



(12) 发明专利申请

(10) 申请公布号 CN 113012167 A

(43) 申请公布日 2021.06.22

(21) 申请号 202110313933.6

(22) 申请日 2021.03.24

(71) 申请人 哈尔滨理工大学

地址 150080 黑龙江省哈尔滨市南岗区52号

(72) 发明人 何勇军 秦健 盖晋平

(51) Int.Cl.

G06T 7/11 (2017.01)

G06N 3/08 (2006.01)

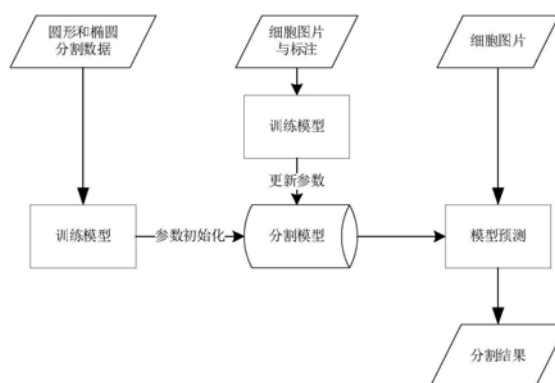
权利要求书2页 说明书5页 附图2页

(54) 发明名称

一种细胞核与细胞质的联合分割方法

(57) 摘要

一种细胞核与细胞质的联合分割方法,本发明涉及细胞病理图片分析诊断技术中,细胞核与细胞质难以分割的问题。提取细胞核与细胞质的形态、纹理和外观特征,为异常细胞的分类和检测提供依据,是细胞病理图像分析诊断的关键工作之一。细胞核的精确分割是提取细胞特征的关键所在。但是由于细胞间存在重叠,细胞形态差异大,细胞质边界对比度较差等,使得目前仍然没有一个较好的分割方法。针对这一问题,结合任务的特点,设计了模型和损失函数,并利用自监督学习对模型参数进行初始化,以引入先验知识。实验表明,这种方法能够有效地实现细胞核与细胞质在细胞病理分析过程中的精确分割。该方法主要应用于细胞分析任务中的细胞核和细胞质的分割任务。



1. 一种细胞核与细胞质的联合分割方法,其特征在于,包含如下步骤:

S1. 采集细胞图片用于标记和数据增强;

S2. 构建改进的UNet模型用于图像分割;

S3. 用自监督学习方法初始化模型参数,引入细胞分割的先验知识;

S4. 计算细胞核与细胞质分割结果之间的相关系数;

S5. 计算模型总体损失函数并更新模型参数;

S6. 采用训练后的模型对图像进行分割。

2. 如权利要求1所述一种细胞核与细胞质的联合分割方法,其特征在于,步骤S1中采集细胞图片用于标记和数据增强,具体为:

通过自动扫描仪获取到细胞图片,将图片裁剪为宽高都为M的小图片,由专业人员利用标注软件标记细胞核与细胞质的轮廓,并生成对应的掩膜图片;对标记数据进行数据增强,数据增强的具体方法包括以下子步骤:

S2-1. 对图像颜色的对比度、饱和度、颜色进行变换;

S2-2. 对图片进行随机区域裁剪,并且随机旋转;

S2-3. 调整范围为原始图像大小的0.1到2.0,如果图像小于原始图像大小,则将用灰色填充该图像。

3. 如权利要求1所述一种细胞核与细胞质的联合分割方法,其特征在于,步骤S2中构建改进的UNet模型用于图像分割具体为:

改进后的UNet模型包含一个下采样模块与两个上采样模块;下采样模块与两个上采样模块之间存在着跳跃连接,跳跃连接可以减少采样过程带来的空间信息损失,使得上采样恢复的特征图中包含更多的语义信息;图片输入到模型后,首先通过下采样模块对图片进行特征提取,之后通过两个上采样模块恢复特征图,最后输出细胞核与细胞质的分割结果;这种模型设计使特征提取模块能够同时关注图像中细胞核和胞质的特征,而两个上采样模块可以各自集中精力处理细胞核和细胞质的分割任务,从而提高分割效果。

4. 如权利要求1所述一种细胞核与细胞质的联合分割方法,其特征在于,步骤S3用自监督学习方法初始化模型参数,引入细胞分割的先验知识具体为:

采用自监督学习方法,可将已有的先验知识引入到模型中,提高模型的收敛速度和识别效果;首先利用软件制作大量人造圆形或椭圆形细胞核图像,并在细胞核外添加细胞质,生成训练数据集;然后使用生成的数据训练改进的UNet模型,将训练后的模型参数作为模型的初始参数,使模型能提前获得胞核形状呈现为圆形或椭圆形的特点。

5. 如权利要求1所述一种细胞核与细胞质的联合分割方法,其特征在于,步骤S4计算细胞核与细胞质分割结果之间的相关系数具体为:

人体细胞由细胞核和细胞质构成,细胞核被细胞质包围,可以通过公式1计算细胞质与细胞核的相关系数来反映这种关系;

$$S = \frac{2|\bar{X} \cap \bar{Y}|}{|\bar{X}| + |\bar{Y}|} + \exp\left(\frac{\|f_n - f_c\|_2}{2\sigma^2}\right) - 1 \quad (1)$$

S表示细胞核和细胞质的关系系数, \bar{X} 和 \bar{Y} 分别表示细胞核和细胞质的分割结果, f_n 和 f_c 分别表示分割结果在细胞核的外轮廓集和细胞质的内轮廓集, σ 表示轮廓的数量,通过 \bar{X}

和 \bar{Y} 的交并比判断两者在空间上是否相交, f_n 和 f_c 距离可以反映出细胞核是否被细胞质所包围。

6. 如权利要求1所述一种细胞核与细胞质的联合分割方法,其特征在于,步骤S5计算模型总体损失函数并更新模型参数具体为:

在细胞分割任务中,存在大量体积较小的细胞核,我们采用Generalized Dice损失来提高小目标的分割效果,具体公式如公式2所示:

$$GDL = 1 - 2 \frac{\sum_{l=1}^2 w_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^2 w_l \sum_n r_{ln} + p_{ln}} \quad (2)$$

其中 w_l 为每个类别的权重, r_{ln} 为类别1在第n个像素值的正确类别, p_{ln} 为模型将像素类别预测为1的概率值,其中 w_l 的计算如公式3所示:

$$w_l = \frac{1}{\left(\sum_{n=1}^N r_{ln} \right)^2} \quad (3)$$

模型的总体损失如公式4所示:

$$\text{Loss} = \text{Loss}_n + \text{Loss}_c + \alpha \cdot S \quad (4)$$

其中 Loss_n 和 Loss_c 分别代表了模型预测细胞核和细胞质时的Generalized Dice损失, S 表示细胞质和细胞核的相关系数, α 为用于调整相关系数影响的超参;最终将新提出的模型整体损失函数加入到模型训练过程中,循环训练不断更新模型参数以达到最佳效果。

7. 如权利要求1所述一种细胞核与细胞质的联合分割方法,其特征在于,步骤S6采用训练后的模型对图像进行分割具体为:

将待分割图片输入到模型中后,模型会同时输出细胞核和细胞质的分割结果,之后按公式5对两个输出进行处理,得到细胞图片分割结果;

$$I(i, j) = \begin{cases} a & \text{if } P_n(i, j) > P_c(i, j) > a \\ b & \text{if } P_c(i, j) > P_n(i, j) > a \\ c & \text{otherwise} \end{cases} \quad (5)$$

其中 $I(i, j)$ 表示细胞分割结果图片 (i, j) 位置的像素值, a, b, c 分别表示分割结果图中细胞核、细胞质和背景的像素值, $P_n(i, j)$ 和 $P_c(i, j)$ 分别表示细胞核和细胞质分割结果在 (i, j) 位置像素正确分类的概率, a 表示设定的阈值。

一种细胞核与细胞质的联合分割方法

技术领域

[0001] 本发明应用于细胞病理图片分析技术中细胞核与细胞质分割问题。

背景技术

[0002] 近年来,随着经济和社会的发展,由于饮食习惯、社会压力、环境污染、作息不规律等因素,癌症在全球范围内大规模爆发,已经成为危害人们生命的严重问题。细胞病理诊断通过采集脱落细胞进行检查,取材简单,应用广泛,又能做出定性诊断,特别适合早期诊断和筛查,值得大面积推广。传统病理诊断完全依赖于病理医生“手动操作、肉眼观察”的人工阅片,存在两大痛点:(1)病理医生对癌症的诊断准确率普遍较低,存在大量的误诊和漏诊;(2)病理诊断的工作量大,且病理医生严重缺乏,无法大面积推广筛查。辅助阅片技术就显得尤为重要,其关键任务之一是提取细胞核与细胞质的形状、纹理和外观特征,以帮助系统有效地实现对异常细胞的分类和检测。

[0003] 细胞的精确分割可以用于计算细胞核与细胞质的形态、纹理及外观特征,这些特征能够有效帮助异常细胞的分类和检测。在细胞图像处理中,由于存在着染色差异、光照不均和垃圾杂质等影响,细胞间相互粘连、重叠,给细胞的精细分割带来困难。通过引入细胞的先验知识到模型中,比如细胞核椭圆度、细胞核与细胞质的位置关系,能够有效提高细胞分割效果。本文提出一种改进后的UNet 模型,将原本U型结构改造作为W型,模型拥有一个下采样模块和两个上采样模块,下采样对于图片进行特征提取,两个上采样模块则分别对于细胞核与细胞质的分割进行解码和分割。首先利用自监督学习初始化模型参数,使模型能够提前掌握细胞核为圆形或椭圆形的先验知识。最后根据细胞核和细胞质之间的空间关系,设计了一种新的损失函数对模型进行优化。

发明内容

[0004] 本发明的目的是为了解决细胞分析任务中的细胞核和细胞质的分割任务,从而提出一种基于改进后UNet模型的分割方法。

[0005] 上述发明目的主要是通过以下技术方案实现的:

[0006] S1、采集细胞图片用于标记和数据增强;

[0007] 通过自动扫描仪获取到细胞图片,将图片裁剪为宽高都为M的小图片,由专业人员利用标注软件标记细胞核与细胞质的轮廓,并生成对应的掩膜图片;对标记数据进行数据增强,数据增强的具体方法包括以下子步骤:

[0008] S1-1. 对图像颜色的对比度、饱和度、颜色进行变换;

[0009] S1-2. 对图片进行随机区域裁剪,并且随机旋转;

[0010] S1-3. 调整范围为原始图像大小的0.1到2.0,如果图像小于原始图像大小,则将用灰色填充该图像。

[0011] S2、构建改进的UNet模型用于图像分割;

[0012] 改进后的UNet模型包含一个下采样模块与两个上采样模块;下采样模块与两个上

采样模块之间存在着跳跃连接,跳跃连接可以减少采样过程带来的空间信息损失,使得上采样恢复的特征图中包含更多的语义信息;图片输入到模型后,首先通过下采样模块对图片进行特征提取,之后通过两个上采样模块恢复特征图,最后输出细胞核与细胞质的分割结果;这种模型设计使特征提取模块能够同时关注图像中细胞核和胞质的特征,而两个上采样模块可以各自集中精力处理细胞核和细胞质的分割任务,从而提高分割效果。

[0013] S3、用自监督学习方法初始化模型参数,引入细胞分割的先验知识;

[0014] 采用自监督学习方法,可将已有的先验知识引入到模型中,提高模型的收敛速度和识别效果;首先利用软件制作大量人造圆形或椭圆形细胞核图像,并在细胞核外添加细胞质,生成训练数据集;然后使用生成的数据训练改进的UNet模型,将训练后的模型参数作为模型的初始参数,使模型能提前获得胞核形状呈现为圆形或椭圆形的特点。

[0015] S4、计算细胞核与细胞质分割结果之间的相关系数;

[0016] 人体细胞由细胞核和细胞质构成,细胞核被细胞质包围,可以通过公式1计算模型预测的细胞质与细胞核的相关系数来反映这种关系;

$$[0017] \quad S = \frac{2|\bar{X} \cap \bar{Y}|}{|\bar{X}| + |\bar{Y}|} + \exp\left(\frac{\|f_n - f_c\|_2}{2\sigma^2}\right) - 1 \quad (1)$$

[0018] S表示细胞核和细胞质的关系系数, \bar{X} 和 \bar{Y} 分别表示细胞核和细胞质的分割结果, f_n 和 f_c 分别表示分割结果在细胞核的外轮廓集和细胞质的内轮廓集, σ 表示轮廓的数量,通过 \bar{X} 和 \bar{Y} 的交并比判断两者在空间上是否相交, f_n 和 f_c 距离可以反映出细胞核是否被细胞质所包围。

[0019] S5、计算模型总体损失函数并更新模型参数;

[0020] 在细胞分割任务中,存在大量体积较小的细胞核,我们采用Generalized Dice 损失来提高小目标的分割效果,具体公式如公式2所示:

$$[0021] \quad GDL = 1 - 2 \frac{\sum_{l=1}^2 w_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^2 w_l \sum_n r_{ln} + p_{ln}} \quad (2)$$

[0022] 其中 w_l 为每个类别的权重, r_{ln} 为类别1在第n个像素值的正确类别, p_{ln} 为模型将像素类别预测为1的概率值,其中 w_l 的计算如公式3所示:

$$[0023] \quad w_l = \frac{1}{\left(\sum_{n=1}^N r_{ln}\right)^2} \quad (3)$$

[0024] 模型的总体损失如公式4所示:

$$[0025] \quad \text{Loss} = \text{Loss}_n + \text{Loss}_c + \alpha \cdot S \quad (4)$$

[0026] 其中 Loss_n 与 Loss_c 分别代表了模型预测细胞核与细胞质时的Generalized Dice损失, S 表示细胞质与细胞核的相关系数, α 为用于调整相关系数影响的超参;最终将新提出的模型整体损失函数加入到模型训练过程中,循环训练不断更新模型参数以达到最佳效果。

[0027] S6、采用训练后的模型对图像进行分割;

[0028] 将待分割图片输入到模型中后,模型会同时输出细胞核和细胞质的分割结果,之

后按公式5对两个输出进行处理,得到细胞图片分割结果;

$$[0029] \quad I(i, j) = \begin{cases} a & \text{if } P_n(i, j) > P_c(i, j) > a \\ b & \text{if } P_c(i, j) > P_n(i, j) > a \\ c & \text{otherwise} \end{cases} \quad (5)$$

[0030] 其中 $I(i, j)$ 表示细胞分割结果图片 (i, j) 位置的像素值, a, b, c 分别表示分割结果图中细胞核、细胞质和背景的像素值, $P_n(i, j)$ 和 $P_c(i, j)$ 分别表示细胞核和细胞质分割结果在 (i, j) 位置像素正确分类的概率, a 表示设定的阈值。

[0031] 发明效果

[0032] 新提出的模型不同于UNet模型和传统的分割模型,新模型能够针对细胞核和细胞质分割的特点进行分割;在细胞结构中,细胞核和细胞质在空间上具有高度的相关性,因此通过设计独特的网络结构和损失函数来处理这类特殊的任务;同时利用自监督学习对分割模型参数进行初始化,使模型能够把握细胞核形状呈现椭圆或圆形的特征,从而使分割结果更为精确。使用本发明可更好地确定细胞轮廓和位置,可有效地帮助人工智能对异常细胞进行分级和诊断。

附图说明

[0033] 图1为细胞核与细胞质分割过程示意图;

[0034] 图2为分割模型结构图;

[0035] 图3为人造圆形和椭圆分割数据示意图;

[0036] 图4为模型效果展示。

[0037] 具体实施方式

[0038] 具体实施方式一:

[0039] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整的描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0040] 如图1所示本文提供的细胞核与细胞质的联合分割方法,主要包含以下步骤:

[0041] S1.采集细胞图片用于标记和数据增强;

[0042] S2.构建改进的UNet模型用于图像分割;

[0043] S3.用自监督学习方法初始化模型参数,引入细胞分割的先验知识;

[0044] S4.计算细胞核与细胞质分割结果之间的相关系数;

[0045] S5.计算模型总体损失函数并更新模型参数;

[0046] S6采用训练后的模型对图像进行分割。

[0047] 本发明实施例中,首先需要在不同的细胞图片中选择大量的细胞数据,并且标记出细胞核与细胞质对应的掩膜图片。在UNet模型的基础上对其进行了改进,包括一个下采样模块和两个上采样模块,通过人工构造的圆形和椭圆形分割图对模型进行训练,使得模型能够提前掌握细胞分割的先验知识。根据任务特点和模型结构设计模型损失函数,可以有效地改善模型的分割效果。

[0048] 下面对本发明实施例进行详细的说明:

[0049] 如图1所示算法的实现包含步骤：

[0050] S1、采集细胞图片用于标记和数据增强；

[0051] 通过自动扫描仪获取到细胞图片，将图片裁剪为宽高都为512*512的小图片，由专业人员利用标注软件标记细胞核与细胞质的轮廓，并生成对应的掩膜图片；对标记数据进行数据增强，数据增强的具体方法包括以下子步骤：

[0052] S1-1对图像颜色的对比度、饱和度、颜色进行变换；

[0053] S1-2对图片进行随机区域裁剪，并且随机旋转；

[0054] S1-3调整范围为原始图像大小的0.1到2.0，如果图像小于原始图像大小，则用灰色填充该图像。

[0055] S2、构建改进的UNet模型用于图像分割；

[0056] 利用Pytorch框架实现改进后的UNet模型，模型在原有UNet模型的基础上添加一个新的上采样层，并且通过跳跃连接将下采样层特征融入到上采样层中，网络结构如图2所示。

[0057] S3、用自监督学习方法初始化模型参数，引入细胞分割的先验知识；

[0058] 首先通过OpenCV生成大小不一、圆度不同的椭圆和正圆作为细胞核，然后通过人工模拟细胞质形状绘制了20个模版，最后用生成的细胞核和细胞质组成一个完整的细胞图像，如图3所示；用生成的细胞图像和标记图像对模型进行训练，训练结束后将模型参数作为模型初始参数用于实际的细胞分割任务的训练中。

[0059] S4、计算细胞核与细胞质分割结果之间的相关系数；

[0060] 通过公式6计算细胞质与细胞核的相关系数，这里的相关系数是越小越好，表示细胞核与细胞质单独预测结果越接近于真实的两者空间上的分布；

$$[0061] \quad S = \frac{2|\bar{X} \cap \bar{Y}|}{|\bar{X}| + |\bar{Y}|} + \exp\left(\frac{\|f_n - f_c\|_2}{2\sigma^2}\right) - 1 \quad (6)$$

[0062] 其中S表示细胞核与细胞质的关系系数， \bar{X} 和 \bar{Y} 分别表示细胞核与细胞质的分割结果， f_n 和 f_c 分别表示分割结果在细胞核的外边界集和细胞质的内边界集， σ 表示边界的数量。

[0063] S5、计算模型总体损失函数并更新模型参数；

[0064] 本文采用Generalized Dice损失分别计算模型预测细胞核与细胞质的损失，具体公式如公式7所示：

$$[0065] \quad GDL = 1 - 2 \frac{\sum_{l=1}^2 w_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^2 w_l \sum_n r_{ln} + p_{ln}} \quad (7)$$

[0066] 其中 w_l 为每个类别的权重， r_{ln} 为类别1在第n个像数值的正确类别， p_{ln} 为模型将像素类别预测为1的概率值，其中 w_l 的计算如公式8所示：

$$[0067] \quad w_l = \frac{1}{\left(\sum_{n=1}^N r_{ln}\right)^2} \quad (8)$$

[0068] 模型的总体损失如公式9所示：

[0069] $Loss = Loss_n + Loss_c + \alpha \cdot S$ (9)

[0070] 其中 $Loss_n$ 和 $Loss_c$ 分别代表了模型预测细胞核和细胞质时的Generalized Dice loss损失, S 表示模型预测细胞质和细胞核的相关系数, α 为用于调整相关系数影响的超参,实施中将其设置为3;最终将新提出的模型整体损失函数加入到模型训练过程中,循环训练不断更新模型参数以达到最佳效果。

[0071] S6、采用训练后的模型对图像进行分割;

[0072] 将待分割图片输入到模型中后,模型会同时输出细胞核和细胞质的分割结果,之后按公式10对两个输出进行处理,得到细胞图片分割结果;

[0073]
$$I(i, j) = \begin{cases} 255 & \text{if } P_n(i, j) > P_c(i, j) > 0.5 \\ 100 & \text{if } P_c(i, j) > P_n(i, j) > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

[0074] 其中 $I(i, j)$ 表示细胞分割结果图片 (i, j) 位置的像素值, $P_n(i, j)$ 和 $P_c(i, j)$ 分别表示细胞核和细胞质分割结果在 (i, j) 位置像素正确分类的概率。

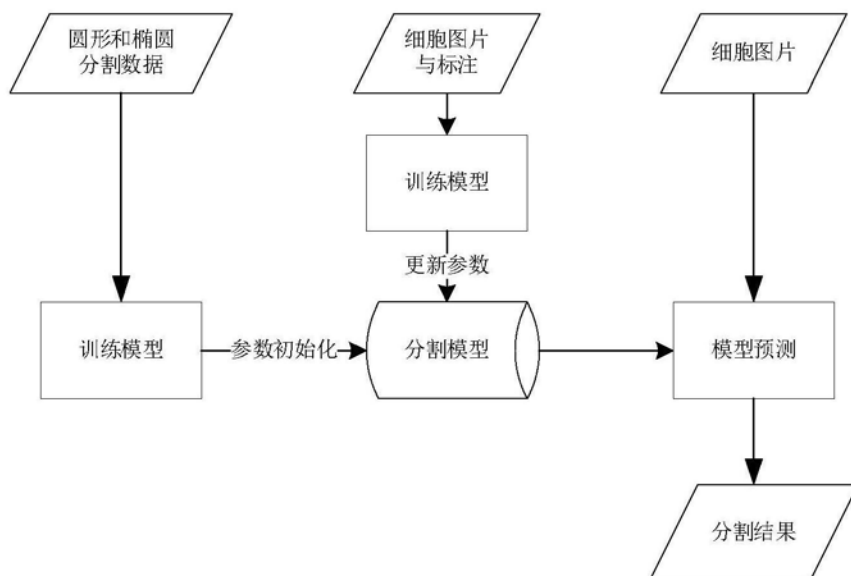


图1

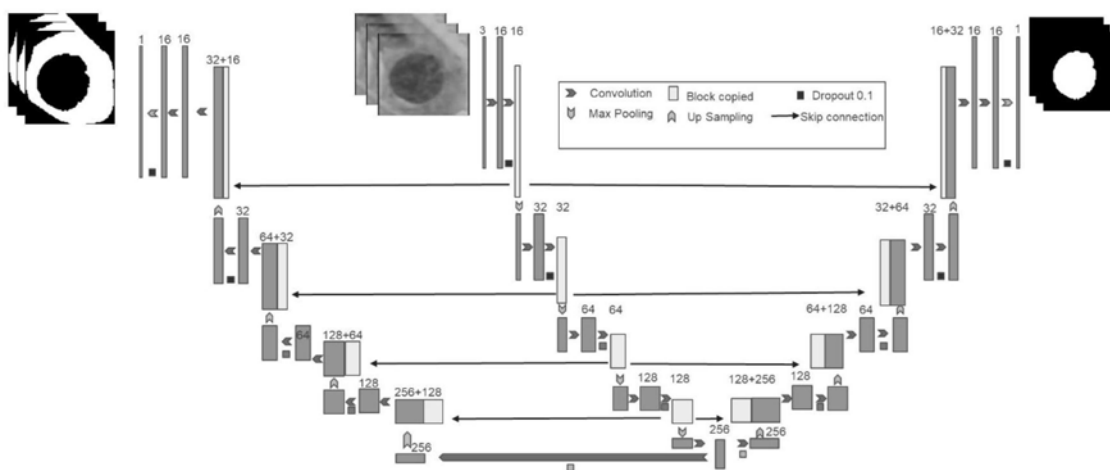


图2

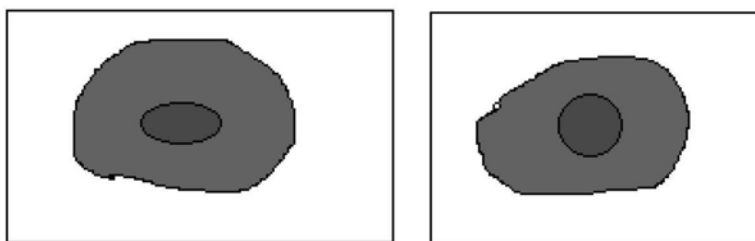


图3

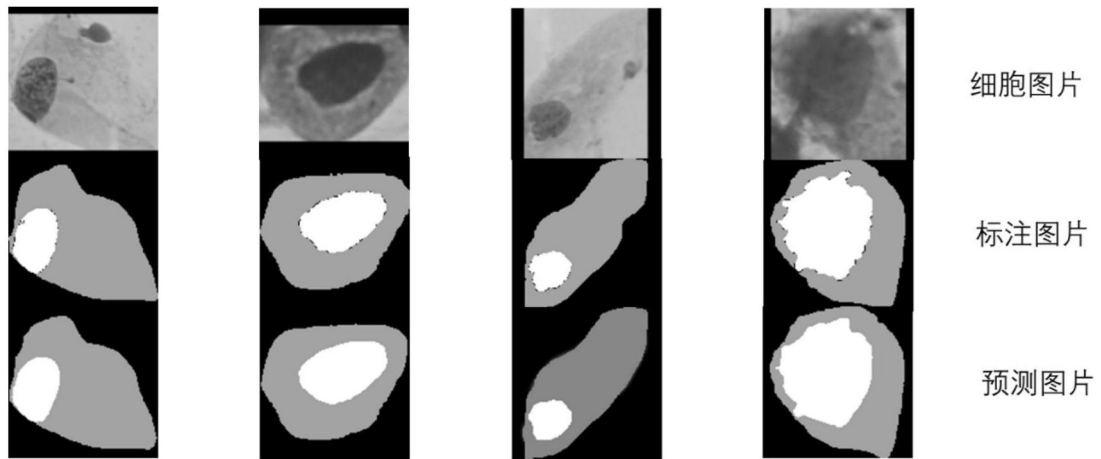


图4