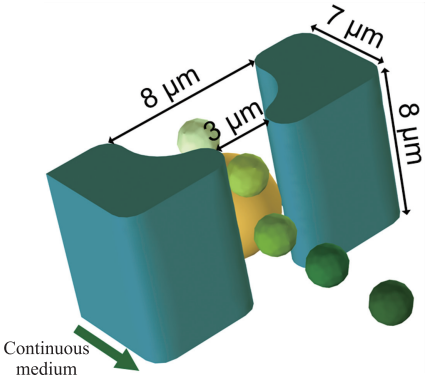
图 1　酵母细胞的复制寿命

随着微流控技术的不断发展,高通量微流控芯片广泛地应用到了酵母衰老研究中。微流控芯片利用微结构将酵母细胞捕获并固定在结构中央,流道中注入培养液对酵母细胞进行培养。

如图 2(a)所示,高通量微流控芯片上含有多个捕获结构,每个捕获结构由对称分布的聚二甲基硅氧烷(Polydimethylsiloxane,PDMS)微柱组成。如图 2(b)所示,微柱结构高度为 8 μm,形成了“漏碗”式结构,流体从 8 μm 的宽口上游注入,从 3 μm 窄口流出,可将酵母细胞稳定固定在捕获结构中心进行长时间培养与监测。在流体的作用下,酵母细胞产生的芽会被流体冲刷,实现芽的剪切。



(a) 微流控芯片捕获阵列示意图

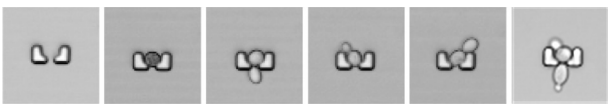


(b) “漏碗”式捕获结构示意图

图 2　用于酵母复制寿命监测的高通量微流控芯片

为实现酿酒酵母细胞复制衰老过程的自动监测和数据分析,本研究采用激光共聚焦显微镜系统对酵母细胞进行长时间序列成像。实验开始时,由注射泵控制气密性针管通入细胞悬浮液,待微流控芯片中的大部分微结构完成细胞捕获后,将通入流体切换至纯净的细胞培养液,并以 10 μL/min 的恒定流速对已捕获的酿酒酵母细胞实现稳定的长时间培养。整个实验过程需 50 h 左右,并由计算机工作站自动化控制,每隔 10 min 使用 20 倍物镜对捕获微结构区进行一次明场成像,记录细胞在此期间的生长、分裂、衰老等过程。

拍摄得到的酵母细胞如图 3 所示。其中图 3(a)中捕获结构未捕获到母细胞;图 3(b)中捕获的母细胞已死亡;在图 3(c)(d)(e)中,母细胞在各方向上出芽;图 3(f)为母细胞在两个方向上均有出芽的现象。显微镜获取到的图像均为灰度图像。



(a)　(b)　(c)　(d)　(e)　(f)

图 3　捕获阵列上的酵母细胞图像

1.2 图像的标注

本文采用卷积神经网络对酵母细胞的出芽特征进行学习,由于微流控芯片培养酵母细胞的特殊性,本文创新性地提出了一种标注方法,对不同方向的芽进行分类,并搭建了训练集与测试集。

对于酵母的衰老实验,研究人员重点关注母细胞自身以及芽的状态。在时序图像中,对于母细胞,研究人员主要关注母细胞是否被捕获以及是否死亡。当母细胞被捕获时,图像中的捕获结构中央出现卵圆形细胞。随后,在微流控芯片捕获结构的束缚下,母细胞被固定在结构中央直至死亡。母细胞死亡后,图像上表现为捕获结构中央的圆形细胞像素变黑且变密集;对于酵母的子细胞,可以很容易发现,母细胞在捕获结构以及流体的作用下,其出芽方向只有 3 个。本文以流体方向为基准,将捕获结构的窄口出芽方向定义为下游颈部(bottle,B),宽口的两个出芽方向定义为上游左侧(left,L)、上游右侧(right,R)。芽的体积不断增大,一段时间后与母细胞分离。由于显微镜的时序拍摄存在间隔,在这段时间内,可能会发生旧芽掉落、新芽生长,而显微镜并未拍摄到的情况。针对以上情况,本文定义了子细胞的两组特征,即不同的出芽方向以及出芽尺寸。出芽方向定义如图 4 所示,其中图 4(a)(b)为颈部出芽,图 4(c)(d)为上游左侧出芽,图 4(e)(f)为上游右侧出芽;图 4(a)(c)(e)为小芽,图 4(b)(d)(f)为临近被剪切去除的芽。

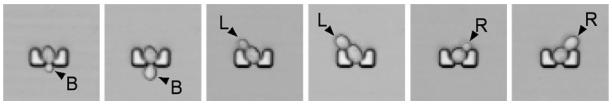


图 4 出芽方向定义

为了提取上述的酵母细胞的图像特征,本文提出数字化编码的图像标注方法,一方面可提高计算效率,另一方面便于后续寿命计算。本文采用二进制方式对图像特征进行数学上的描述,对一张图片,采用多个二进制变量对图像进行特征编码。

本文建立了一个 1×8 的向量对母细胞和芽的特征进行编码。首先对于母细胞,本文采用子向量[STATUS]来描述捕获与存活状态。[STATUS]包含 2 个元素,定义[STATUS]=[NE D],NE(Not Empty)用来表示母细胞是否存在于捕获结构上,D(Dead)用来表示母细胞是否死亡,变量分量均为二值取值(0 或 1)。以母细胞存活状态为例,本文以 1 代表母细胞死亡,以 0 代表母细胞存活;其次,对于母细胞衰老过程中产生的芽,本文用向量[B L R]来编码特征。[B L R]维度为 1×6,分别代表不同方向的出芽特征,[B]、[L]、[R]类似于 STATUS,维度均为 1×2,用于描述相应出芽方向上是否出芽以及芽的大小。以向量[B]为例,该向量含有 2 个元素,以 b1、b2 代表,即[B]=[b1 b2],b1 用于描述是否出芽,b2 用于描述出芽大小。[B]=[1 0]即代表母细胞于下游颈部方向产生一个细胞尺寸较小的芽;[B]=[1 1]则代表母细胞于下游颈部方向较大的芽临近剪切。

酵母细胞的图像标注如图 5 所示。对于 S,也就是前文中的 STATUS 向量,上标用来表示母细胞是否死亡,另一元素用于表示母细胞是否被捕获;对于 L、R、B,上标用来表征芽的尺寸大小,另一元素用来表示在该方向上是否存在芽。基于以上提出的标注方法,整张酵母细胞图像被编码为只含二进制元素的向量,其元素反映了母细胞的捕获与存活特征,以及出芽方向与出芽尺寸特征。

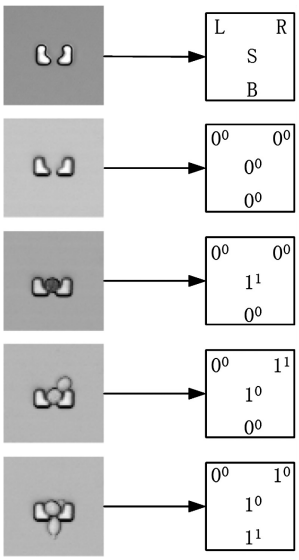


图 5 酵母细胞图像标注

已有的出芽数目简单标记方法[17]仅标注当前图像上出现的子细胞数目,而本文提出的基于酵母细胞状态特征和出芽方向、尺寸特征的标注方法可以对各出芽方向上的芽进行自动化追踪,区分旧芽与新芽的产生与剪切时间。连续两位数编码方式以两个变量记录连续两帧图像之间的出芽特征差异,包括数目变化特征与新旧芽替换特征[18],本质上为出芽数目简单标记的改善方法。该方法需对两帧图像特征进行比较分析,标注时更为繁琐,优点是在分析复制寿命时更为准确。相对于该方法,本文提出的标注方法仅对单张图像上酵母细胞的出芽特征进行标注,其图像标注数据更易于理解,可通过卷积神经网络自动化提取特征,在提高标注效率的同时更有利于后续在不同的出芽方向上进行复制寿命统计。

1.3　图像扩增与数据集搭建

深度学习算法的特征提取步骤本质上是损失函数进行梯度下降运算,在这个过程中,容易出现过拟合,即训练的神经网络对于训练集有着非常高的性能,而在测试集上性能较差的现象,这会使得网络得到一个方差较高的函数。为了解决过拟合,许多正则化方法被应用到网络结构中,使得神经网络能够扩展到小型数据集中。而数据增强,则可以在数据的根源上改善该问题,特别是对于数据量稀少的生物医学领域。对于神经网络,数据集越庞大,网络迭代得到的模型性能越好,但是数据的收集和标注是一个十分艰巨的任务。本论文实验中获得的微流控芯片上酵母细胞图像数据量有限,因此需要采取合适的方法扩增图像数据集,同时减少对于图像质量的影响。

对于酵母的衰老图像,本文中使用的图像扩增主要为几何变换,如图 6 所示,包括旋转,随机平移。由于在显微镜拍摄酵母细胞的时序图像时,载物台会发生一定范围内的偏移,导致捕获结构和细胞无法精确保持在图像中央,因此几何变换能有效改善位置偏差带来的问题。值得注意的是,由于芽的尺寸是标注特征之一,旋转可能会导致图像的失真。为了避免这一问题,本文选择了 90,180,270 三个角度对图像进行旋转。最后,训练集中存在类别失衡问题,也就是某一类别的样本数过多导致算法倾向于该类群体的预测,而对于样本数较少的类别,网络的性能较为低下。为了解决这一问题,本文在数据集中引入了过采样数据集处理方法。考虑到标注中的出芽方向特征,本文对少数类别样本进行图像和标注的同时翻转,随后采取重采样改善数据集中的类别不平衡问题。

得到酵母细胞图像和标注数据后,数据集按照 4:1 的比例划分为训练集与验证集。训练集直接用于网络的梯度下降,验证集用于防止网络过拟合,另外本文也搭建了测试集则用于验证网络的有效性。

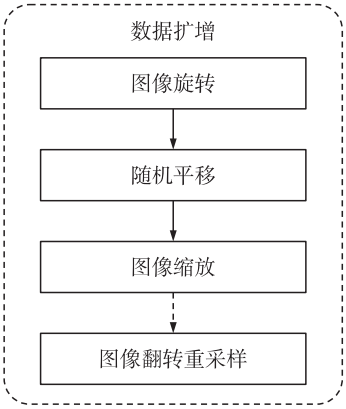


图 6　图像扩增

2　卷积神经网络

2.1　残差结构

深度卷积神经网络主要通过卷积、池化、激活层来实现输入到输出的非线性转换与特征提取。研究人员一般在现有的神经网络中添加线性转换,使得网络在梯度下降时实现非线性转换与线性转换之间的平衡。

为了更加准确地提取酵母细胞的出芽特征,本文采用快捷链接(Short Cut Connection)实现线性与非线性转换之间的平衡。本文搭建的快捷链接基本结构如图 7 所示,包含 2 个卷积层。普通卷积神经网络经过 2 次卷积、池化和激活后会直接输出映射 $F(x)$,而在快捷链接层中,非线性映射 $F(x)$ 会拟合另一个恒等映射,原始映射被改写为 $F(x)+x$。

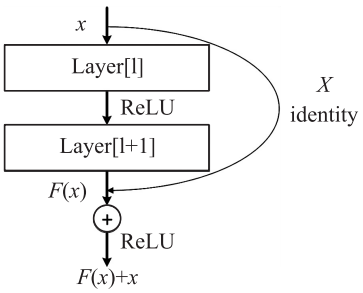


图 7　快捷链接层

将该映射以公式表达,

$$y=F(x,\{W_i\})+x \tag{1}$$

式中:$y=F(x,\{W_i\})$ 为网络迭代的非线性映射,x 为该快捷链接层输入,W_i 为权重参数,y 为该快捷链接层输出,将两层卷积层展开,得到 F 的计算公式。

$$F=W_2\sigma(W_1x) \tag{2}$$

式中:W_1,W_2 分别为两层卷积层的权重参数,σ 为 ReLU(Rectified Linear Unit)激活层。

若 x 与 F 的维度不同,则需要加入线性映射 W_s,得到下述公式。

$$y=F(x,\{W_i\})+W_sx \tag{3}$$

将该类残差映射与原始映射对比,在极端情况下,如果恒等映射是最优的,线性变换会使得网络准确率下降的情况下,网络直接将 F 置为 0 比通过非线性层迭代来拟合恒等映射更容易,此时的快捷链接层就相当于一个恒等层,并且不会增加网络的计算复杂度。

2.2　残差神经网络搭建与训练

基于上文中的残差结构,本文搭建了残差神经网络 ResNet18(图 8),用于提取酵母细胞的出芽特征。

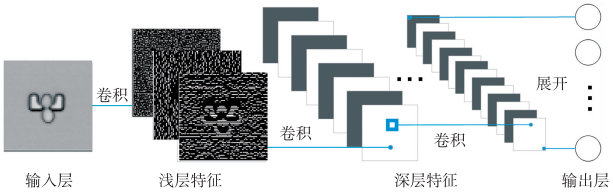


图 8 神经网络示意图

ResNet18 结构如表 1 所示,细胞图像首先经浅层卷积层提取基本特征,进行池化,随后经过 4 个残差网络层提取深层特征。最后网络使用全局池化并将提取到的深层特征展开,使用全连接层进行加权处理,得到每个任务最终的输出。

表 1 残差神经网络结构

Layer name	ResNet18	Number
Conv1	7×7,64,stride 2	1
Pool	3×3 max pool,stride 2	1
Conv2	$\begin{bmatrix}33,64\\33,64\end{bmatrix}$	2
Conv3	$\begin{bmatrix}33,128\\33,128\end{bmatrix}$	2
Conv4	$\begin{bmatrix}33,256\\33,256\end{bmatrix}$	2
Conv5	$\begin{bmatrix}33,512\\33,512\end{bmatrix}$	2
Pool	Average pool	1
FC	8 outputs,regression	1

浅层卷积网络层使用 7×7 卷积核,步长为 2,填充为 3,该浅层卷积层用于提取基本特征,如垂直、水平等边缘特征。每个卷积层后都有一个批标准化层(Batch Normalization,简称 BN 层)和非线性激活层(Activation)。BN 层将各特征的取值范围统一,防止网络某一层输入数据的分布发生变换,从而导致网络需要重新对该层数据分布进行学习,降低学习速度。其具体计算公式如下:

$$\hat{x}_i=\frac{x_i-\mu}{\sqrt{\sigma^2+\varepsilon}} \tag{4}$$

$$y_i=\gamma\hat{x}_i+\beta \tag{5}$$

式中:x_i 为第 i 个样本,μ 为当前批量样本的均值,σ 为标准差,在进行训练时这两个值为当前批量数据的均值与标准差,而在网络进行预测时这两个变量是固定的,通过计算整个训练集的均值与标准差得到;ε 为一个无限接近 0 的常数,防止计算公式中分母为 0;γ 与 β 为待学习的比例因子与偏移量,用于控制输出 y_i 的方差与均值。BN 层将输入的直流偏置置为 0,所以在构建卷积层时不需要再添加偏置。

在网络结构中批标准化层处于卷积层与激活函数之间,该层的使用使得网络在进行迭代时可以更快地适应不同层的数据分布,可以使用更大的学习率进行迭代,训练过程更加稳定,训练速度也得到极大的提高。此外,批标准化层取代了以往使用较多的 dropout 层结构,具有正则化的效果,有效地抑制了网络过拟合。

非线性激活层采用 ReLU 函数,相比于以往使用较多的 Logistic 与 Tanh 函数,该函数的导数较为简单,十分适用于卷积神经网络中的梯度下降。

浅层卷积获得特征并对数据进行非线性转换后,数据维度为 112×112×64,此时的特征维度较大,本文采用最大池化将特征维度大小降为 56×56×64。

随后,本文采用 4 个残差结构 Conv2-Conv5 提取深层特征,当中的每个残差结构均采用 3×3 卷积核,表中的 Number 代表残差结构的数量。本文使用的残差结构快捷链接每次均跳过 2 个卷积层,其中 Conv2 快捷链接使用恒等映射,Conv3-Conv5 中的快捷链接因输入与输出维度不同,需加入线性映射,线性映射中使用的卷积核大小为 1×1,步长为 2。在经过残差网络结构提取深层特征后,本文采用全局平均池化进一步减少特征维度,并展开为向量,此时数据维度为 1×1×512,也就是说,网络此时提取了 512 个有效特征。最后,本文采用全连接层对这 512 个特征进行权重赋值计算,得到最后 8 个输出,并计算网络预测值与真实值之间的均方误差,利用该误差进行反向传播和梯度下降,更新网络参数。

网络输入为 224×224 的灰度图,网络最后一层输出为 1×8,对应前文中提到的标注方法。为了加快网络训练的速度,本文采用 Adam 优化器对网络参数进行梯度下降迭代。由于训练集中的数据量往往过于庞大,若每次网络在整个训练集上更新权重参数,每次前向与后向传播需要大量的计算时间。Adam 优化器对批量数据进行迭代,即每次进行前向与反向传播时,只有部分训练集的数据用于计算。每个周期结束时,训练集中的所有数据均被网络计算过。另一方面,为了寻找最佳拟合点,防止过拟合和欠拟合,本文设置 3 个迭代步骤,3 个不同的学习率,在周期 1~6,学习率设置为 0.003;在周期 7~12,学习率设置为 0.001;在周期 12~18,初始学习率设置为 0.0001,并且每 2 个周期学习率乘以 0.3。

3 网络性能评估

3.1 评估指标

在机器学习的分类问题中,常用到的指标有准确率(Accuracy)、查准率(Precision)、召回率

(Recall)和 F1 score。在得到上述指标之前,需要首先计算 TP(True Positive)、FN(False Negative)、FP(False Positive)、TN(True Negative)组成的混淆矩阵(Confusion matrix),具体定义如表 2 所示。举例来说,若一个样本实际为正样本,网络预测其也为正样本,那么该样本计入 TP。

表 2　混淆矩阵

	预测为正	预测为负
真实为正	TP	FN
真实为负	FP	TN

准确率定义为网络正确分类的样本占总样本数的比例,计算公式如下:

$$\text{Accuracy}=\frac{\text{TP}+\text{FN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \tag{6}$$

查准率定义为网络预测的正样本中真实为正样本的比例,计算公式如下:

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{7}$$

召回率定义为在真实为正的样本中被网络预测为正的比例,计算公式如下:

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{8}$$

准确率作为最容易理解的指标,当样本类别失衡时,无法合理反应网络的性能,因此引入查准率与召回率,前者用于衡量网络是否有误判,而后者用于衡量网络的预测是否有遗漏,二者互相矛盾,单一指标并不能完全反应网络的性能。鉴于以上情况,本文采用二者的加权平均评估网络性能,也就是 F1 score,其计算公式如下:

$$\text{F1}_{\text{score}}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{9}$$

3.2　结果分析

本文对 20 000 张训练用的图像以及 4 000 张测试用的图像进行性能测试评估,得到的结果如表 3~表 5 所示。

从表 3 可以看出,网络在训练集和测试集上的平均准确率都达到了 95%以上。而对于训练集和测试集的 F1 score,每项任务的评估结果均在 90%以上,可以证明本文训练好的网络能有效提取酵母细胞的出芽特征。从表 4、表 5 可知,网络对于捕获状态和酵母细胞存活的识别准确率较高,而对于尺寸的识别程度低于对各方向上出芽判断的识别程度,这是因为在人工标注数据时,标注人员很难一直以一个确定的标准来衡量尺寸大小;另外,对比不同方向上的出芽判断和尺寸判断结果,可得知网络在识别颈部方向上的芽时性能要高于其他两个方向,这是因为在捕获结构和流体的作用下,酵母细胞更倾向在窄口出芽,导致该方向出芽样本较多。以上问题均来自数据集的不合理,未来可通过数据集的改善优化上述问题。

表 3　网络预测准确率　　单位:%

Task		Training set	Test set
STATUS	NE	100	100
	D	99.97	99.73
B	b1	99.64	99.38
	b2	98.28	96.91
L	l1	99.47	99.10
	l2	99.20	98.80
R	r1	99.60	99.08
	r2	99.39	98.93

表 4　网络在训练集上的性能　　单位:%

Task		Precision	Recall	F1 score
STATUS	NE	100	100	100
	D	99.92	99.95	99.93
B	b1	99.62	99.49	99.56
	b2	96.92	96.36	96.64
L	l1	99.25	97.35	98.29
	l2	95.14	95.86	95.50
R	r1	99.38	98.06	98.72
	r2	96.63	97.58	97.10

表 5　网络在测试集上的性能　　单位:%

Task		Precision	Recall	F1 score
STATUS	NE	100	100	100
	D	100	98.66	99.32
B	b1	98.90	99.50	99.23
	b2	94.02	94.87	94.44
L	l1	97.09	97.39	97.24
	l2	92.15	95.14	93.62
R	r1	95.68	96.67	96.17
	r2	90.94	94.43	92.65

4　酵母细胞复制寿命提取

网络输出后,整个酵母细胞衰老过程的时序图像被编码为二维矩阵,矩阵由定义的特征组成。

对于该特征矩阵,可从各出芽方向上进行数据分析,其计算流程如图 9 所示。以单个酵母细胞的时序衰老图像(图 10)为例进行复制寿命提取,首先,需要对母细胞的特征进行分析,也就是 NE 与 D 变量,获得酵母细胞被捕获以及死亡的时间(图 11)。随后,在母细胞被捕获并且存活的时间段内对三个方向的芽进行统计。此时确定一个芽需要对两个变量同时分析。以 B 方向为例,确定 B 方向上某一时间段内

存在一个芽,需要在该时间段内对 b1 与 b2 曲线进行分析,确认两个变量的曲线脉冲有重合区域,以此描述一个完整的出芽过程。

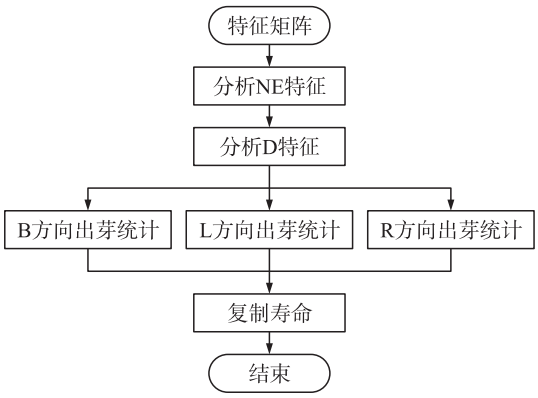


图 9 复制寿命提取算法流程图

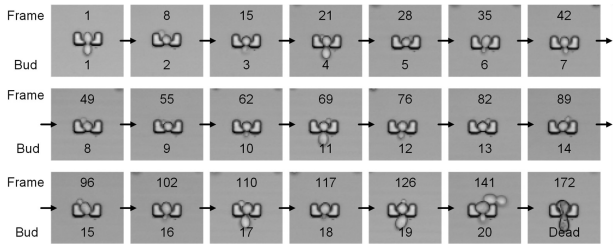


图 10 单个酵母细胞的时序显微图像

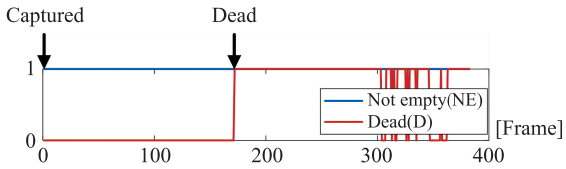


图 11 获取酵母细胞的捕获与死亡时间

采用以上方法对各方向上的芽进行统计后,对三个方向上的出芽进行整合统计(图 12),可知该酵母细胞的复制寿命为 20,即一共产生了 20 个芽,与原时序图像中的出芽数目相同,验证了本文提出的出芽特征提取方法的有效性。

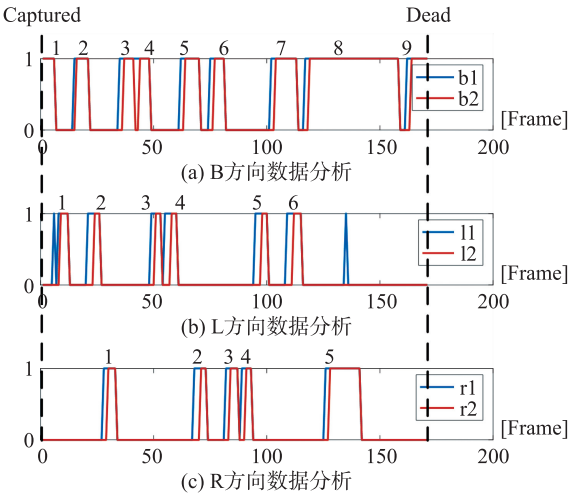


图 12 对不同方向上出芽数目进行分析

5 结论及意义

本文针对基于微流控芯片的酵母细胞衰老研究,提出了酵母细胞衰老过程中的出芽特征识别方法,结合深度学习神经网络算法,设计了新颖的编码方法,并通过残差结构和多标签分类方法构建了用于出芽特征检测的网络模型。本文提出的神经网络模型可以从实验显微图像提取酵母细胞捕获与存活状态、出芽方向与出芽尺寸特征。本文提出的模型评估指标较高,验证了该网络的特征识别具有较高的可信度。

此外,本文在提出的出芽特征识别算法基础上,从各出芽方向上统计芽的数目,进一步处理特征向量可最终获取酵母细胞的复制寿命。本文创新性地将卷积神经网络残差结构应用于微流控芯片-单细胞图像,有效提取了单细胞出芽特征,推进了酵母细胞衰老研究。本文也为生物医学领域的图像数据处理提供一种新的思路,研究人员可根据芯片结构的不同,更改待提取特征,从而满足不同实验的数据处理需求。

参考文献:

[1] Ajoolabady A, Aslkhodapasandhokmabad H, Aghanejad A, et al. Mitophagy Receptors and Mediators: Therapeutic Targets in the Management of Cardiovascular Ageing[J]. Ageing Research Reviews, 2020, 62: 101129.

[2] Munshi M N, Meneilly G S, Rodríguez-Mañas L, et al. Diabetes in Ageing: Pathways for Developing the Evidence Base for Clinical Guidance[J]. The Lancet Diabetes & Endocrinology, 2020, 8(10): 855-867.

[3] Botstein D, Fink G R. Yeast: An Experimental Organism for 21st Century Biology[J]. Genetics, 2011, 189(3): 695-704.

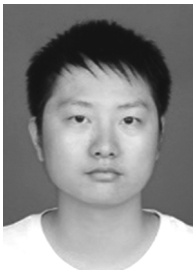
[4] Zhu Z, Chen W, Tian B, et al. Using Microfluidic Impedance Cytometry to Measure C. Elegans Worms and Identify Their Developmental Stages[J]. Sensors and Actuators B: Chemical, 2018, 275: 470-482.

[5] Yamauchi T, Oi A, Kosakamoto H, et al. Gut Bacterial Species Distinctively Impact Host Purine Metabolites During Aging in Drosophila[J]. iScience, 2020, 23(9): 101477.

[6] Wang S, Lai X, Deng Y, et al. Correlation Between Mouse Age and Human Age in Anti-Tumor Research: Significance and Method Establishment[J]. Life Sciences, 2020, 242: 117242.

[7] Jo M C, Qin L. Microfluidic Platforms for Yeast-Based Aging Studies[J]. Small, 2016, 12(42): 5787-5801.

[8] Mortimer R K,Johnson J R. Life Span of Individual Yeast Cells[J]. Nature,1959,183(4677):1751-1752.
[9] Crane M M,Clark I B,Bakker E,et al. A Microfluidic System for Studying Ageing and Dynamic Single-Cell Responses in Budding Yeast[J]. PLoS One,2014,9(6):e100042.
[10] Liu P,Young T Z,Acar M. Yeast Replicator:A High-Throughput Multiplexed Microfluidics Platform for Automated Measurements of Single-Cell Aging[J]. Cell Reports,2015,13(3):634-644.
[11] Xie Z,Zhang Y,Zou K,et al. Molecular Phenotyping of Aging in Single Yeast Cells Using a Novel Microfluidic Device[J]. Aging Cell,2012,11(4):599-606.
[12] Koschwanez J,Holl M,Marquardt B,et al. Identification of Budding Yeast Using a Fiber-Optic Imaging Bundle[J]. Review of Scientific Instruments,2004,75(5):1363-1365.
[13] Schindelin J,Rueden C T,Hiner M C,et al. The ImageJ Ecosystem:An Open Platform for Biomedical Image Analysis[J]. Molecular Reproduction and Development,2015,82(7-8):518-529.
[14] Pelet S,Dechant R,Lee S S,et al. An Integrated Image Analysis Platform to Quantify Signal Transduction in Single Cells[J]. Integrative Biology,2012,4(10):1274-1282.
[15] Zhang H,Fang C,Xie X,et al. High-Throughput,High-Resolution Deep Learning Microscopy Based on Registration-Free Generative Adversarial Network[J]. Biomedical Optics Express,2019,10(3):1044-1063.
[16] Aydin A S,Dubey A,Dovrat D,et al. CNN Based Yeast Cell Segmentation in Multi-Modal Fluorescent Microscopy Data[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu,USA,IEEE,2017:753-759.
[17] Ghafari M,Clark J,Guo H B,et al. Complementary Performances of Convolutional and Capsule Neural Networks on Classifying Microfluidic Images of Dividing Yeast Cells[J]. PLoS One,2021,16(3):e0246988.
[18] Wang Y,Zhu Z,Liu K,et al. A High-Throughput Microfluidic Diploid Yeast Long-Term Culturing(DYLC)Chip Capable of bud Reorientation and Concerted Daughter Dissection for Replicative Lifespan Determination[J]. Journal of Nanobiotechnology,2022,20(1):171.



肖　秦(1998—),男,汉族,籍贯湖南,东南大学微电子学院硕士在读,xiaoqin98@seu.edu.cn;

张雨昕(2002—),女,汉族,籍贯辽宁,东南大学电子科学与工程学院电子科学与技术专业本科生在读,213202332@seu.edu.cn;

王颖瀛(1994—),女,汉族,籍贯江苏,东南大学电子科学与工程学院微电子学与固体电子学博士在读,yy_wang1994@seu.edu.cn;

刘　可(1999—),女,汉族,籍贯山西,东南大学电子科学与工程学院微电子学与固体电子学硕士在读,liuke@seu.edu.cn;

朱　真(1984—),男,汉族,东南大学电子科学与工程学院、微电子学院,MEMS教育部重点实验室,教授,博士学位,主要研究方向:集成微流控、生物传感、电子学等工程技术方法进行生物学、精准医学领域的单细胞、组织器官、模式动物分析研究,zhuzhen@seu.edu.cn。