



(21) 申请号 202410510512.6

(22) 申请日 2024.04.26

(65) 同一申请的已公布的文献号

申请公布号 CN 118096757 A

(43) 申请公布日 2024.05.28

(73) 专利权人 天津市肿瘤医院(天津医科大学
肿瘤医院)

地址 300060 天津市河西区体院北环湖西
路

(72) 发明人 郭晓静 刘再毅 贾玉棉 林佳泰
韩楚 刘芳芳 高广深 杨柳

(74) 专利代理机构 北京风雅颂专利代理有限公
司 11403

专利代理师 孙晓凤

(51) Int.Cl.

G06T 7/00 (2017.01)

G06T 7/11 (2017.01)

G06T 7/30 (2017.01)

G06V 10/25 (2022.01)

G06V 10/26 (2022.01)

G06V 20/70 (2022.01)

(56) 对比文件

WO 2023183699 A2, 2023.09.28

CN 111602136 A, 2020.08.28

US 2020157501 A1, 2020.05.21

US 2020302603 A1, 2020.09.24

US 2022188976 A1, 2022.06.16

审查员 韩亚楠

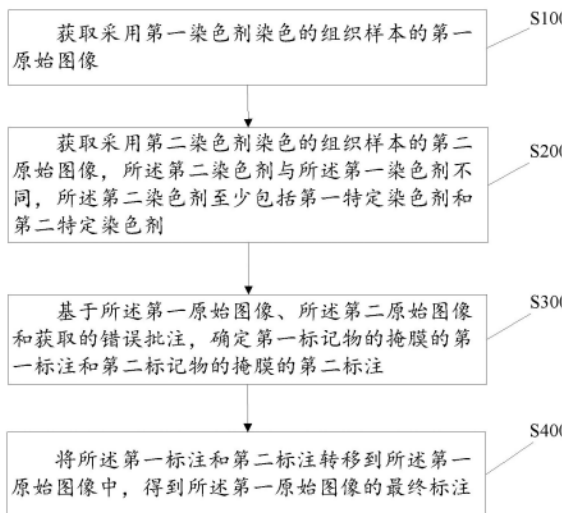
权利要求书2页 说明书12页 附图2页

(54) 发明名称

基于双染数字病理图像的图像标注方法和
图像标注设备

(57) 摘要

本申请提供一种基于双染数字病理图像的
图像标注方法和图像标注设备,所述方法包括获
取采用第一染色剂染色的组织样本的第一原始
图像;获取采用第二染色剂染色的组织样本的第
二原始图像;基于所述第一原始图像、所述第二
原始图像和获取的错误标注,确定第一标记物
的掩膜的第一标注和第二标记物的掩膜的第二
标注;将所述第一标注和第二标注转移到所述第
一原始图像中,得到所述第一原始图像的最终
标注。可以同时两种标记物进行分析,满足了实
际需求,又无需对同样的组织样本进行多次染
色、洗脱和重新染色的重复操作,提升了掩膜
标注的准确性。



1. 一种基于双染数字病理图像的图像标注方法,其特征在于,包括:

获取采用第一染色剂染色的组织样本的第一原始图像;

获取采用第二染色剂染色的组织样本的第二原始图像,所述第二染色剂与所述第一染色剂不同,所述第二染色剂至少包括第一特定染色剂和第二特定染色剂,所述第一特定染色剂用于标记第一标记物,所述第二特定染色剂用于标记第二标记物,所述第一标记物为细胞角蛋白CK,所述第二标记物为肌上皮细胞P63,所述第一特定染色剂用于标记CK,所述第二特定染色剂用于标记P63;

基于所述第一原始图像、所述第二原始图像和获取的错误标注,确定第一标记物的掩膜的第一标注和第二标记物的掩膜的第二标注;

将所述第一标注和所述第二标注转移到所述第一原始图像中,得到所述第一原始图像的最终标注,所述最终标注用于区分导管原位癌和微浸润性癌;

其中,所述基于所述第一原始图像、所述第二原始图像和获取的错误标注,确定第一标记物的掩膜的第一标注和第二标记物的掩膜的第二标注,包括:

基于所述第一原始图像和所述第二原始图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注;

获取错误标注,去除所述第一初始标注和第二初始标注中的所述错误标注,得到所述第一标注和所述第二标注;

其中,所述基于所述第一原始图像和所述第二原始图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注,包括:

将第一原始图像和第二原始图像进行第一次配准,得到第一配准图像和第二配准图像;

获取第一配准图像的全切片组织样本掩膜;

分别确定所述全切片组织样本掩膜和所述第二配准图像的兴趣区域;

基于所述兴趣区域,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注,包括:

将所述全切片组织样本掩膜的兴趣区域和所述第二配准图像的兴趣区域进行第二次配准,得到第三配准图像和第四配准图像;

基于所述第四配准图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注,包括:

获取所述第四配准图像的多通道灰度图,所述多通道灰度图至少包括第一特定染色剂通道灰度图和第二特定染色剂通道灰度图;

基于所述第一特定染色剂通道灰度图,确定第一标记物的掩膜的第一初始标注,包括:将所述第一特定染色剂通道灰度图输入训练后的第一模型中,得到第一标记物的掩膜的第一初始标注;

基于所述第二特定染色剂通道灰度图,确定第二标记物的掩膜的第二初始标注,包括:

基于所述第二特定染色剂通道灰度图,确定第二标记物的初始掩膜;

确定所述第四配准图像中所有细胞的边界,包括:将所述第四配准图像输入训练后的第二模型中,得到所述第四配准图像中所有细胞的边界;

基于所述初始掩膜和所述所有细胞的边界,将所述初始掩膜对应的细胞边界确定为第

二标记物的掩膜的第二初始标注。

2. 根据权利要求1所述的图像标注方法,其特征在于,所述分别确定所述全切片组织样本掩膜和所述第二配准图像的感兴趣区域,包括:

对所述全切片组织样本掩膜进行图片块裁剪得到多个第一图像块;

对所述第二配准图像进行图片块裁剪得到多个第二图像块;

确定所述多个第一图像块和多个第二图像块中具有相同坐标的相同图像块;

将所述相同图像块中的第一图像块的集合确定为所述全切片组织样本掩膜的感兴趣区域,将所述相同图像块中的第二图像块的集合确定为所述第二配准图像的感兴趣区域。

3. 根据权利要求1所述的图像标注方法,其特征在于,所述第一模型为上皮组织语义分割模型,所述第二模型为肌上皮细胞检测模型。

4. 一种图像标注设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,其特征在于,所述处理器执行所述程序时实现如权利要求1至3任意一项所述的方法。

基于双染数字病理图像的图像标注方法和图像标注设备

技术领域

[0001] 本申请涉及数字病理学技术领域,尤其涉及一种基于双染数字病理图像的图像标注方法和图像标注设备。

背景技术

[0002] 在对数字病理图像进行处理的过程中,通常会利用洗脱重染的技术获取HE染色和IHC染色的图像,但是,重新进行的IHC染色通常是单免疫组化染色。基于单免疫组化染色的数字病理图像仅能得到一个生物标记物的掩膜,无法同时得到两个或多个生物标记物的掩膜,如此不能满足实际需求。

发明内容

[0003] 有鉴于此,本申请的目的在于提出一种基于双染数字病理图像的图像标注方法和图像标注设备。

[0004] 基于上述目的,本申请第一方面提供了一种基于双染数字病理图像的图像标注方法,包括:

[0005] 获取采用第一染色剂染色的组织样本的第一原始图像;

[0006] 获取采用第二染色剂染色的组织样本的第二原始图像,所述第二染色剂与所述第一染色剂不同,所述第二染色剂至少包括第一特定染色剂和第二特定染色剂,所述第一特定染色剂用于标记第一标记物,所述第二特定染色剂用于标记第二标记物;

[0007] 基于所述第一原始图像、所述第二原始图像和获取的错误标注,确定第一标记物的掩膜的第一标注和第二标记物的掩膜的第二标注;

[0008] 将所述第一标注和第二标注转移到所述第一原始图像中,得到所述第一原始图像的最终标注。

[0009] 可选地,所述基于所述第一原始图像、所述第二原始图像和获取的错误标注,确定第一标记物的掩膜的第一标注和第二标记物的掩膜的第二标注,包括:

[0010] 基于所述第一原始图像和所述第二原始图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注;

[0011] 获取错误标注,去除所述第一初始标注和第二初始标注中的错误标注,得到所述第一标注和所述第二标注。

[0012] 可选地,所述基于所述第一原始图像和所述第二原始图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注,包括:

[0013] 将第一原始图像和第二原始图像进行第一次配准,得到第一配准图像和第二配准图像;

[0014] 获取第一配准图像的全切片组织样本掩膜;

[0015] 分别确定所述全切片组织样本掩膜和所述第二配准图像的感兴趣区域;

[0016] 基于所述感兴趣区域,确定第一标记物的掩膜的第一初始标注和第二标记物的掩

膜的第二初始标注。

[0017] 可选地,所述分别确定所述全切片组织样本掩膜和所述第二配准图像的感兴趣区域,包括:

[0018] 对所述全切片组织样本掩膜进行图片块裁剪得到多个第一图像块;

[0019] 对所述第二配准图像进行图片块裁剪得到多个第二图像块;

[0020] 确定所述多个第一图像块和多个第二图像块中具有相同坐标的相同图像块;

[0021] 将所述相同图像块中的第一图像块的集合确定为所述全切片组织样本掩膜的感兴趣区域,将所述相同图像块中的第二图像块的集合确定为所述第二配准图像的感兴趣区域。

[0022] 可选地,所述基于所述感兴趣区域,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注,包括:

[0023] 将所述全切片组织样本掩膜的感兴趣区域和所述第二配准图像的感兴趣区域进行第二次配准,得到第三配准图像和第四配准图像;

[0024] 基于所述第四配准图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注。

[0025] 可选地,所述基于所述第四配准图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注,包括:

[0026] 获取所述第四配准图像的多通道灰度图,所述多通道灰度图至少包括第一特定染色剂通道灰度图和第二特定染色剂通道灰度图;

[0027] 基于所述第一特定染色剂通道灰度图,确定第一标记物的掩膜的第一初始标注;

[0028] 基于所述第二特定染色剂通道灰度图,确定第二标记物的掩膜的第二初始标注。

[0029] 可选地,所述基于所述第二特定染色剂通道灰度图,确定第二标记物的掩膜的第二初始标注,包括:

[0030] 基于所述第二特定染色剂通道灰度图,确定第二标记物的初始掩膜;

[0031] 确定所述第四配准图像中所有细胞的边界;

[0032] 基于所述初始掩膜和所述所有细胞的边界,将所述初始掩膜对应的细胞边界确定为第二标记物的掩膜的第二初始标注。

[0033] 可选地,所述基于所述第一特定染色剂通道灰度图,确定第一标记物的掩膜的第一初始标注,包括:将所述第一特定染色剂通道灰度图输入训练后的第一模型中,得到第一标记物的掩膜的第一初始标注;

[0034] 所述确定所述第四配准图像中所有细胞的边界,包括:将所述第四配准图像输入训练后的第二模型中,得到所述第四配准图像中所有细胞的边界。

[0035] 可选地,所述第一标记物为细胞角蛋白CK,所述第二标记物为肌上皮细胞P63,所述第一模型为上皮组织语义分割模型,所述第二模型为肌上皮细胞检测模型。

[0036] 基于同一发明构思,本申请第二方面提供了一种图像标注设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现如上述第一方面任意一项所述的方法。

[0037] 从上面所述可以看出,本申请提供的基于双染数字病理图像的图像标注方法和图像标注设备,基于采用第一染色剂染色的组织样本的第一原始图像和采用第二染色剂染色

的组织样本的第二原始图像以及获取的错误标注,可以得到第一标记物的掩膜的第一标注和第二标记物掩膜的第二标注,并将第一标注和第二标注转移到第一原始图像中得到最终标注,即仅通过一次组织样本的染色、洗脱和重新染色就可以在第二原始图像中得到包括两种标记物掩膜的标注,既可以同时对两种标记物进行分析,满足了实际需求,又无需对同样的组织样本进行多次染色、洗脱和重新染色的重复操作,提升了掩膜标注的准确性;同时在确定第一标注和第二标注时引入了错误标注,可以降低全自动标注造成的噪声问题,进一步提升标注的准确性。

附图说明

[0038] 为了更清楚地说明本申请或相关技术中的技术方案,下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0039] 图1为本申请实施例的图像标注方法的第一种流程示意图;

[0040] 图2为本申请实施例的图像标注方法的第二种流程示意图;

[0041] 图3为本申请实施例的图像标注设备的示意图。

具体实施方式

[0042] 为使本申请的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本申请进一步详细说明。

[0043] 需要说明的是,除非另外定义,本申请实施例使用的技术术语或者科学术语应当为本申请所属领域内具有一般技能的人士所理解的通常意义。本申请实施例中使用的“第一”、“第二”以及类似的词语并不表示任何顺序、数量或者重要性,而只是用来区分不同的组成部分。“包括”或者“包含”等类似的词语意指出现该词前面的元件或者物件涵盖出现在该词后面列举的元件或者物件及其等同,而不排除其他元件或者物件。“连接”或者“相连”等类似的词语并非限定于物理的或者机械的连接,而是可以包括电性的连接,不管是直接的还是间接的。“上”、“下”、“左”、“右”等仅用于表示相对位置关系,当被描述对象的绝对位置改变后,则该相对位置关系也可能相应地改变。

[0044] 需要说明的是,除非另外定义,以下实施例中所用的技术术语具有与本发明所属领域技术人员普遍理解的相同含义。以下实施例中所用的试验试剂,如无特殊说明,均为常规生化试剂;所述实验方法,如无特殊说明,均为常规方法。

[0045] 术语“掩膜”是指组织样本图像中包围感兴趣区域(诸如肿瘤细胞)的闭合多边形区域。术语“基准真值”是指通过直接观察样本来提供掩膜以及可以分配给组织样本的标签。术语“HE数字病理图像”或“HE图像”是指HE染色切片经数字扫描仪扫描后生成的数字图像。术语“IHC数字病理图像”或“IHC图像”是指免疫组化染色切片经数字扫描仪扫描后生成的数字图像。术语“单染IHC数字病理图像”是指使用单一免疫组织化学染色剂染色切片经数字扫描仪扫描后生成的数字图像。术语“CK/P63双染IHC数字病理图像”是指CK/P63双重免疫组化染色切片经数字扫描仪扫描后生成的数字图像。

[0046] 其中,HE指的是苏木精和曙红,是非特定染色剂。IHC指的是免疫组织化学染色剂,

是特定染色剂。CK指的是细胞角蛋白,主要存在于乳腺和肺的上皮细胞中,可以作为上皮细胞的特定标记物。P63指的是肌上皮细胞里面具有特高特异性以及敏感性的肿瘤标志物,可以作为肌上皮细胞的特定标记物。

[0047] 组织样本的数字图像被用于多种情景中,例如可以作为用于构建机器学习模型的训练示例,也可以帮助诊断,为临床决策提供支持以及对提供组织样本的患者做出预测,诸如预测生存或对治疗的反应。不管在什么情景下,从组织样本的数字图像中生成精确的基准真值掩膜(ground truth mask)是非常关键的环节,基准真值掩膜的准确与否关系到机器学习模型训练的准确性以及为临床提供支撑或预测的准确性等。

[0048] 若采用人工创建准确的基准真值掩膜并且分配标签需要花费病例学家大量的时间,工作量巨大。因此现有在数字病理学领域引入了机器学习模型来得到基准真值掩膜。但是,若完全使用机器学习模型来得到基准真值掩膜,则可能会因病理图像中一些未知的染色问题(如染色污渍问题)导致最终得到的掩膜存在噪声。

[0049] 在当今的病理学中,从组织样本中得到基准真值掩膜,并基于基准真值掩膜进行癌症分期和诊断通常是在HE染色的组织样本上进行的。HE染色剂是非特定染色剂,其突出组织的整体形态。与之相反,存在特定染色剂(包括免疫组织化学染色剂IHC、细胞核染色剂或与组织类型或讨论的可能疾病状态相关的一些其他特定染色剂等),其突出特定抗原,可以作为肿瘤标志物。

[0050] 病理学家通常可以在HE染色的病理图像上描绘肿瘤轮廓以支持和帮助诊断。但是,有时对于疑难病例,直接在HE染色的病理图像上很难得到精确地肿瘤轮廓,此时就需要特定染色剂(例如IHC染色剂)对组织样本进行染色,得到的IHC染色的病理图像对细胞和细胞核形态具有更具体的洞察力,并且由于染色引起的感兴趣区域的对比度更高,因此更易于得到更为精确的肿瘤轮廓,最终得到更为精确的基准真值掩膜。

[0051] 乳腺微浸润性癌(Micro invasive carcinoma)的定义为浸润性癌病灶 $\leq 1\text{mm}$ 的乳腺癌。研究表明,微浸润性癌比导管原位癌有更强的侵袭性,预后与小灶的浸润性癌相似。因此,准确诊断微浸润性癌,以与导管原位癌进行区分,对于准确预测患者预后及临床治疗方案的制定具有重要临床意义。

[0052] 面对如何将微浸润性癌与导管原位癌进行有效区分时,发明人发现,可以使用现有的通过对单染IHC数字病理图像进行标注来得到微浸润性癌及导管原位癌的掩膜,进而对二者进行区分。但是,在实践中发现,微浸润性癌与导管原位癌在单染IHC数字病理图像中很难进行明确区分,以致很难将微浸润性癌与导管原位癌进行有效区分。这是因为,导管原位癌与微浸润性癌的区别仅在于导管原位癌周围有肌上皮包绕,微浸润性癌周围肌上皮缺失,这种较小的差别在单染IHC数字病理图像中很难被精确标注出来。

[0053] 基于此,发明人发现,若能在数字病理图像中分别对肌上皮细胞和上皮细胞(即肿瘤细胞)进行标注,就可以明显地区分导管原位癌与微浸润性癌。因此,发明人尝试对同样的组织样本的单染IHC数字病理图像依次进行肌上皮细胞标注和上皮细胞(即肿瘤细胞)标注,但是,这种依次标注的方式需要对同样的组织样本进行多次染色、洗脱和重新染色的操作,需要耗费大量的时间,并且经过多次染色、洗脱和重新染色之后的组织样本可能会发生形变,导致多次染色得到的掩膜并不准确。

[0054] 基于此,参见图1,在一些实施例中,本申请提供了一种基于双染数字病理图像的

图像标注方法,具体包括如下步骤:

[0055] 步骤S100、获取采用第一染色剂染色的组织样本的第一原始图像;

[0056] 步骤S200、获取采用第二染色剂染色的组织样本的第二原始图像,所述第二染色剂与所述第一染色剂不同,所述第二染色剂至少包括第一特定染色剂和第二特定染色剂,所述第一特定染色剂用于标记第一标记物,所述第二特定染色剂用于标记第二标记物;

[0057] 步骤S300、基于所述第一原始图像、所述第二原始图像和获取的错误标注,确定第一标记物的掩膜的第一标注和第二标记物的掩膜的第二标注;

[0058] 步骤S400、将所述第一标注和第二标注转移到所述第一原始图像中,得到所述第一原始图像的最终标注。

[0059] 具体地,所述第一染色剂可以为苏木精和曙红(即HE),所述第一原始图像可以为HE数字病理图像。

[0060] 所述步骤S200获取采用第二染色剂染色的组织样本的第二原始图像具体包括:清洗步骤S100中的组织样本中的第一染色剂,然后采用第二染色剂对该清洗后的组织样本进行染色,将第二染色剂染色后的组织样本切片经数字扫描仪扫描后生成所述第二原始图像。

[0061] 所述第二染色剂可以为与第一染色剂不同的特定染色剂,例如所述第二染色剂可以为免疫组织化学染色剂中的至少两种,例如可以为用于染色前列腺癌PIN4的染色剂、或用于染色淋巴结上皮细胞转移的细胞角蛋白AE1/AE3的染色剂、用于染色CK的染色剂、用于染色P63的染色剂中的至少两种,其可以作为肿瘤标志物用于对特定的肿瘤细胞进行染色和标记。所述第二原始图像可以为IHC数字病理图像。

[0062] 所述第二染色剂至少包括第一特定染色剂和第二特定染色剂,所述第一特定染色剂用于染色和标记第一标记物,所述第二特定染色剂用于染色和标记第二标记物,使得本申请中,通过第一特定染色剂和第二特定染色剂对组织样本进行染色,得到双染色的组织样本,进而使得后续可以对组织样本中的第一标记物和第二标记物的掩膜进行标注,提高标注精度,减少标注时间。

[0063] 在得到第一原始图像和第二原始图像之后,获取错误标注。所述错误标注可以为获取到的人工确定的错误标注。例如,可以为获取到的人工通过触控组件(例如鼠标和键盘等)确定的错误标注。

[0064] 然后基于所述第一原始图像、所述第二原始图像和获取的错误标注,确定第一标记物的掩膜的第一标注和第二标记物的掩膜的第二标注。如此,确定的第一标注和第二标注中已经剔除了人工确定的错误标注,可以降低全自动标注造成的噪声问题,提升标注的准确性。

[0065] 最后,将所述第一标注和第二标注转移到所述第一原始图像中,得到所述第一原始图像的最终标注,仅通过一次组织样本的染色、洗脱和重新染色就可以在所述第一原始图像中得到同时包括两种标记物掩膜的标注,既可以同时对两种标记物进行分析,满足了实际需求,又无需对同样的组织样本进行多次染色、洗脱和重新染色的重复操作,提升了掩膜标注的准确性。

[0066] 具体实施时,所述第一特定染色剂可以为用于对CK染色的染色剂,例如棕染DAB(3,3-二氨基联苯胺)显色液,用于标记第一标记物CK,最终得到的CK的标注即为上皮细胞

(即肿瘤细胞)的标注。

[0067] 所述第二特定染色剂可以为用于对P63染色的染色剂,例如红染DAB显色液,用于标记第二标记物P63,最终得到的P63的标注即为肌上皮细胞的标注。

[0068] 所述第一染色剂可以为苏木精和曙红(即HE)。

[0069] 所述图像标注方法具体包括:

[0070] 获取采用HE染色的组织样本的HE数字病理图像;

[0071] 获取采用红染DAB显色液和棕染DAB显色液染色的组织样本的IHC数字病理图像;

[0072] 基于所述HE数字病理图像、IHC数字病理图像和获取的错误标注,确定上皮细胞的掩膜的第一标注和肌上皮细胞的掩膜的第二标注;

[0073] 将所述第一标注和第二标注转移到所述HE数字病理图像中,得到所述HE数字病理图像的最终标注。

[0074] 通过借助CK和p63双重免疫组化染色,在HE数字病理图像上同时标注上皮细胞(肿瘤细胞)和肌上皮细胞,判定癌巢周围是否存在肌上皮细胞,当确定癌巢周围存在肌上皮细胞,则可确定为导管原位癌;当确定癌巢周围不存在肌上皮细胞,则可确定为微浸润性癌,如此仅通过一次HE数字病理图像标注就可以精确区分导管原位癌和微浸润性癌,可以为临床提供准确的支撑或预测,也可以为机器学习模型训练提供精确的样本集。

[0075] 并且,本图像标注方法通过程序自动化的方式完成大部分初始标注工作,并引入获取到的错误标注来对初始标注进行筛选,删选后的标注即为最终标注,最终标注可以作为基准真值掩膜。

[0076] 在一些实施例中,所述步骤S300基于所述第一原始图像、所述第二原始图像和获取的错误标注,确定第一标记物的掩膜的第一标注和第二标记物的掩膜的第二标注,包括:

[0077] 步骤S310、基于所述第一原始图像和所述第二原始图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注;

[0078] 步骤S320、获取错误标注,去除所述第一初始标注和第二初始标注中的错误标注,得到所述第一标注和所述第二标注。

[0079] 具体地,所述错误标注可以为获取到的人工确定的错误标注。

[0080] 基于错误标注,去除所述第一初始标注和第二初始标注中的错误标注,得到所述第一标注和所述第二标注,如此使得最终确定的第一标注和第二标注中已经剔除了人工确定的错误标注,可以降低完全自动标注造成的噪声问题,提升标注的准确性。

[0081] 本申请中,通过程序自动化的方式得到第一初始标注和第二初始标注,然后引入获取到的错误标注来对初始标注进行筛选,删选后的标注即为最终的第一标注和第二标注,可以提升第一标注和第二标注的准确性。

[0082] 在一些实施例中,所述步骤S310基于所述第一原始图像和所述第二原始图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注,包括:

[0083] 步骤S311、将第一原始图像和第二原始图像进行第一次配准,得到第一配准图像和第二配准图像;

[0084] 步骤S312、获取第一配准图像的全切片组织样本掩膜;

[0085] 步骤S313、分别确定所述全切片组织样本掩膜和所述第二配准图像的感兴趣区域;

[0086] 步骤S314、基于所述感兴趣区域,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注。

[0087] 具体地,所述步骤S311将第一原始图像和第二原始图像进行第一次配准,具体可以以任何方便的方式执行配准,例如使用公知的图像处理技术来确定图像中各种关键特征(诸如高对比度区域、拐角、边界等)的X/Y坐标,使得图像中的每个关键特征的X/Y像素位置可以与另一图像中相同关键特征的X/Y像素位置相关或匹配。示例性地,可以使用SIFT(特征提取方法)来确定具有锐度或颜色分布梯度的区域,以识别图像中的关键特征或位置并且确定匹配特征向量。

[0088] 将第一原始图像和第二原始图像进行第一次配准,以解决第一原始图像和第二原始图像的图像大小及位置不一致、图像变形等问题。

[0089] 所述步骤S312获取第一配准图像的全切片组织样本掩膜,具体为在低倍率(例如1x或10x)下,将所述第一配准图像灰度化并通过阈值分割来提取全切片组织样本掩膜。

[0090] 步骤S313分别确定所述全切片组织样本掩膜和所述第二配准图像的感兴趣区域,具体包括:

[0091] 步骤S3131、对所述全切片组织样本掩膜进行图片块裁剪得到多个第一图像块;

[0092] 步骤S3132、对所述第二配准图像进行图片块裁剪得到多个第二图像块;

[0093] 步骤S3133、确定所述多个第一图像块和多个第二图像块中具有相同坐标的相同图像块;

[0094] 步骤S3134、将所述相同图像块中的第一图像块的集合确定为所述全切片组织样本掩膜的感兴趣区域,将所述相同图像块中的第二图像块的集合确定为所述第二配准图像的感兴趣区域。

[0095] 具体地,在高倍率(例如40x或50x)下,通过滑动窗口的方式对全切片组织样本掩膜进行图片块裁剪得到多个第一图像块,对所述第二配准图像进行图片块裁剪得到多个第二图像块。

[0096] 然后,确定所述多个第一图像块和多个第二图像块中具有相同坐标的相同图像块,将所述相同图像块中的第一图像块的集合确定为所述全切片组织样本掩膜的感兴趣区域,将所述相同图像块中的第二图像块的集合确定为所述第二配准图像的感兴趣区域。

[0097] 示例性地,对所述全切片组织样本掩膜进行图片块裁剪得到多个第一图像块分别为A、B、C、D、E;对所述第二配准图像进行图片块裁剪得到多个第二图像块分别为Q、W、E、R、T。

[0098] 基于这多个第一图像块和多个第二图像块,确定所述多个第一图像块和多个第二图像块中具有相同坐标的相同图像块分别为A和W,C和R,因此将相同图像块中的第一图像块A和C的集合确定为所述全切片组织样本掩膜的感兴趣区域,将相同图像块中的第二图像块W和R的集合确定为所述第二配准图像的感兴趣区域。

[0099] 本申请中,通过图片块裁剪和确定相同图像块的方式得到全切片组织样本掩膜的感兴趣区域和第二配准图像的感兴趣区域,使得第二配准图像的感兴趣区域的每一个图像块均能在全切片组织样本掩膜的感兴趣区域中找到相同坐标的图像块,确保第二配准图像和全切片组织样本掩膜的感兴趣区域基本相同,提升了确定的感兴趣区域的准确性。

[0100] 然后,将所述全切片组织样本掩膜的感兴趣区域和所述第二配准图像的感兴趣区

域进行第二次配准,得到第三配准图像和第四配准图像。所述配准步骤可以以任何方便的方式执行配准,具体配准方式在此不做限定。

[0101] 通过将所述全切片组织样本掩膜的感兴趣区域和所述第二配准图像的感兴趣区域进行第二次配准,实现感兴趣区域在像素级别的位置对齐,确保第二次配准后的第三配准图像和第四配准图像可以完全配准,如此将所述第一标注和第二标注转移到所述第一原始图像后得到的最终标注才能是非常准确的。

[0102] 本申请中,基于图片块裁剪和确定相同图像块的方式得到全切片组织样本掩膜的感兴趣区域和第二配准图像的感兴趣区域,并再次对感兴趣区域进行配准,可以提升配准的精确性,进而确保最终得到的第一原始图像的最终标注的准确性。

[0103] 在一些实施例中,基于所述第四配准图像,确定第一标记物的掩膜的第一初始标注和第二标记物的掩膜的第二初始标注,具体包括:

[0104] 获取所述第四配准图像的多通道灰度图,所述多通道灰度图至少包括第一特定染色剂通道灰度图和第二特定染色剂通道灰度图;

[0105] 基于所述第一特定染色剂通道灰度图,确定第一标记物的掩膜的第一初始标注;

[0106] 基于所述第二特定染色剂通道灰度图,确定第二标记物的掩膜的第二初始标注。

[0107] 具体地,所述获取所述第四配准图像的多通道灰度图包括:将所述第四配准图像送入颜色反卷积模型分离出多通道灰度图,所述多通道灰度图至少包括第一特定染色剂通道灰度图和第二特定染色剂通道灰度图。

[0108] 然后,基于所述第一特定染色剂通道灰度图,确定第一标记物的掩膜的第一初始标注,具体可以将所述第一特定染色剂通道灰度图输入训练后的第一模型中,得到第一标记物的掩膜的第一初始标注。示例性地,所述第一特定染色剂可以为棕染DAB显色液,所述第一标记物可以为CK,所述第一模型为上皮组织语义分割模型。

[0109] 然后,基于所述第二特定染色剂通道灰度图,确定第二标记物的掩膜的第二初始标注,包括:基于所述第二特定染色剂通道灰度图,确定第二标记物的初始掩膜,具体可以通过第二特定染色剂通道灰度图的阈值分割得到所述第二标记物的初始掩膜。然后确定所述第四配准图像中所有细胞的边界,具体可以将所述第四配准图像输入训练后的第二模型中,得到所述第四配准图像中所有细胞的边界。最后基于所述初始掩膜和所述所有细胞的边界,将所述初始掩膜对应的细胞边界确定为第二标记物的掩膜的第二初始标注。示例性地,所述第二特定染色剂可以为红染DAB显色液,所述第二标记物可以为P63,所述第二模型可以为肌上皮细胞检测模型。

[0110] 本申请中,仅基于所述第一特定染色剂通道灰度图就可以确定第一标记物的掩膜的第一初始标注,这是因为,第一方面,使用第一特定染色剂对第一标记物进行标记是非常成熟且可靠的方式,可以得到非常精确的标注;第二方面,不管是导管原位癌还是微浸润性癌,其第一标记物的掩膜是肯定存在的,而且二者的第一标记物的掩膜没有太大差别,因此仅使用第一特定染色剂对第一标记物进行标记以得到第一标记物的掩膜及第一初始标注就可以满足需求,无需进行更为精细地标注。

[0111] 本申请中,基于第二特定染色剂通道灰度图确定第二标记物的初始掩膜,然后基于初始掩膜和所述所有细胞的边界来确定第二标记物的掩膜的第二初始标注,如此可以基于掩膜和细胞边界的双重方式来精确地确定第二标记物的掩膜及其第二初始标注,确保第

二初始标注的准确性,以准确地区分导管原位癌和微浸润性癌。这是因为,第二标记物的掩膜的第二次初始标注的存在与否、区域大小等是区分导管原位癌还是微浸润性癌的关键因素,因此必须非常精确地确定第二标记物的掩膜及其第二次初始标注,才能精确地区分导管原位癌和微浸润性癌。

[0112] 在一些实施例中,所述方法还包括:将所述第一原始图像的最终标注作为基准真值掩膜样本集,基于该基准真值掩膜样本集分别对第一模型和第二模型进行多次训练,以得到训练后的第一模型和第二模型。示例性地,所述第一模型可以为上皮组织语义分割模型,所述第二模型可以为肌上皮细胞检测模型。

[0113] 通过将第一原始图像的最终标注作为基准真值掩膜样本集对第一模型和第二模型进行训练,以对预训练好的模型进行微调,然后再以下一批基准真值掩膜样本集再次对模型进行训练,如此反复执行,以得到准确度符合要求的训练好的模型。

[0114] 值得注意的是,本申请中所涉及得到模型均为现有的机器学习模型,例如本申请所述涉及到的模型均可以为已经预训练过的卷积神经网络模型。

[0115] 在一些实施例中,以所述第一特定染色剂为棕染DAB显色液,所述第二特定染色剂为红染DAB显色液,所述第一标记物为CK,所述第二标记物为P63,所述第一染色剂可以为苏木精和曙红为例来进一步描述本申请所述的基于双染数字病理图像的图像标注方法。

[0116] 参见图2,所述的基于双染数字病理图像的图像标注方法,具体包括如下步骤:

[0117] 步骤1:HE染色及扫描

[0118] (1)切片及烤片:石蜡组织块常规切片,厚度为4 μ m,正电荷防脱玻片捞片,65 $^{\circ}$ C烤片60min。

[0119] (2)HE染色:切片二甲苯1、11常规脱蜡各10min,梯度酒精至水化,稳定前液2min,苏木素浸染15min,流水漂洗,分化30秒,返蓝2min,浸染伊红1min,流水漂洗10秒,梯度乙醇脱水(80%乙醇,90%乙醇,95%乙醇 *2,100%乙醇 *2)各1min,二甲苯 *2透明各1min,中性树胶封片。

[0120] (3)切片扫描:HE染色切片于Leica Aperio GT450 扫描仪进行扫描存档,得到HE图像(即所述第一原始图像)。

[0121] 步骤2:洗脱

[0122] (1)HE染色切片去除盖玻片:扫描完毕的HE染色切片可加热至60 $^{\circ}$ C至盖玻片轻易取下,切片去除盖玻片后浸泡二甲苯10min,彻底去除切片上的中性树胶,切片经100%乙醇3次,每次10min、蒸馏水水化待用。

[0123] (2)HE褪色:将玻片置于75%酒精中,观察伊红褪色完全后取出冲洗,随后置于草酸参与配置的分化液中褪去苏木色素,流水冲洗后等待上机。

[0124] 步骤3:CK/P63双重IHC染色及扫描

[0125] (1)免疫组化检测:检测平台为BOND全自动免疫组化和原位杂交染色系统,检测试剂盒为BOND Polymer Refine Red Detection显色试剂。褪色后的切片重新贴上免疫组化专用标签(CK-P63)。染色程序为:使用碱性清洗剂(例如BOND ER碱性清洗剂)碱性修复30min,过氧化物封闭液5min,一抗P63孵育30min,初步二抗10min,次级二抗10min,棕染DAB显色对细胞核定位;随后再次滴加一抗CK孵育15min,初步二抗碱性磷酸酶(简称AP)10min,次级二抗碱性磷酸酶10min,红染DAB对细胞质染色定位,苏木素复染6min,水洗,玻片架中

取出玻片晾干、中性树胶封片。

[0126] (2)再次将免疫组化染色片置于Leica Aperio GT450扫描仪上扫描存档,得到IHC图像(即所述第二原始图像)。

[0127] 步骤4:针对HE图像与IHC图像的全切片级别第一次配准。使用现有的配准技术把上述步骤准备好的HE图像、IHC图像进行初步的非刚性配准以解决HE染色和IHC染色的图像大小及位置不一致、图像变形等问题。

[0128] 步骤5:全切片病理组织掩膜提取及感兴趣区域(region of interest,ROI)裁剪。在低倍率下,把配准后的HE全切片图像(即所述第一配准图像)灰度化并通过阈值分割提取全切片病理组织掩膜(即所述全切片组织样本掩膜)。然后在高倍率下(40x)通过滑动窗口的方式对配对后的HE和IHC染色图像(即所述全切片组织样本掩膜和第二配准图像)进行图片块的裁剪及配对,相同坐标的同一配对的图像块确定为感兴趣区域(即ROI)。

[0129] 步骤6:针对配对的ROI进行第二次配准。为了保证HE染色与IHC染色的ROI图像能够共享相同的掩膜,对裁剪后的ROI进行二次配准,实现像素级别的位置对齐。

[0130] 步骤7:基于颜色反卷积的CK/P63通道分离。将IHC染色的ROI图像(即所述第四配准图像)送入颜色反卷积模型分离出CK/P63通道的灰度图(即所述多通道灰度图)。

[0131] 步骤8:利用CK/P63通道的上皮组织/肌上皮细胞伪标注自动获取:

[0132] (1)上皮组织伪标注:通过对CK通道的灰度图(即所述第一特定染色剂通道灰度图)进行阈值分割得到上皮组织掩膜作为伪标注(即所述第一标记物的掩膜的第一初始标注);

[0133] (2)肌上皮细胞伪标注:

[0134] 步骤a、通过P63通道灰度图(即所述第二特定染色剂通道灰度图)的阈值分割得到对应染色细胞的掩膜(即所述第二标记物的初始掩膜);

[0135] 步骤b、使用预训练好的细胞核分割模型算出所有细胞的边界;

[0136] 步骤c、使用步骤a的掩膜去击中/选择对应细胞的边界作为肌上皮细胞的伪标注(即所述第二标记物的掩膜的第二初始标注)。

[0137] 步骤9:伪标注的人工筛选,即获取错误标注,去除所述第一初始标注和第二初始标注中的错误标注,得到所述第一标注和所述第二标注。将所述第一标注和第二标注转移到所述第一原始图像中,得到所述第一原始图像的最终标注可以作为基准真值掩膜。

[0138] 还可以包括:把HE图像、IHC图像、CK/P63通道灰度图以及上皮组织掩膜/肌上皮细胞标注进行拼接并在屏幕上展示以供筛选。通过简单选择的方式把良好的伪标注纳入为基准真值掩膜分别可用于上皮组织语义分割模型训练以及肌上皮细胞检测模型训练。

[0139] 相较于传统的单免疫组化染色,本申请中,基于双重染色IHC图像的标注流程可以通过一张IHC图像获得两种生物标记物的掩膜,避免多次洗脱重新染色的需求;标注自动生成过程中,人工参与最后的标注把关既避免了全程手工标注的负担又保证用于模型训练的标注的正确性。

[0140] 需要说明的是,本申请实施例的方法可以由单个设备执行,例如一台计算机或服务器等。本实施例的方法也可以应用于分布式场景下,由多台设备相互配合来完成。在这种分布式场景的情况下,这多台设备中的一台设备可以只执行本申请实施例的方法中的某一个或多个步骤,这多台设备相互之间会进行交互以完成所述的方法。

[0141] 需要说明的是,上述对本申请的一些实施例进行了描述。在一些情况下,在上述实施例中记载的动作或步骤可以按照不同于上述实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可能是有利的。

[0142] 基于同一发明构思,与上述任意实施例方法相对应的,本申请还提供了一种图像标注设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现上任意一实施例所述的图像标注方法。

[0143] 图3示出了本实施例所提供的一种更为具体的图像标注设备硬件结构示意图,该设备可以包括:处理器1010、存储器1020、输入/输出接口1030、通信接口1040和总线1050。其中处理器1010、存储器1020、输入/输出接口1030和通信接口1040通过总线1050实现彼此之间在设备内部的通信连接。

[0144] 处理器1010可以采用通用的CPU(Central Processing Unit,中央处理器)、微处理器、应用专用集成电路(Application Specific Integrated Circuit,ASIC)、或者一个或多个集成电路等方式实现,用于执行相关程序,以实现本说明书实施例所提供的技术方案。

[0145] 存储器1020可以采用ROM(Read Only Memory,只读存储器)、RAM(Random Access Memory,随机存取存储器)、静态存储设备,动态存储设备等形式实现。存储器1020可以存储操作系统和其他应用程序,在通过软件或者固件来实现本说明书实施例所提供的技术方案时,相关的程序代码保存在存储器1020中,并由处理器1010来调用执行。

[0146] 输入/输出接口1030用于连接输入/输出模块,以实现信息输入及输出。输入输出/模块可以作为组件配置在设备中(图中未示出),也可以外接于设备以提供相应功能。其中输入设备可以包括键盘、鼠标、触摸屏、麦克风、各类传感器等,输出设备可以包括显示器、扬声器、振动器、指示灯等。

[0147] 通信接口1040用于连接通信模块(图中未示出),以实现本设备与其他设备的通信交互。其中通信模块可以通过有线方式(例如USB、网线等)实现通信,也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信。

[0148] 总线1050包括一通路,在设备的各个组件(例如处理器1010、存储器1020、输入/输出接口1030和通信接口1040)之间传输信息。

[0149] 需要说明的是,尽管上述设备仅示出了处理器1010、存储器1020、输入/输出接口1030、通信接口1040以及总线1050,但是在具体实施过程中,该设备还可以包括实现正常运行所必需的其他组件。此外,本领域的技术人员可以理解的是,上述设备中也可以仅包含实现本说明书实施例方案所必需的组件,而不必包含图中所示的全部组件。

[0150] 上述实施例的图像标注设备用于实现前述任一实施例中相应的图像标注方法,并且具有相应的方法实施例的有益效果,在此不再赘述。

[0151] 所属领域的普通技术人员应当理解:以上任何实施例的讨论仅为示例性的,并非旨在暗示本申请的范围被限于这些例子;在本申请的思路下,以上实施例或者不同实施例中的技术特征之间也可以进行组合,步骤可以以任意顺序实现,并存在如上所述的本申请实施例的不同方面的许多其它变化,为了简明它们没有在细节中提供。

[0152] 另外,为简化说明和讨论,并且为了不会使本申请实施例难以理解,在所提供的附

图中可以示出或不示出与集成电路 (IC) 芯片和其它部件的公知的电源/接地连接。此外,可以以框图的形式示出装置,以便避免使本申请实施例难以理解,并且这也考虑了以下事实,即关于这些框图装置的实施方式的细节是高度取决于将要实施本申请实施例的平台(即,这些细节应当完全处于本领域技术人员的理解范围内)。在阐述了具体细节(例如,电路)以描述本申请的示例性实施例的情况下,对本领域技术人员来说显而易见的是,可以在没有这些具体细节的情况下或者这些具体细节有变化的情况下实施本申请实施例。因此,这些描述应被认为是说明性的而不是限制性的。

[0153] 尽管已经结合了本申请的具体实施例对本申请进行了描述,但是根据前面的描述,这些实施例的很多替换、修改和变型对本领域普通技术人员来说将是显而易见的。例如,其它存储器架构(例如,动态RAM (DRAM))可以使用所讨论的实施例。

[0154] 本申请实施例旨在涵盖落入所附权利要求的宽泛范围之内的所有这样的替换、修改和变型。因此,凡在本申请实施例的精神和原则之内,所做的任何省略、修改、等同替换、改进等,均应包含在本申请的保护范围之内。

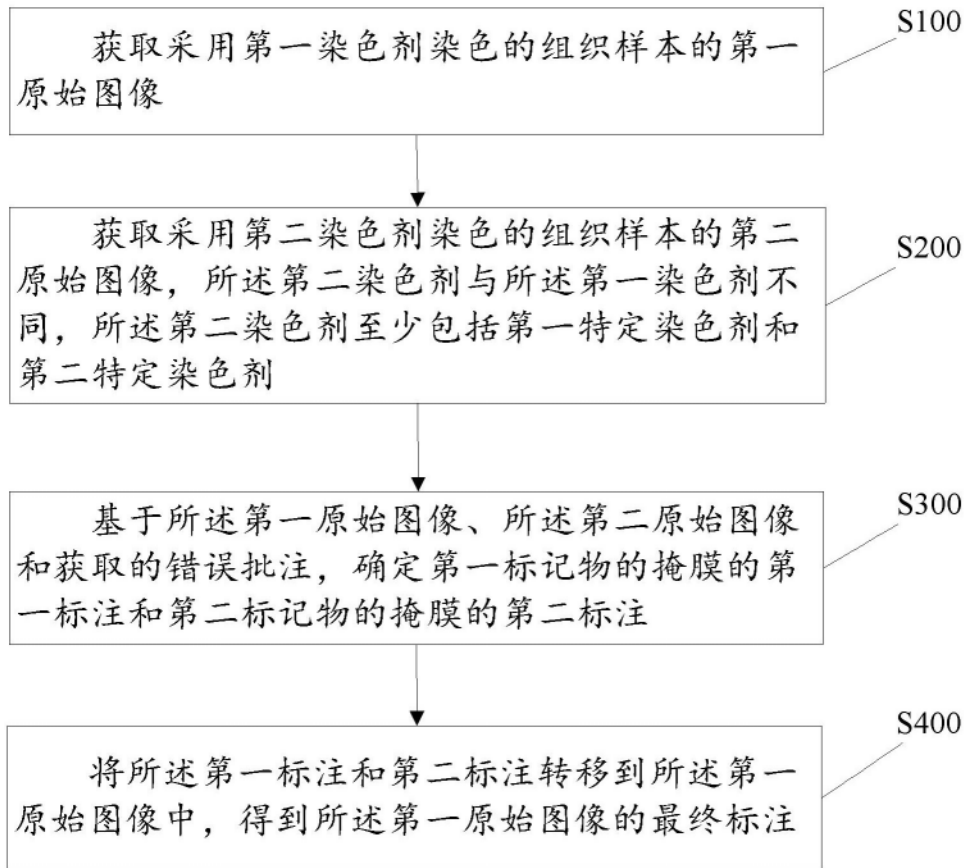


图1

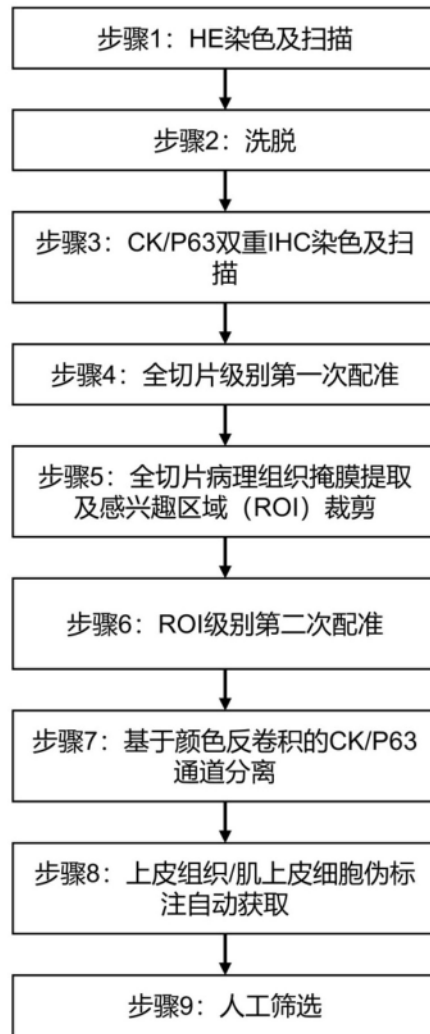


图2

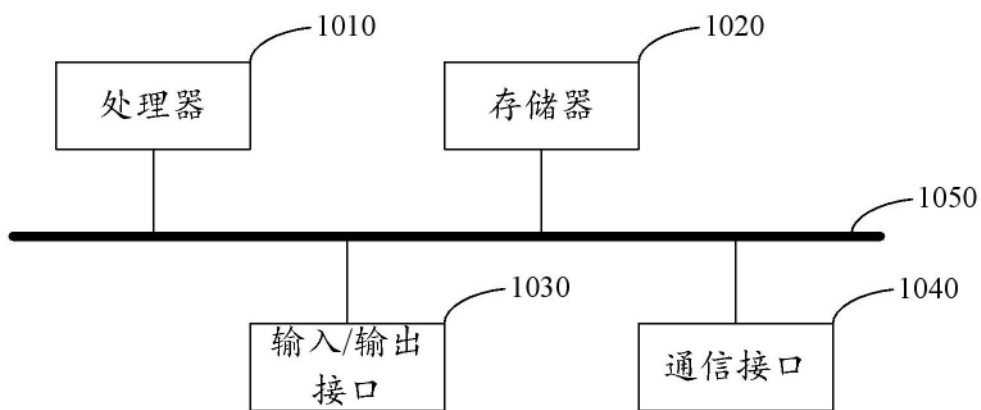


图3