



(12) 发明专利申请

(10) 申请公布号 CN 112419253 A

(43) 申请公布日 2021. 02. 26

(21) 申请号 202011279541.4

(22) 申请日 2020.11.16

(71) 申请人 中山大学

地址 510275 广东省广州市海珠区新港西路135号

(72) 发明人 李伟忠 杨欢 柯尊富 陈丽丽

(74) 专利代理机构 广州三环专利商标代理有限公司 44202

代理人 胡枫 曹万菊

(51) Int. Cl.

G06T 7/00 (2017.01)

G06T 7/11 (2017.01)

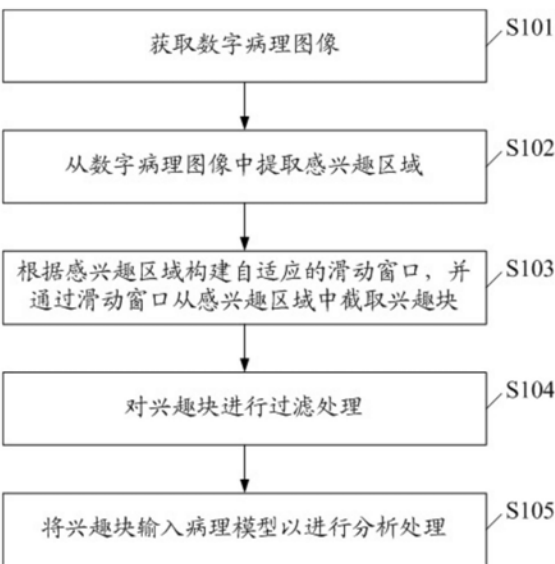
权利要求书3页 说明书9页 附图4页

(54) 发明名称

数字病理图像分析方法、系统、设备及存储介质

(57) 摘要

本发明公开了一种数字病理图像分析方法，包括：获取数字病理图像；从所述数字病理图像中提取感兴趣区域；根据所述感兴趣区域构建自适应的滑动窗口，并通过所述滑动窗口从所述感兴趣区域中截取兴趣块；对所述兴趣块进行过滤处理；将所述兴趣块输入病理模型以进行分析处理。本发明还公开了一种数字病理图像分析系统，一种计算机设备及一种计算机可读存储介质。采用本发明，可对病理图像进行预处理，有效提升人工智能算法的分类准确度，适用性更广、容错性更高、处理难度更低。



1. 一种数字病理图像分析方法,其特征在于,包括:

获取数字病理图像;

从所述数字病理图像中提取感兴趣区域;

根据所述感兴趣区域构建自适应的滑动窗口,并通过所述滑动窗口从所述感兴趣区域中截取兴趣块;

对所述兴趣块进行过滤处理;

将所述兴趣块输入病理模型以进行分析处理。

2. 如权利要求1所述的数字病理图像分析方法,其特征在于,所述根据感兴趣区域构建自适应的滑动窗口并通过滑动窗口从感兴趣区域中截取兴趣块的步骤包括:

判断所述感兴趣区域的最小外接矩形的边长是否小于滑动窗口的边长,判断为是时,则删除所述感兴趣区域,判断为否时,则保留所述感兴趣区域;

根据所述感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定所述感兴趣区域对应的滑动窗口的滑动步长;

根据所述滑动窗口的边长及滑动步长,采用所述滑动窗口法遍历所述感兴趣区域以提取兴趣块。

3. 如权利要求2所述的数字病理图像分析方法,其特征在于,所述根据感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定感兴趣区域对应的滑动窗口的滑动步长的步骤包括:

计算总体感兴趣区域的数量 N ,其中, $N=N_1+N_2+\cdots+N_i+\cdots+N_k$, $i \in (1,2,\cdots,k)$, N_i 为每个类别的感兴趣区域数量, k 为类别数量;

计算各感兴趣区域的面积 s_j ,其中, $s_j=w_j \cdot h_j$, w_j 为第 j 个感兴趣区域的最小外接矩的宽, h_j 为第 j 个感兴趣区域的最小外接矩的高;

计算各类别感兴趣区域的面积均值 \bar{S}^i ,其中, $\bar{S}^i = \frac{1}{N_i} \sum_{j=1}^{N_i} s_j$;

计算各类别感兴趣区域的面积均值 \bar{S}^i 与总体感兴趣区域平均面积的组间系数 ω^i ,其

$$\text{中, } \omega^i = \frac{N_i \cdot \bar{S}^i}{\sum_{i=1}^k N_i \cdot \bar{S}^i};$$

计算各感兴趣区域的面积 s_j 与各类别感兴趣区域的面积均值 \bar{S}^i 的组内系数 ω^j ,其中,

$$\omega^j = \frac{s_j}{\bar{S}^i};$$

根据所述组间系数 ω^i 、组内系数 ω^j 及初始化滑动步长 l_0 ,计算第 j 个感兴趣区域对应的滑动窗口的滑动步长 l_j ,其中, $l_j = \omega^i \cdot \omega^j \cdot l_0$ 。

4. 如权利要求1所述的数字病理图像分析方法,其特征在于,所述对兴趣块进行过滤处理的步骤包括:

采用阈值算法分别计算每个兴趣块的病理组织覆盖比;

判断所述兴趣块的病理组织覆盖比是否低于预设阈值,

判断为是时,则删除所述兴趣块,

判断为否时,则保留所述兴趣块。

5.如权利要求1所述的数字病理图像分析方法,其特征在于,所述从数字病理图像中提取感兴趣区域的步骤包括:

针对无注释的数字病理图像,将组织覆盖的区域提取为感兴趣区域;

针对带注释的数字病理图像,根据注释文件提取感兴趣区域。

6.一种数字病理图像分析系统,其特征在于,包括:

获取模块,用于获取数字病理图像;

提取模块,用于从所述数字病理图像中提取感兴趣区域;

截取模块,用于根据所述感兴趣区域构建自适应的滑动窗口,并通过所述滑动窗口从所述感兴趣区域中截取兴趣块;

过滤模块,用于对所述兴趣块进行过滤处理;

分析模块,用于将所述兴趣块输入病理模型以进行分析处理。

7.如权利要求6所述的数字病理图像分析系统,其特征在于,所述截取模块包括:

区域过滤单元,用于判断所述感兴趣区域的最小外接矩形的边长是否小于滑动窗口的边长,判断为是时,则删除所述感兴趣区域,判断为否时,则保留所述感兴趣区域;

自适应单元,用于根据所述感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定所述感兴趣区域对应的滑动窗口的滑动步长;

切分单元,用于根据所述滑动窗口的边长及滑动步长,采用所述滑动窗口法遍历所述感兴趣区域以提取兴趣块。

8.如权利要求7所述的数字病理图像分析系统,其特征在于,所述自适应单元包括:

区域计算子单元,用于计算总体感兴趣区域的数量 N ,其中, $N=N_1+N_2+\cdots+N_i+\cdots+N_k$, $i \in (1,2,\cdots,k)$, N_i 为每个类别的感兴趣区域数量, k 为类别数量;

面积计算子单元,用于计算各感兴趣区域的面积 s_j ,其中, $s_j=w_j \cdot h_j$, w_j 为第 j 个感兴趣区域的最小外接矩的宽, h_j 为第 j 个感兴趣区域的最小外接矩的高;

均值计算子单元,用于计算各类别感兴趣区域的面积均值 \bar{S}^i ,其中, $\bar{S}^i = \frac{1}{N_i} \sum_{j=1}^{N_i} s_j$;

组间系数计算子单元,用于计算各类别感兴趣区域的面积均值 \bar{S}^i 与总体感兴趣区域平均面积的组间系数 ω^i ,其中, $\omega^i = \frac{N_i \cdot \bar{S}^i}{\sum_{i=1}^k N_i \cdot \bar{S}^i}$;

组内系数计算子单元,用于计算各感兴趣区域的面积 s_j 与各类别感兴趣区域的面积均值 \bar{S}^i 的组内系数 ω^j ,其中, $\omega^j = \frac{s_j}{\bar{S}^i}$;

步长计算子单元,用于根据所述组间系数 ω^i 、组内系数 ω^j 及初始化滑动步长 l_0 ,计算第 j 个感兴趣区域对应的滑动窗口的滑动步长 l_j ,其中, $l_j = \omega^i \cdot \omega^j \cdot l_0$ 。

9.一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至5中任一项所述的方法的步骤。

10.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序

被处理器执行时实现权利要求1至5中任一项所述的方法的步骤。

数字病理图像分析方法、系统、设备及存储介质

技术领域

[0001] 本发明涉及形态学图像处理技术领域,尤其涉及一种数字病理图像分析方法、一种数字病理图像分析系统、一种计算机设备及一种计算机可读存储介质。

背景技术

[0002] 肺癌的发病率仅次于前列腺癌(男性)和乳腺癌(女性),而致死率居癌症之首,虽然癌症整体死亡率呈现逐年下降的趋势,但肺癌具有早期症状不明显、高异质性与受环境暴露及生活习惯影响(吸烟)等特点,因而发病率高,且易漏诊或错诊而导致预后较差。目前,低剂量计算机断层扫描(low-dose computerized tomography,LDCT)是肺癌筛查的主要方法,但假阳性过高(26%~58%)的问题使其无法作为肺癌确诊的依据。病理学诊断是肺癌诊断的金标准,而随着免疫治疗和分子靶向疗法等新生肿瘤技术的快速发展,引入免疫组化和基因检测技术来提升肺癌分型的准确度,更能适应肺癌个性化治疗的需求。数字切片扫描仪的诞生是病理学发展的里程碑,传统玻片的数字化使得病理诊断的智能化成为可能,但亿级像素的超高分辨率图像构成了算法与数据之间的无形阻碍,如何预处理与后处理数字病理切片,使其能够被人工智能算法所使用,达到提升诊断精度的目的,是本发明主要解决的问题。

[0003] 目前,针对分类任务,数字切片预处理的大体思路是一致的,即抽取金字塔结构的全扫描数字病理图像(WSI)的某一层(layer),提取感兴趣的区域(Region of Interest, ROI),将其切分成小块(tile)来预测;后处理的思路是将小块层级的预测结果整合为切片层级的预测结果。虽然整体思路一致,但具体的技术路线和参数设置各有不同,而这种差别在一定程度上影响人工智能算法的预测结果。

[0004] 具体地,现有的预处理技术方案包括:

[0005] 1、ROI区域切分成一定程度上重叠的tiles:滑动窗口的步长是根据横向或纵向相邻的两个tiles之间的重叠程度来设定,具体地, $s = l \cdot (1-p)$,其中,s为步长,l为滑窗(正方形)的边长,p为tile重叠程度且p的取值范围是0~1。

[0006] 2、ROI区域切分成互不重叠的tiles:首先,设定tile的大小(比如 256×256 像素),然后使用滑动窗口法(slidingwindow)遍历ROI区域来提取tiles,滑动窗口的尺寸和步长均等于tile的大小,该方案是方案1的特殊情况,即 $s = l$ 。但是,固定步长的方法无法适应切片尺寸或ROI区域大小的差异,一个大标本(原始尺寸或ROI区域较大的切片)可生成 $10^3 \sim 10^5$ 个tiles,信息过于冗余,引入过多噪声或导致过拟合问题,影响分类性能;而一个小标本(原始尺寸或ROI区域较小的切片)可能仅仅生成 $1 \sim 10^3$ 个tiles,以致信息丰度不足;同时,小标本的重要性往往不亚于大标本,且在人工智能的病理学应用场景中占有特殊地位,tile量级的差距很可能导致或加剧类别失衡问题,降低少数类的预测精度。

[0007] 3、ROI区域内随机截取tiles:在ROI区域内随机选取若干个点作为tile顶点(左上角顶点),提取tiles。但是,随机截取tiles具有不可控性,可能造成信息冗余或遗失,失去了精确刻画ROI边界的意义。

[0008] 由上可知,现有的预处理技术方案未能满足实际应用,仍有待改进。

发明内容

[0009] 本发明所要解决的技术问题在于,提供一种数字病理图像分析方法、系统、计算机设备及计算机可读存储介质,可对病理图像进行预处理,有效提升人工智能算法的分类准确度。

[0010] 为了解决上述技术问题,本发明提供了一种数字病理图像分析方法,包括:获取数字病理图像;从所述数字病理图像中提取感兴趣区域;根据所述感兴趣区域构建自适应的滑动窗口,并通过所述滑动窗口从所述感兴趣区域中截取兴趣块;对所述兴趣块进行过滤处理;将所述兴趣块输入病理模型以进行分析处理。

[0011] 作为上述方案的改进,所述根据感兴趣区域构建自适应的滑动窗口并通过滑动窗口从感兴趣区域中截取兴趣块的步骤包括:判断所述感兴趣区域的最小外接矩形的边长是否小于滑动窗口的边长,判断为是时,则删除所述感兴趣区域,判断为否时,则保留所述感兴趣区域;根据所述感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定所述感兴趣区域对应的滑动窗口的滑动步长;根据所述滑动窗口的边长及滑动步长,采用所述滑动窗口法遍历所述感兴趣区域以提取兴趣块。

[0012] 作为上述方案的改进,所述根据感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定感兴趣区域对应的滑动窗口的滑动步长的步骤包括:计算总体感兴趣区域的数量 N ,其中, $N=N_1+N_2+\cdots+N_i+\cdots+N_k$, $i \in (1, 2, \cdots, k)$, N_i 为每个类别的感兴趣区域数量, k 为类别数量;计算各感兴趣区域的面积 s_j ,其中, $s_j=w_j \cdot h_j$, w_j 为第 j 个感兴趣区域的最小外接矩的宽, h_j 为第 j 个感兴趣区域的最小外接矩的高;计算各类别感兴趣区域的面积均值 \bar{S}^i ,其中, $\bar{S}^i = \frac{1}{N_i} \sum_{j=1}^{N_i} s_j$;计算各类别感兴趣区域的面积均值 \bar{S}^i 与总体感兴趣区域平均

面积的组间系数 ω^i ,其中, $\omega^i = \frac{N_i \cdot \bar{S}^i}{\sum_{i=1}^k N_i \cdot \bar{S}^i}$;计算各感兴趣区域的面积 s_j 与各类别感兴趣区域

的面积均值 \bar{S}^i 的组内系数 ω^j ,其中, $\omega^j = \frac{s_j}{\bar{S}^i}$;根据所述组间系数 ω^i 、组内系数 ω^j 及初始

化滑动步长 l_0 ,计算第 j 个感兴趣区域对应的滑动窗口的滑动步长 l_j ,其中, $l_j = \omega^i \cdot \omega^j \cdot l_0$ 。

[0013] 作为上述方案的改进,所述对兴趣块进行过滤处理的步骤包括:采用阈值算法分别计算每个兴趣块的病理组织覆盖比;判断所述兴趣块的病理组织覆盖比是否低于预设阈值,判断为是时,则删除所述兴趣块,判断为否时,则保留所述兴趣块。

[0014] 作为上述方案的改进,所述从数字病理图像中提取感兴趣区域的步骤包括:针对无注释的数字病理图像,将组织覆盖的区域提取为感兴趣区域;针对带注释的数字病理图像,根据注释文件提取感兴趣区域。

[0015] 相应地,本发明还提供了一种数字病理图像分析系统,包括:获取模块,用于获取数字病理图像;提取模块,用于从所述数字病理图像中提取感兴趣区域;截取模块,用于根

据所述感兴趣区域构建自适应的滑动窗口,并通过所述滑动窗口从所述感兴趣区域中截取兴趣块;过滤模块,用于对所述兴趣块进行过滤处理;分析模块,用于将所述兴趣块输入病理模型以进行分析处理。

[0016] 作为上述方案的改进,所述截取模块包括:区域过滤单元,用于判断所述感兴趣区域的最小外接矩形的边长是否小于滑动窗口的边长,判断为是时,则删除所述感兴趣区域,判断为否时,则保留所述感兴趣区域;自适应单元,用于根据所述感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定所述感兴趣区域对应的滑动窗口的滑动步长;切分单元,用于根据所述滑动窗口的边长及滑动步长,采用所述滑动窗口法遍历所述感兴趣区域以提取兴趣块。

[0017] 作为上述方案的改进,所述自适应单元包括:区域计算子单元,用于计算总体感兴趣区域的数量 N ,其中, $N=N_1+N_2+\dots+N_i+\dots+N_k$, $i \in (1, 2, \dots, k)$, N_i 为每个类别的感兴趣区域数量, k 为类别数量;面积计算子单元,用于计算各感兴趣区域的面积 s_j ,其中, $s_j=w_j \cdot h_j$, w_j 为第 j 个感兴趣区域的最小外接矩的宽, h_j 为第 j 个感兴趣区域的最小外接矩的高;均值计算

子单元,用于计算各类别感兴趣区域的面积均值 \bar{S}^i ,其中, $\bar{S}^i = \frac{1}{N_i} \sum_{j=1}^{N_i} s_j$;组间系数计算子

单元,用于计算各类别感兴趣区域的面积均值 \bar{S}^i 与总体感兴趣区域平均面积的组间系数

ω^i ,其中, $\omega^i = \frac{N_i \cdot \bar{S}^i}{\sum_{i=1}^k N_i \cdot \bar{S}^i}$;组内系数计算子单元,用于计算各感兴趣区域的面积 s_j 与各类

别感兴趣区域的面积均值 \bar{S}^i 的组内系数 ω^j ,其中, $\omega^j = \frac{s_j}{\bar{S}^i}$;步长计算子单元,用于根据所

述组间系数 ω^i 、组内系数 ω^j 及初始化滑动步长 l_0 ,计算第 j 个感兴趣区域对应的滑动窗口的滑动步长 l_j ,其中, $l_j = \omega^i \cdot \omega^j \cdot l_0$ 。

[0018] 相应地,本发明还提供了一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现上述数字病理图像分析方法的步骤。

[0019] 相应地,本发明还提供了一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现上述数字病理图像分析方法的步骤。

[0020] 实施本发明,具有如下有益效果:

[0021] 本发明数字病理图像分析方法中创造性地采用自适应性的滑动窗口提取感兴趣区域的兴趣块,其中,所述自适应的滑动窗口的步长可根据感兴趣区域的实际情况进行调节,能够保证信息的丰度,从而控制冗余度。

[0022] 具体地,自适应性体现在以下两个方面:(1)考虑组间感兴趣区域面积和数量的差异,为滑动窗口的滑动步长设计组间系数;(2)考虑组内各感兴趣区域面积的差异,为每个感兴趣区域设计滑动步长的组内系数。因此,本发明在提取兴趣块的过程中,需要同时考虑组内和组间差异,从而合理设计滑动窗口的滑动步长,既避免了珍稀样本的信息丢失,又削弱了大样本的信息冗余,合理规避类别失衡问题,利于后期分析处理过程中的模型训练和优化。

附图说明

- [0023] 图1是本发明数字病理图像分析方法的第一实施例流程图；
- [0024] 图2是本发明数字病理图像分析方法的第二实施例流程图；
- [0025] 图3是本发明中数字病理图像的示意图；
- [0026] 图4是本发明数字病理图像分析系统的结构示意图；
- [0027] 图5是本发明中截取模块的结构示意图；
- [0028] 图6是本发明中自适应单元的结构示意图。

具体实施方式

[0029] 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明作进一步地详细描述。

[0030] 参见图1，图1显示了本发明数字病理图像分析方法的第一实施例流程图，其包括：

[0031] S101，获取数字病理图像。

[0032] 所述数字病理图像需由病理医生进行初步处理后，再输入数字病理图像分析系统。

[0033] S102，从数字病理图像中提取感兴趣区域。

[0034] 具体地，所述从数字病理图像中提取感兴趣区域的方法包括：

[0035] (1) 针对无注释的数字病理图像，将组织覆盖的区域提取为感兴趣区域；

[0036] (2) 针对带注释的数字病理图像，根据注释文件提取感兴趣区域。

[0037] 需要说明的是，对于肺癌分类问题，类标签为“正常 (Normal Lung, NL)”的数字病理图像，是不需要病理医生额外添加注释的，因此，针对无注释的数字病理图像 (即类标签为“正常”的数字病理图像)，所有组织覆盖的区域均属于感兴趣区域；而其他类的切片，则需病理医生添加注释，再根据注释文件来提取感兴趣区域。其中，所述注释文件中记录了数字病理图像中感兴趣区域的类别，所述类别包括肺腺癌、肺鳞癌、小细胞肺癌、肺结核、机化性肺炎、正常肺等，但不以此为限制。

[0038] S103，根据感兴趣区域构建自适应的滑动窗口，并通过滑动窗口从感兴趣区域中截取兴趣块。

[0039] 现有技术中，通常采用滑动步长固定的滑动窗口对感兴趣区域进行切分。与现有技术不同的是，本发明中，采用自适应性的滑动窗口提取感兴趣区域的兴趣块。其中，所述自适应的滑动窗口是指，滑动窗口的步长可根据感兴趣区域的实际情况进行调节，能够保证信息的丰度，从而控制冗余度。

[0040] 例如，可根据感兴趣区域的大小和数量来调整滑动窗口的滑动步长，从而使得兴趣块的截取不受感兴趣区域面积大小或数量多少的影响。

[0041] S104，对兴趣块进行过滤处理。

[0042] 需要说明的是，不同的兴趣块中所记录的信息量不同，为了提高效率，在对兴趣块进行分析处理前，需过滤掉信息量较小的兴趣块。具体地，所述对兴趣块进行过滤处理的步骤包括：

[0043] (1) 采用阈值算法分别计算每个兴趣块的病理组织覆盖比。

[0044] 具体地，可采用QTSU法 (大津阈值分割法) 来判断每个兴趣块的病理组织覆盖比，

但不以此为限制,只要可实现覆盖比的计算即可。

[0045] (2) 判断所述兴趣块的病理组织覆盖比是否低于预设阈值,判断为是时,则删除所述兴趣块,判断为否时,则保留所述兴趣块。

[0046] 需要说明的是,所述预设阈值可根据实际需求进行设置。针对后续用于训练和验证病理模型的训练集和验证集,所述预设阈值可设为0.5;而针对后续用于测试病理模块的测试集,所述预设阈值可设为0.1。

[0047] S105,将兴趣块输入病理模型以进行分析处理。

[0048] 在实际应用中,可将兴趣块输入病理模型以对病理模型进行训练、测试和验证,从而实现对兴趣块的有效分析。其中,所述病理模型可由技术人员根据实际情况进行构建,本发明中不作限定,而本发明仅对输入的病理图像进行预处理,以使病理模型能获得更为优质、有效、针对性强的训练集、测试集和验证集。

[0049] 因此,通过本发明可有效提升人工智能算法的分类准确度,适用性更广、容错性更高、处理难度更低且更贴合临床实际,使科研人员不必在图像处理环节过度消耗,能够集中精力研究算法与实验设计。

[0050] 参见图2,图2显示了本发明数字病理图像分析方法的第二实施例流程图,其包括:

[0051] S201,获取数字病理图像。

[0052] S202,从数字病理图像中提取感兴趣区域。

[0053] S203,判断感兴趣区域的最小外接矩形的边长是否小于滑动窗口的边长,判断为是时,则删除感兴趣区域,判断为否时,则保留感兴趣区域。

[0054] 如图3所示,数字病理图像中的感兴趣区域可以是任意封闭曲线围成的特定形状(如曲线A),但为了处理方便,本发明将所述ROI区域的特定形状的最小外接矩形作为参考区域(如曲线B)。

[0055] 所述滑动窗口的边长可根据实际情况进行预先设置,本发明中不作限定。相应地,由于本发明中的病理模型采用卷积神经网络,而卷积神经网络的输入通常是正方形图像(如, 224×224 , 256×256 等),因此在切分兴趣块时,可将也滑动窗口设计为正方形。

[0056] 需要说明的是,兴趣块是由滑动窗口进行切分的,即滑动窗口的尺寸等于兴趣块的尺寸。因此,若所述感兴趣区域的最小外接矩形的长或宽小于滑动窗口的边长时,则可放弃所述感兴趣区域,否则,保留所述感兴趣区域。

[0057] S204,根据感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定感兴趣区域对应的滑动窗口的滑动步长。

[0058] 与现有技术不同的是,本发明根据该类别下所有感兴趣区域与总体感兴趣区域的关系来确定滑动窗口遍历该类别感兴趣区域所使用的滑动步长,再根据具体某个感兴趣区域与其所在类别感兴趣区域的关系来确定其滑动窗口的滑动步长。优选地,所述关系包括大小、数量、均值等关系,但不以此为限制。

[0059] 具体地,所述根据感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定感兴趣区域对应的滑动窗口的滑动步长的步骤包括:

[0060] (1) 计算总体感兴趣区域的数量N。

[0061] 计算所有可提取兴趣块的感兴趣区域的数量,记作N。

[0062] $N = N_1 + N_2 + \dots + N_i + \dots + N_k, i \in (1, 2, \dots, k)$

[0063] 其中, N_i 为每个类别的感兴趣区域数量, k 为类别数量;

[0064] (2) 计算各感兴趣区域的面积 s_j 。

[0065] $s_j = w_j \cdot h_j$

[0066] 其中, w_j 为第 j 个感兴趣区域的最小外接矩的宽, h_j 为第 j 个感兴趣区域的最小外接矩的高;

[0067] (3) 计算各类别感兴趣区域的面积均值 \bar{S}^i 。

[0068]
$$\bar{S}^i = \frac{1}{N_i} \sum_{j=1}^{N_i} s_j$$

[0069] (4) 计算各类别感兴趣区域的面积均值 \bar{S}^i 与总体感兴趣区域平均面积的组间系数 ω^i 。

[0070]
$$\omega^i = \frac{N_i \cdot \bar{S}^i}{\sum_{i=1}^k N_i \cdot \bar{S}^i}$$

[0071] (5) 计算各感兴趣区域的面积 s_j 与各类别感兴趣区域的面积均值 \bar{S}^i 的组内系数 ω^j 。

[0072]
$$\omega^j = \frac{s_j}{\bar{S}^i}$$

[0073] (6) 根据所述组间系数 ω^i 、组内系数 ω^j 及初始化滑动步长 l_0 , 计算第 j 个感兴趣区域对应的滑动窗口的滑动步长 l_j 。

[0074] $l_j = \omega^i \cdot \omega^j \cdot l_0$

[0075] 本发明在设定滑动窗口的初始化滑动步长的基础上, 采用自适应性滑动窗口法提取兴趣区域的兴趣块, 具体地, 自适应性体现在以下两个方面:

[0076] 一、考虑组 (类) 间感兴趣区域面积和数量的差异, 为滑动窗口的滑动步长设计组间系数;

[0077] 二、考虑组 (类) 内各感兴趣区域面积的差异, 为每个感兴趣区域设计滑窗步长的组内系数。

[0078] 因此, 通过滑动窗口的滑动步长的自适应性调节, 可保证大面积的感兴趣区域的信息不至于被反复提取, 小面积的感兴趣区域的信息也不至于被忽略, 且弱化了可能出现的组间失衡问题, 利于后期分析处理过程中的模型训练和优化。

[0079] 需要说明的是, l_0 为初始化 (默认) 滑动步长, 若兴趣块的边长为 a , 则相邻两个兴趣块之间的重叠度为 θ , 则 $l_0 = a \cdot \theta$ 。

[0080] 本发明中, 兴趣块的边长 $a = 256$ 。相应地, 针对后续用于训练和验证病理模型的训练集和验证集, 所述重叠度 θ 可设为 0.5; 而针对后续用于测试病理模块的测试集, 所述重叠度 θ 可设为 0.1。

[0081] S205, 根据滑动窗口的边长及滑动步长, 采用滑动窗口法遍历感兴趣区域以提取兴趣块。

[0082] 因此, 本发明在提取兴趣块的过程中, 需要同时考虑组内和组间差异 (如, 组间系数、组内系数), 从而合理设计滑动窗口的滑动步长, 既避免了珍稀样本的信息丢失, 又削弱

了大样本的信息冗余,合理规避类别失衡问题,有助于构建无偏模型。

[0083] S206,对兴趣块进行过滤处理。

[0084] S207,将兴趣块输入病理模型以进行分析处理。

[0085] 综上所述,本发明数字病理图像分析方法中创造性地采用自适应性的滑动窗口提取感兴趣区域的兴趣块,其中,所述自适应的滑动窗口的步长可根据感兴趣区域的实际情况进行调节,能够保证信息的丰度,从而控制冗余度。

[0086] 进一步,在实际应用中还可对病理模型输出的预测结果进行后处理。具体地,所采用的后处理技术方案包括:

[0087] 1、Tile计数法:分别统计预测结果中每一类别的兴趣块数量,以兴趣块数量最多的类别所对应的类标签作为切片的预测类别。

[0088] 2、概率均值最大法:分别统计预测结果中每一类别的兴趣块的概率之和并求均值,以概率均值最大的类别所对应的类标签作为切片的预测类别。

[0089] 3、特征工程法:首先,根据兴趣块层级的预测结果,人工设计和提取特征;然后,利用机器学习分类算法对特征进行学习与拟合,得到训练好的分类器;最后,将待整合的兴趣块层级结果输入到该分类器中,得出切片层级的结论。

[0090] 4、相互验证法:在病理模型中构建两个网络模块,第一个网络模块执行肺部病理组织类型预测,第二个网络模块执行肿瘤组织的免疫组化(IHC)蛋白表达预测,通过免疫组化蛋白表达状态预测与病理组织类型预测的相互参照和验证,提升肺癌病理分型准确率。相互参照和验证时,对于每种组织类别,先由病理专家根据实验室诊断经验设置初始阈值,并由采用网络搜索策略寻找最优阈值;再计算每一类别兴趣块的数量占总体兴趣块数量的比值,如果比值高于该类别的最优阈值,则该类别被视为该片的类别标签。其中,所述最优阈值集合如下:

$$[0091] \quad \begin{cases} LUAD = 0.1 \\ LUSC = 0.1 \\ SCLC = 0.1 \\ PTB = 0.3 \\ OP = 0.4 \\ NL = 0.9 \end{cases}$$

[0092] 其中,LUAD为肺腺癌,LUSC为肺鳞癌,SCLC为小细胞肺癌,PTB为肺结核,OP为机化性肺炎,NL为正常肺。由上可知,各类别的阈值具有明显的分层性:对于肿瘤类别,采用较低阈值,而对于非肿瘤的带病类别,则采较高阈值,而正常组织类别,采用最高阈值。

[0093] 因此,相互验证法将临床病理诊断实际使用的阈值与兴趣块计数法相结合,利用网络搜索策略确定最优阈值集,各个阶段使用同一个阈值集,保持判断标准的客观性和一致性,并将块水平的预测结果整合为片水平的预测结果,再利用阈值法得出片的病理类型推断。预测过程更简单客观,便于深度学习算法投入临床使用,不需要人工设计特征,避免叠加分类器所造成的错误级联效应。

[0094] 参见图4,图4显示了本发明数字病理图像分析系统100的具体结构,其包括获取模块1、提取模块2、截取模块3、过滤模块4及分析模块5,具体地:

[0095] 获取模块1,用于获取数字病理图像。所述数字病理图像需由病理医生进行初步处

理后,再输入数字病理图像分析系统,并由获取模块1进行获取。

[0096] 提取模块2,用于从所述数字病理图像中提取感兴趣区域。其中,针对无注释的数字病理图像,将组织覆盖的区域提取为感兴趣区域;针对带注释的数字病理图像,根据注释文件提取感兴趣区域。

[0097] 截取模块3,用于根据所述感兴趣区域构建自适应的滑动窗口,并通过所述滑动窗口从所述感兴趣区域中截取兴趣块。其中,所述自适应的滑动窗口是指,滑动窗口的步长可根据感兴趣区域的实际情况进行调节,能够保证信息的丰度,从而控制冗余度。优选地,截取模块3可根据感兴趣区域的大小和数量来调整滑动窗口的滑动步长,从而使得兴趣块的截取不受感兴趣区域面积大小或数量多少的影响。

[0098] 过滤模块4,用于对所述兴趣块进行过滤处理。具体地,过滤模块4先采用阈值算法分别计算每个兴趣块的病理组织覆盖比,再判断所述兴趣块的病理组织覆盖比是否低于预设阈值,判断为是时,则删除所述兴趣块,判断为否时,则保留所述兴趣块。

[0099] 分析模块5,用于将所述兴趣块输入病理模型以进行分析处理。其中,所述病理模型可由技术人员根据实际情况进行构建。

[0100] 因此,通过本发明中的截取模块3可对感兴趣区域中的兴趣块进行重新截取,能有效提升人工智能算法的分类准确度,适用性更广、容错性更高、处理难度更低且更贴合临床实际,使科研人员不必在图像处理环节过度消耗,能够集中精力研究算法与实验设计。

[0101] 如图5所示,所述截取模块3包括:

[0102] 区域过滤单元31,用于判断所述感兴趣区域的最小外接矩形的边长是否小于滑动窗口的边长,判断为是时,则删除所述感兴趣区域,判断为否时,则保留所述感兴趣区域。

[0103] 自适应单元32,用于根据所述感兴趣区域、各类别感兴趣区域及总体感兴趣区域之间的关系,确定所述感兴趣区域对应的滑动窗口的滑动步长。

[0104] 切分单元33,用于根据所述滑动窗口的边长及滑动步长,采用所述滑动窗口法遍历所述感兴趣区域以提取兴趣块。

[0105] 与现有技术不同的是,本发明中的自适应单元32可根据该类别下所有感兴趣区域与总体感兴趣区域的关系来确定滑动窗口遍历该类别感兴趣区域所使用的滑动步长,再根据具体某个感兴趣区域与其所在类别感兴趣区域的关系来确定其滑动窗口的滑动步长,灵活性及针对性强。优选地,所述关系包括大小、数量、均值等关系,但不以此为限制。

[0106] 如图6所示,所述自适应单元32包括:

[0107] 区域计算子单元321,用于计算总体感兴趣区域的数量 N ,其中, $N=N_1+N_2+\cdots+N_i+\cdots+N_k$, $i \in (1,2,\cdots,k)$, N_i 为每个类别的感兴趣区域数量, k 为类别数量;

[0108] 面积计算子单元322,用于计算各感兴趣区域的面积 s_j ,其中, $s_j=w_j \cdot h_j$, w_j 为第 j 个感兴趣区域的最小外接矩的宽, h_j 为第 j 个感兴趣区域的最小外接矩的高;

[0109] 均值计算子单元323,用于计算各类别感兴趣区域的面积均值 \bar{S}^i ,其中,

$$\bar{S}^i = \frac{1}{N_i} \sum_{j=1}^{N_i} s_j;$$

[0110] 组间系数计算子单元324,用于计算各类别感兴趣区域的面积均值 \bar{S}^i 与总体感兴

趣区域平均面积的组间系数 ω^i , 其中, $\omega^i = \frac{N_i \cdot \bar{S}^i}{\sum_{i=1}^k N_i \cdot \bar{S}^i}$;

[0111] 组内系数计算子单元325, 用于计算各感兴趣区域的面积 s_j 与各类别感兴趣区域的面积均值 \bar{S}^i 的组内系数 ω^j , 其中, $\omega^j = \frac{s_j}{\bar{S}^i}$;

[0112] 步长计算子单元326, 用于根据所述组间系数 ω^i 、组内系数 ω^j 及初始化滑动步长 l_0 , 计算第 j 个感兴趣区域对应的滑动窗口的滑动步长 l_j , 其中, $l_j = \omega^i \cdot \omega^j \cdot l_0$ 。

[0113] 因此, 本发明在提取兴趣块的过程中, 需要同时考虑组内和组间差异 (如, 组间系数、组内系数), 从而合理设计滑动窗口的滑动步长, 既避免了珍稀样本的信息丢失, 又削弱了大样本的信息冗余, 合理规避类别失衡问题, 有助于构建无偏模型。

[0114] 相应地, 本发明还提供了一种计算机设备, 包括存储器和处理器, 所述存储器存储有计算机程序, 所述处理器执行所述计算机程序时实现上述数字病理图像分析方法的步骤。同时, 本发明还提供了一种计算机可读存储介质, 其上存储有计算机程序, 所述计算机程序被处理器执行时实现上述数字病理图像分析方法的步骤。

[0115] 以上所述是本发明的优选实施方式, 应当指出, 对于本技术领域的普通技术人员来说, 在不脱离本发明原理的前提下, 还可以做出若干改进和润饰, 这些改进和润饰也视为本发明的保护范围。

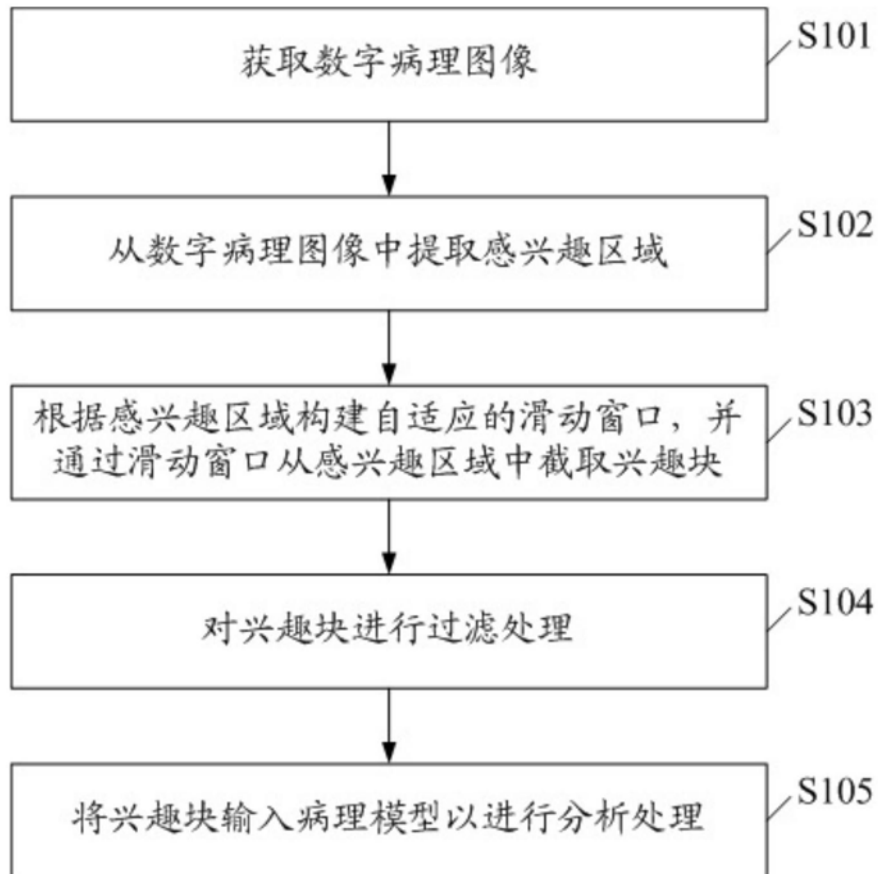


图1

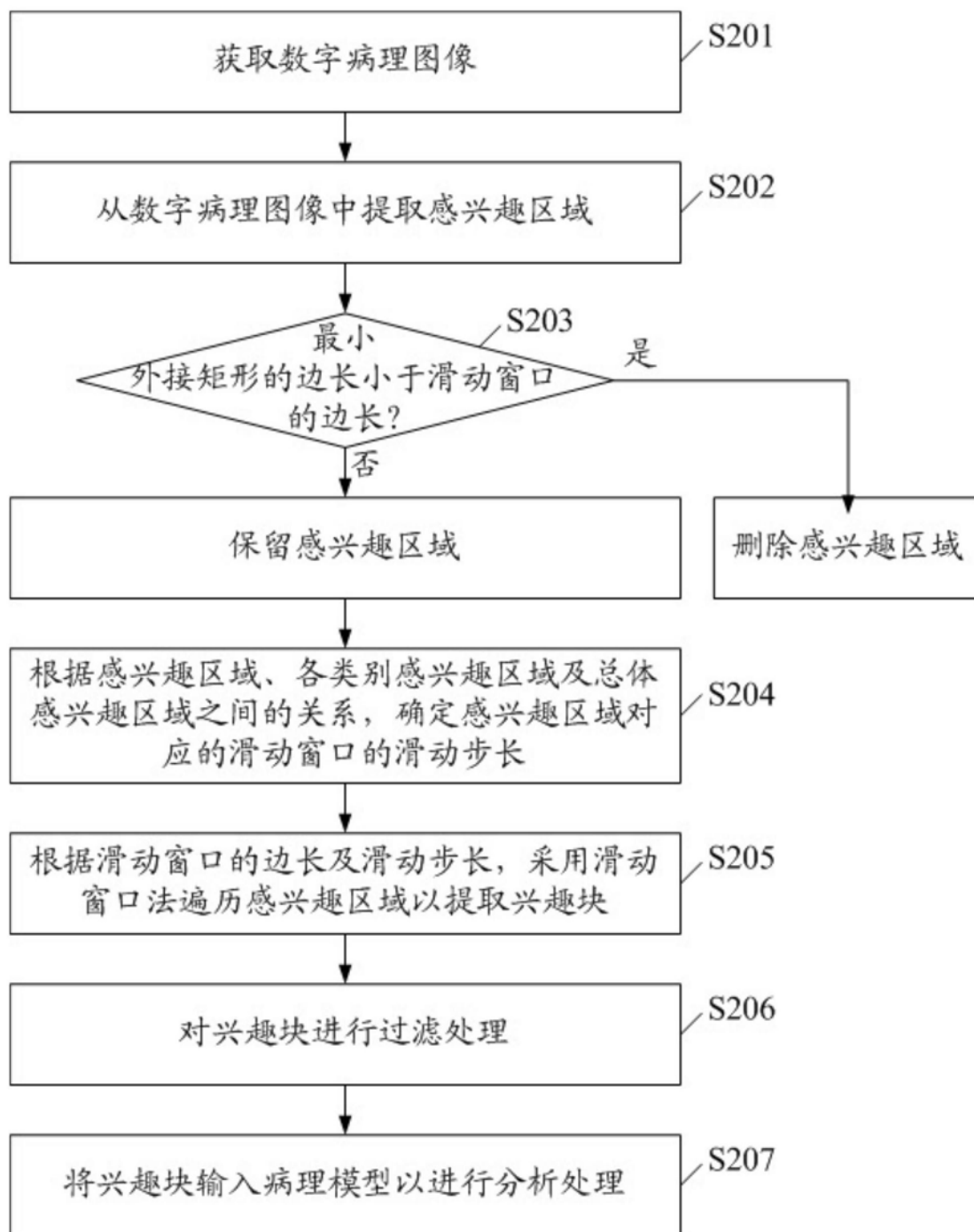


图2

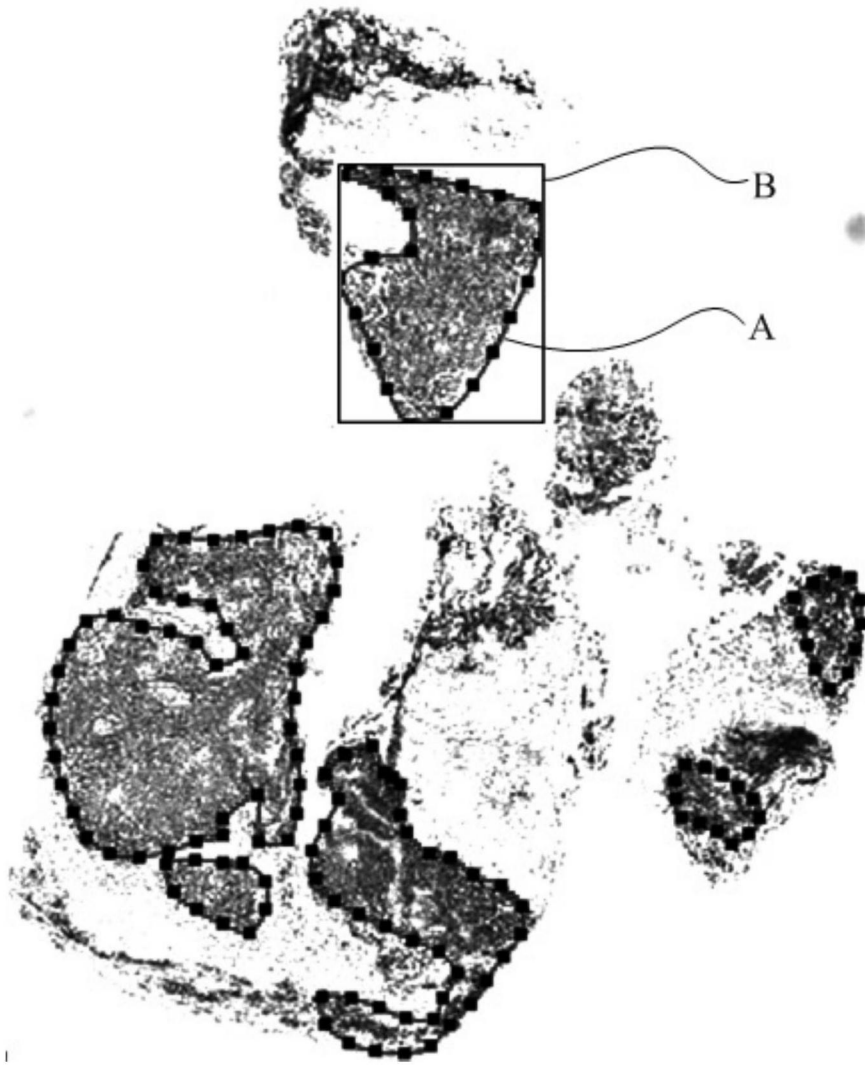


图3

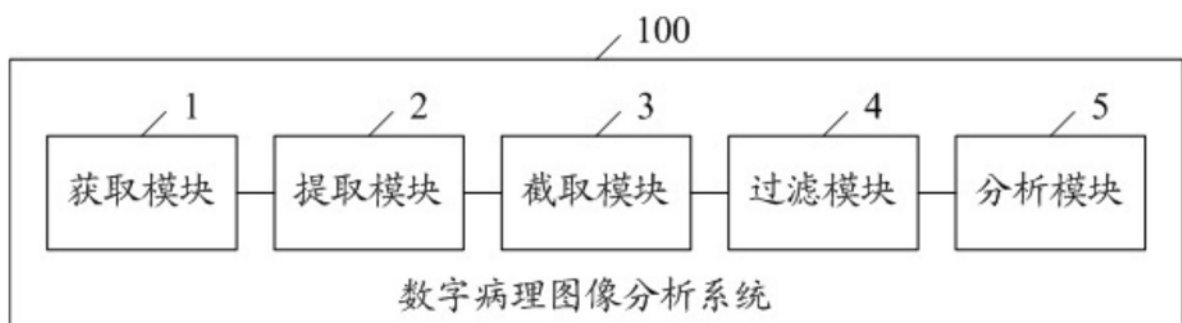


图4

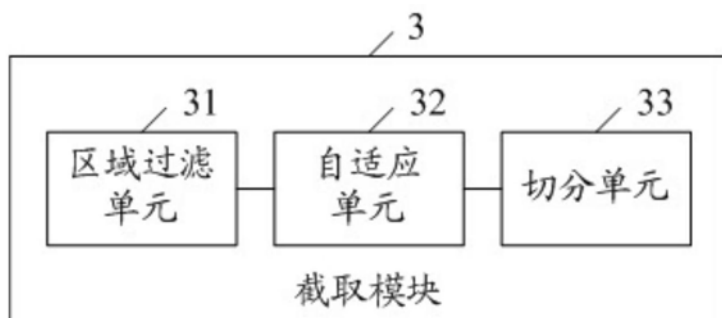


图5

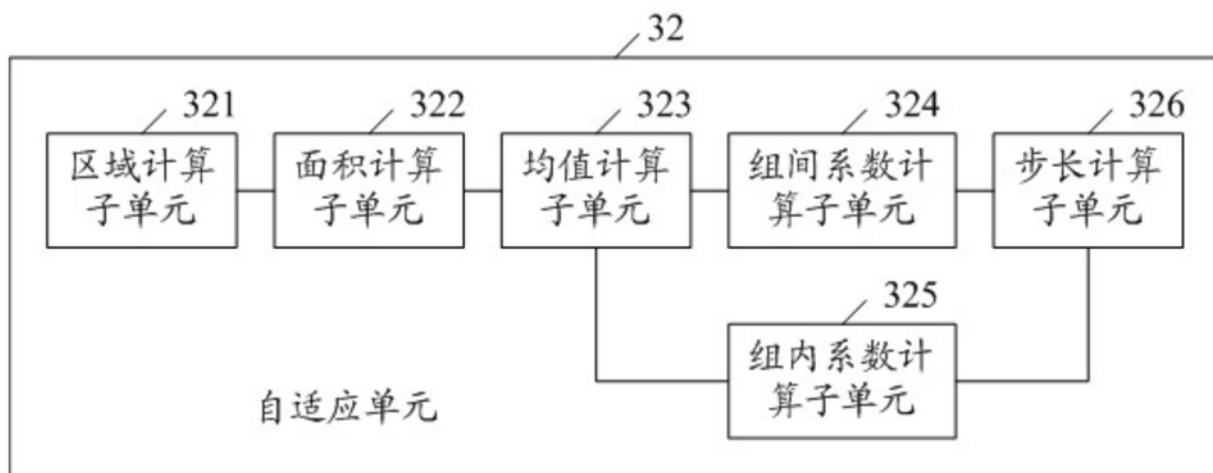


图6