



(12) 发明专利申请

(10) 申请公布号 CN 114332526 A

(43) 申请公布日 2022. 04. 12

(21) 申请号 202111333677.3

G06N 3/04 (2006.01)

(22) 申请日 2021.11.11

G06N 3/08 (2006.01)

G16H 30/00 (2018.01)

(71) 申请人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区  
科技中一路腾讯大厦35层

(72) 发明人 张闻华 张军 韩骁

(74) 专利代理机构 北京三高永信知识产权代理  
有限责任公司 11138

代理人 李文静

(51) Int.Cl.

G06V 10/764 (2022.01)

G06V 10/82 (2022.01)

G06V 10/74 (2022.01)

G06V 10/774 (2022.01)

G06K 9/62 (2022.01)

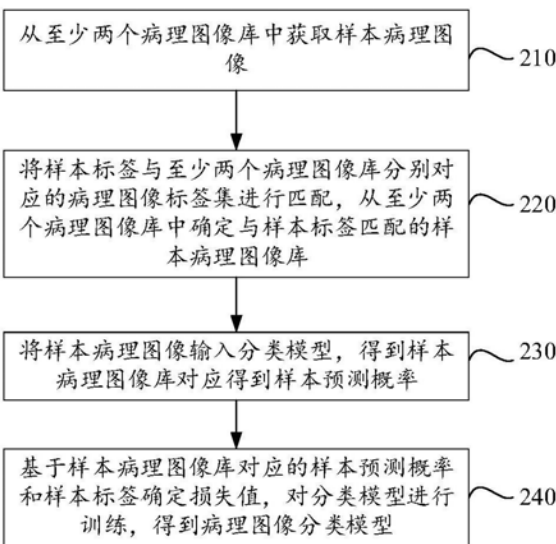
权利要求书3页 说明书18页 附图7页

(54) 发明名称

病理图像分类方法、装置、设备、存储介质及  
程序产品

(57) 摘要

本申请公开了一种病理图像分类方法、装置、设备、存储介质及程序产品,涉及机器学习领域。该方法包括:从病理图像库中获取样本病理图像,样本病理图像标注有样本标签;将样本标签与病理图像库分别对应的病理图像标签集进行匹配,确定样本病理图像库;将样本病理图像输入分类模型,得到样本预测概率;基于样本病理图像库对应的样本预测概率和样本标签确定损失值,对分类模型进行训练,得到病理图像分类模型,病理图像分类模型用于对目标病理图像进行分类,得到预测结果。通过以上方式,可以加强病理图像库之间的关联性,更好地实现集中化的病理图像分类,也便于更好地管理和使用病理图像。本申请可应用于云技术、人工智能、智慧交通等各种场景。



1. 一种病理图像分类方法,其特征在于,所述方法包括:

从至少两个病理图像库中获取样本病理图像,所述样本病理图像标注有样本标签,所述至少两个病理图像库分别对应有病理图像标签集;

将所述样本标签与所述至少两个病理图像库分别对应的病理图像标签集进行匹配,从所述至少两个病理图像库中确定与所述样本标签匹配的样本病理图像库;

将所述样本病理图像输入分类模型,得到样本病理图像库对应的样本预测概率,所述分类模型是基于所述至少两个病理图像库构建的待训练模型;

基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值,对所述分类模型进行训练,得到病理图像分类模型,所述病理图像分类模型用于对目标病理图像进行分类,得到与所述至少两个病理图像库分别对应的预测结果。

2. 根据权利要求1所述的方法,其特征在于,所述样本病理图像库对应的样本病理图像标签集中包括至少一个样本病理图像标签;

所述基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值,包括:

获取所述分类模型的病理图像结果中与所述样本病理图像库对应的样本预测概率,所述样本预测概率中包括所述样本病理图像库的样本病理图像标签分别对应的预测概率;

将所述样本病理图像库对应所述样本预测概率的标签预测结果与所述样本标签输入损失函数,得到所述样本病理图像对应的损失值。

3. 根据权利要求1所述的方法,其特征在于,所述病理图像标签集中包括至少一个病理图像标签;

所述将所述样本标签与所述至少两个病理图像库分别对应的病理图像标签集进行匹配,从所述至少两个病理图像库中确定与所述样本病理图像匹配的样本病理图像库,包括:

将所述样本标签与所述至少两个病理图像库分别对应的病理图像标签集中的病理图像标签进行匹配;

响应于所述病理图像库对应的病理图像标签集中包括所述样本标签,将所述病理图像库确定为所述样本病理图像库。

4. 根据权利要求1至3任一所述的方法,其特征在于,所述对所述分类模型进行训练,得到病理图像分类模型之后,还包括:

将所述目标病理图像输入所述病理图像分类模型中,确定所述目标病理图像与病理图像标签总集中的病理图像标签分别对应的标签预测概率,所述病理图像标签总集为所述至少两个病理图像库分别对应的病理图像标签集的集合;

基于所述至少两个病理图像库之间的划分标准,对所述病理图像标签总集进行划分,得到所述目标病理图像在所述至少两个病理图像库中分别对应的标签预测概率;

基于所述目标病理图像在所述至少两个病理图像库中分别对应的标签预测概率,确定所述目标病理图像对应的病理图像分类结果。

5. 根据权利要求4所述的方法,其特征在于,所述将所述目标病理图像输入所述病理图像分类模型中,确定所述目标病理图像与病理图像标签总集中的病理图像标签分别对应的标签预测概率,包括:

将所述目标病理图像输入所述病理图像分类模型;

通过所述病理图像分类模型预测所述目标病理图像与所述病理图像标签总集中的病理图像标签分别对应的预测情况,得到所述标签预测概率。

6. 根据权利要求4所述的方法,其特征在于,所述基于所述目标病理图像在所述至少两个病理图像库中分别对应的标签预测概率,确定所述目标病理图像对应的病理图像分类结果,包括:

响应于所述目标病理图像在所述至少两个病理图像库中分别对应的标签预测概率达到预测概率标准,确定所述目标病理图像分别对应所述至少两个病理图像库的病理图像分类结果。

7. 根据权利要求1至3任一所述的方法,其特征在于,所述基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值,对所述分类模型进行训练,得到病理图像分类模型,包括:

基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值;

基于所述样本标签对应的损失值,对所述分类模型的模型参数进行调整,得到候选分类模型;

响应于基于所述损失值对所述候选分类模型的训练达到训练目标,获取所述病理图像分类模型。

8. 根据权利要求7所述的方法,其特征在于,所述响应于基于所述损失值对所述候选分类模型的训练达到训练目标,获取所述病理图像分类模型,包括:

响应于所述损失值达到收敛状态,将最近一次迭代训练得到的所述候选分类模型作为所述病理图像分类模型;

或者,

响应于所述损失值的获取次数达到次数阈值,将最近一次迭代训练得到的候选分类模型作为所述病理图像分类模型。

9. 根据权利要求1至3任一所述的方法,其特征在于,所述从至少两个病理图像库中获取样本病理图像,包括:

从所述至少两个病理图像库中随机选择至少一个病理图像作为所述样本病理图像;

或者,

从所述至少两个病理图像库中轮询选择病理图像作为所述样本病理图像。

10. 一种病理图像分类装置,其特征在于,所述装置包括:

获取模块,用于从至少两个病理图像库中获取样本病理图像,所述样本病理图像标注有样本标签,所述至少两个病理图像库分别对应有病理图像标签集;

匹配模块,用于将所述样本标签与所述至少两个病理图像库分别对应的病理图像标签集进行匹配,从所述至少两个病理图像库中确定与所述样本标签匹配的样本病理图像库;

确定模型,用于将所述样本病理图像输入分类模型,得到样本病理图像库对应的样本预测概率,所述分类模型是基于所述至少两个病理图像库构建的待训练模型;

训练模块,用于基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值,对所述分类模型进行训练,得到病理图像分类模型,所述病理图像分类模型用于对目标病理图像进行分类,得到与所述至少两个病理图像库分别对应的预测结果。

11. 一种计算机设备,其特征在于,所述计算机设备包括处理器和存储器,所述存储器

中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如权利要求1至9任一所述的病理图像分类方法。

12.一种计算机可读存储介质,其特征在于,所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由处理器加载并执行以实现如权利要求1至9任一所述的病理图像分类方法。

13.一种计算机程序产品,其特征在于,包括计算机程序或指令,所述计算机程序或指令被处理器执行时实现如权利要求1至9任一所述的病理图像分类方法。

## 病理图像分类方法、装置、设备、存储介质及程序产品

### 技术领域

[0001] 本申请实施例涉及机器学习领域，特别涉及一种病理图像分类方法、装置、设备、存储介质及程序产品。

### 背景技术

[0002] 病理图像分类是通过病理图像的属性或特征对病理图像进行级别划分的过程。识别不同级别病理图像的价值，从而对病理图像进行正确地分析，是得到正确医学分析结果的基础。

[0003] 相关技术中，通常是基于特定病理图像集对应的特定分类标签对病理图像进行分类的，且一个病理图像集往往对应训练一个模型，通过将待分类的病理图像输入特定病理图像集对应的模型中，确定该待分类病理图像在该特定病理图像集中的分类标签。

[0004] 然而，通过上述方法，每个病理图像集训练得到的模型是独立的，不同的病理图像集之间的关联性较差，当待分类病理图像同时符合多个病理图像集对应的分类标签时，无法同时得到该待分类病理图像的在多个病理图像集中的分类标签。

### 发明内容

[0005] 本申请实施例提供了一种病理图像分类方法、装置、设备、存储介质及程序产品，能够加强病理图像库之间的关联性，更好地实现集中化的病理图像分类。所述技术方案如下。

[0006] 一方面，提供了一种病理图像分类方法，所述方法包括：

[0007] 从至少两个病理图像库中获取样本病理图像，所述样本病理图像标注有样本标签，所述至少两个病理图像库分别对应有病理图像标签集；

[0008] 将所述样本标签与所述至少两个病理图像库分别对应的病理图像标签集进行匹配，从所述至少两个病理图像库中确定与所述样本标签匹配的样本病理图像库；

[0009] 将所述样本病理图像输入分类模型，得到样本病理图像库对应的样本预测概率，所述分类模型是基于所述至少两个病理图像库构建的待训练模型；

[0010] 基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值，对所述分类模型进行训练，得到病理图像分类模型，所述病理图像分类模型用于对目标病理图像进行分类，得到与所述至少两个病理图像库分别对应的预测结果。

[0011] 另一方面，提供了一种病理图像分类装置，所述装置包括：

[0012] 获取模块，用于从至少两个病理图像库中获取样本病理图像，所述样本病理图像标注有样本标签，所述至少两个病理图像库分别对应有病理图像标签集；

[0013] 匹配模块，用于将所述样本标签与所述至少两个病理图像库分别对应的病理图像标签集进行匹配，从所述至少两个病理图像库中确定与所述样本标签匹配的样本病理图像库；

[0014] 确定模型，用于将所述样本病理图像输入分类模型，得到样本病理图像库对应的

样本预测概率,所述分类模型是基于所述至少两个病理图像库构建的待训练模型;

[0015] 训练模块,用于基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值,对所述分类模型进行训练,得到病理图像分类模型,所述病理图像分类模型用于对目标病理图像进行分类,得到与所述至少两个病理图像库分别对应的预测结果。

[0016] 另一方面,提供了一种计算机设备,所述计算机设备包括处理器和存储器,所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如上述本申请实施例中任一所述病理图像分类方法。

[0017] 另一方面,提供了一种计算机可读存储介质,所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由处理器加载并执行以实现如上述本申请实施例中任一所述的病理图像分类方法。

[0018] 另一方面,提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述实施例中任一所述的病理图像分类方法。

[0019] 本申请实施例提供的技术方案带来的有益效果至少包括:

[0020] 从病理图像库中获取样本病理图像,将样本病理图像对应的样本标签与病理图像库对应的病理图像标签集进行匹配,从而从病理图像库中确定样本病理图像所属的样本病理图像库,之后,将样本病理图像输入基于上述病理图像库构建的分类模型中,输出得到样本病理图像库对应的样本预测概率,从而基于样本标签和样本预测概率确定损失值,实现对分类模型的训练过程,得到病理图像分类模型。在应用时,将待分类的目标病理图像输入上述病理图像分类模型中,同时输出与病理图像库相关的多个病理图像标签作为目标病理图像对应的预测结果,实现目标病理图像的分类。通过上述方法,可以避免在对目标病理图像进行分类时,需要将目标病理图像输入多个病理图像库分别对应的多个分类模型中才能得到预测结果的问题,通过本申请中的病理图像分类模型,可以加强病理图像库之间的关联性,更好地实现集中化、专业化、标准化的病理图像分类,也便于更好地管理和使用病理图像。

## 附图说明

[0021] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例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一些实施例,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。

[0022] 图1是本申请一个示例性实施例提供的实施环境示意图;

[0023] 图2是本申请一个示例性实施例提供的病理图像分类方法的流程图;

[0024] 图3是本申请另一个示例性实施例提供的病理图像分类方法的流程图;

[0025] 图4是本申请另一个示例性实施例提供的病理图像分类方法的流程图;

[0026] 图5是本申请一个示例性实施例提供的苏木精—伊红染色方法的示意图;

[0027] 图6是本申请另一个示例性实施例提供的病理图像分类方法的流程图;

- [0028] 图7是本申请一个示例性实施例提供的对细胞核病理图像进行处理的示意图；
- [0029] 图8是本申请一个示例性实施例提供的应用病理图像分类方法获取细胞核分类标签的示意图；
- [0030] 图9是本申请一个示例性实施例提供的病理图像分类装置的结构框图；
- [0031] 图10是本申请另一个示例性实施例提供的病理图像分类装置的结构框图；
- [0032] 图11是本申请一个示例性实施例提供的服务器的结构框图。

## 具体实施方式

[0033] 为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

[0034] 首先，针对本申请实施例中涉及的名词进行简单介绍。

[0035] 人工智能(Artificial Intelligence, AI):是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。

[0036] 人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0037] 机器学习(Machine Learning, ML):是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、示教学习等技术。

[0038] 相关技术中,通常是基于特定病理图像集对应的特定分类标签对病理图像进行分类的,且一个病理图像集往往对应训练一个模型,通过将待分类的病理图像输入特定病理图像集对应的模型中,确定该待分类病理图像在该特定病理图像集中的分类标签。然而,通过上述方法,每个病理图像集训练得到的模型是独立的,不同的病理图像集之间的关联性较差,当待分类病理图像同时符合多个病理图像集对应的分类标签时,无法利用多个病理图像集之间的关联性同时得到该待分类病理图像在多个病理图像集中的分类标签。

[0039] 本申请实施例中,提供了一种病理图像分类方法,可以考虑多个病理图像库之间的关联性,通过病理图像分类模型对一个待分类病理图像进行分析,同时得到待分类病理图像在多个病理图像库中的分类标签。可选地,将本申请训练得到的病理图像分类方法应用于医学领域中为例进行说明。

[0040] 在医学领域中,病理图像库一般难以标注,所以病理图像库中存储的病理图像的体量一般都较小,每一个病理图像库对应训练一个分类模型,当需要对待分类的病理图像进行多标签分类时,通常会将该病理图像输入多个病理图像库对应的多个分类模型中,得

到该病理图像在不同医学病理图像库中对应的医学病理图像标签。示意性的,采用本申请中提供的病理图像分类方法,通过多个医学病理图像库构建分类模型,将医学病理图像库中的医学病理图像作为样本病理图像对该分类模型进行训练,得到训练好的病理图像分类模型。在应用时,将待分类医学病理图像输入该病理图像分类模型中,可以同时得到该待分类医学病理图像对应的多个医学标签,使得获取医学病理图像标签的过程更便捷。

[0041] 其次,对本申请实施例中涉及的实施环境进行说明,示意性的,请参考图1,该实施环境中涉及终端110、服务器120,终端110和服务器120之间通过通信网络130连接。

[0042] 在一些实施例中,终端110中安装有具有病理图像获取功能的应用程序。在一些实施例中,终端110用于向服务器120发送目标病理图像。服务器120可通过病理图像分类模型预测得到预测概率,根据预测概率对目标病理图像进行分类后,输出得到预测结果,并将预测结果反馈至终端110进行显示。

[0043] 其中,病理图像分类模型是采用该病理图像分类方法,通过从病理图像库中获取得到的样本病理图像训练得到的。示意性的,从n个病理图像库中获取样本病理图像后,将样本病理图像对应的样本标签与n个病理图像库分别对应的n个病理图像标签集进行匹配,可以确定样本病理图像所属的病理图像库,即确定样本病理图像库;将样本病理图像输入待训练的分类模型中,得到样本病理图像库对应的样本预测概率;基于样本预测概率与样本标签确定损失值,通过损失值对分类模型进行训练,得到病理图像分类模型。上述过程是病理图像分类模型训练过程的不唯一情形的举例。

[0044] 值得注意的是,上述终端包括但不限于手机、平板电脑、便携式膝上笔记本电脑、智能语音交互设备、智能家电、车载终端等移动终端,也可以实现为台式电脑等。

[0045] 上述服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云病理图像库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络(Content Delivery Network, CDN)、以及大病理图像和人工智能平台等基础云计算服务的云服务器。

[0046] 其中,云技术(Cloud technology)是指在广域网或局域网内将硬件、应用程序、网络等系列资源统一起来,实现病理图像的计算、储存、处理和共享的一种托管技术。云技术基于云计算商业模式应用的网络技术、信息技术、整合技术、管理平台技术、应用技术等的总称,可以组成资源池,按需所用,灵活便利。云计算技术将变成重要支撑。技术网络系统的后台服务需要大量的计算、存储资源,如视频网站、图片类网站和更多的门户网站。伴随着互联网行业的高度发展和应用,将来每个物品都有可能存在自己的识别标志,都需要传输到后台系统进行逻辑处理,不同程度级别的病理图像将会分开处理,各类行业病理图像皆需要强大的系统后盾支撑,只能通过云计算来实现。

[0047] 在一些实施例中,上述服务器还可以实现为区块链系统中的节点。区块链(Blockchain)是分布式病理图像存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链,本质上是一个去中心化的病理图像库,是一串使用密码学方法相关联产生的病理图像块,每一个病理图像块中包含了一批次网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层。

[0048] 结合上述名词简介和应用场景,对本申请提供的病理图像分类方法进行说明,以



该方法应用于服务器为例,如图2所示,该方法包括如下步骤。

[0049] 步骤210,从至少两个病理图像库中获取样本病理图像。

[0050] 病理图像库用于存储病理图像,是病理图像的集合。病理图像库中既可以存储一种类型的病理图像(如:上皮细胞病理图像或者癌细胞病理图像),也可以存储多种类型的病理图像(如:上皮细胞病理图像和癌细胞病理图像)。可选地,根据病理图像库中存储病理图像的病理图像类型,对病理图像库进行划分,例如:病理图像库存储细胞病理图像,该病理图像库为细胞病理图像库。至少两个病理图像库既可以是存储有同种病理图像类型的病理图像库,也可以是存储有不同种病理图像类型的病理图像库,还可以为存储多种病理图像类型病理图像的病理图像库。

[0051] 可选地,从至少两个病理图像库中获取样本病理图像包括以下至少一种方法。

[0052] 1、从至少两个病理图像库中任意选择至少一个病理图像作为样本病理图像。

[0053] 示意性的,在至少两个病理图像库中存储有多个病理图像,从多个病理图像中以随机选择的方式,任意选择至少一个病理图像作为样本病理图像,即,以等概率选择的方式从多个病理图像中选择至少一个病理图像作为样本病理图像。

[0054] 2、从至少两个病理图像库中轮询选择病理图像作为样本病理图像。

[0055] 轮询是从病理图像库中依次选择病理图像后,再重复上述依次选择病理图像的过程。示意性的,预先确定3个病理图像库作为样本病理图像的病理图像来源,3个病理图像库分别为病理图像库1,病理图像库2以及病理图像库3。当选择样本病理图像时,从病理图像库1中选择一个样本病理图像后,从病理图像库2中选择一个样本病理图像,再从病理图像库3中选择一个样本病理图像,将上述过程作为一次选择过程,之后重复上述选择过程,即:从病理图像库1中再选择一个样本病理图像后,从病理图像库2中再选择一个样本病理图像等。待轮询的次数达到次数阈值或者被选择的病理图像库中不存在可供选择的病理图像时,停止样本病理图像的选择过程。需要注意的是,在不同的病理图像库中选择样本病理图像的个数并不唯一,对于轮询结束的条件也并不唯一,以上仅为示意性的举例,本申请实施例对此不加以限定。

[0056] 可选地,每一个病理图像库对应有一个病理图像标签集,病理图像标签集中包括至少一个病理图像标签,用于指示该病理图像库中部分病理图像或全部病理图像的相同特征。示意性的,A病理图像库中包括病理图像A1、病理图像A2、病理图像A3、病理图像A4和病理图像A5,A病理图像库对应的病理图像标签集中包括病理图像标签a1和a2,其中病理图像A1、病理图像A2和病理图像A3对应标注有病理图像标签a1,即:病理图像A1、病理图像A2和病理图像A3具有相同特征,可以通过病理图像标签a1体现;病理图像A4和病理图像A5对应标注有病理图像标签a2,即:病理图像A4和病理图像A5具有相同特征,可以通过病理图像标签a2体现。

[0057] 样本病理图像是从至少两个病理图像库中获取得到的,样本病理图像标注有样本标签。

[0058] 步骤220,将样本标签与至少两个病理图像库分别对应的病理图像标签集进行匹配,从至少两个病理图像库中确定与样本标签匹配的样本病理图像库。

[0059] 在一个可选的实施例中,将样本标签与至少两个病理图像库分别对应的病理图像标签集中的病理图像标签进行匹配;响应于病理图像库对应的病理图像标签集中包括样本

标签,将病理图像库确定为样本病理图像库。

[0060] 样本病理图像是从至少两个病理图像库中获取到的,因此样本病理图像对应的样本标签和至少两个病理图像库对应的至少两个病理图像标签集存在一定的对应关系。可选地,将样本标签与至少两个病理图像库分别对应的病理图像标签集中的每个病理图像标签一一进行匹配,得到标签匹配结果,其中,标签匹配结果至少包括以下一种情况。

[0061] 1、标签匹配结果为标签匹配成功。

[0062] 示意性的,将样本标签与病理图像标签集中的一个病理图像标签进行匹配后,当样本标签与该病理图像标签相同时,视为标签匹配成功。例如:样本标签为“淋巴细胞”,病理图像标签为“淋巴细胞”,即符合标签匹配成功的条件;或者,当样本标签与该病理图像标签的相似度超过预设的相似度阈值时,视为标签匹配相同。例如,预设的相似度阈值为0.8,样本标签与病理图像标签为文本信息,确定样本标签与病理图像标签之间的相似度的过程,即进行文本相似度比较的过程,当相似度超过0.8时,即符合标签匹配成功的条件。可选地,在将样本标签与该病理图像标签进行匹配后,将样本标签与病理图像标签集中的其他病理图像标签继续进行匹配。

[0063] 2、标签匹配结果为标签匹配失败。

[0064] 示意性的,将样本标签与病理图像标签集中的一个病理图像标签进行匹配后,当样本标签与该病理图像标签不同时,视为标签匹配失败。例如:样本标签为“上皮细胞”,病理图像标签为“淋巴细胞”,即样本标签与病理图像标签不同,符合标签匹配失败的条件。

[0065] 在一个可选的实施例中,标签匹配成功为标签相同,当样本标签与病理图像库对应的病理图像标签集进行匹配时,当样本标签与病理图像标签集中至少一个病理图像标签相同,即视为标签匹配成功。示意性的,样本标签为m,将样本标签m与3个病理图像库对应的病理图像标签集进行匹配,其中,3个病理图像标签集分别为病理图像标签集M(其中包括:病理图像标签m、病理图像标签l和病理图像标签n);病理图像标签集L(其中包括:病理图像标签q、病理图像标签l和病理图像标签n);病理图像标签集N(其中包括:病理图像标签m、病理图像标签o和病理图像标签n),基于样本标签m与病理图像标签集M和病理图像标签集N中的病理图像标签m相同,则视为样本标签与病理图像标签集M和病理图像标签集N匹配成功,则将病理图像标签集M和病理图像标签集N对应的病理图像库确定为样本病理图像库。以上仅为示意性的举例,本申请实施例对此不加以限定。

[0066] 可选地,当样本病理图像为从至少两个病理图像库中选择的多个病理图像时,将样本病理图像对应的样本标签依次与至少两个病理图像库分别对应的病理图像标签集进行匹配,根据样本病理图像对应的样本标签,将样本标签与病理图像标签集中的病理图像标签进行匹配,示意性的,当样本标签与病理图像标签相同时,即视为标签匹配成功,确定该病理图像标签所属的病理图像标签集,并随之确定该病理图像标签集对应的病理图像库,将该病理图像库作为样本病理图像库。

[0067] 步骤230,将样本病理图像输入分类模型,得到样本病理图像库对应的样本预测概率。

[0068] 其中,分类模型是基于至少两个病理图像库构建的待训练模型。

[0069] 分类是通过病理图像的属性或者特征对已有病理图像进行分类的过程,分类模型是为了实现分类过程而采用的模型,又称作分类器。通过分类模型,可以根据病理图像的属

性和特征等对病理图像进行划分,常见的分类模型包括:逻辑回归模型、决策树模型、多层感知机模型等。

[0070] 可选地,病理图像的种类多种多样,相同种类的病理图像往往存储在同一个病理图像库中,对于不同种类的病理图像库,通常会训练得到不同的分类模型,例如:基于上皮细胞病理图像库中的上皮细胞病理图像训练得到上皮细胞病理图像分类模型;基于癌细胞病理图像库中的癌细胞病理图像训练得到癌细胞病理图像分类模型等。

[0071] 示意性的,分类模型为一个具有一定病理图像分类功能的通用分类模型,即:通用分类领域可以用于对多个不同领域或者某个特定领域的病理图像进行基本的分类。不同的通用分类模型可能对应不同的病理图像库,可选地,基于不同病理图像库之间的关联性,对不同病理图像库对应的通用分类模型进行模型融合,得到可用于多个病理图像库的分类模型;或者,为多个不同病理图像库对应的通用分类模型提供一个框架,该框架可以利用多个病理图像集之间的关系,对病理图像进行多标签分类,将该框架作为可用于多个病理图像库的分类模型。

[0072] 在确定分类模型后,为了提高分类模型的分类精确度,通常会在该分类模型的基础上,基于病理图像库对分类模型进行训练,例如:将分类模型作为待训练模型,对该分类模型在一个或者多个病理图像库下进行训练,使得该分类模型能够较好地识别病理图像的分类类别,提高分类模型的分类精确度。

[0073] 在一个可选的实施例中,至少两个病理图像库为存储有相同种类病理图像的同种类病理图像库,例如:至少两个病理图像库为存储有细胞病理图像的病理图像库,细胞病理图像包括上皮细胞病理图像、癌细胞病理图像、淋巴细胞病理图像等。分类模型为一个与病理图像库种类相同的通用分类模型,分类模型是基于至少两个病理图像库构建得到的。样本病理图像是从至少两个病理图像库中随机获取到的病理图像,将该样本病理图像输入分类模型中,该分类模型可以将样本病理图像与至少两个病理图像库中的病理图像进行匹配,输出得到包括样本预测概率的病理图像结果。其中,病理图像结果是分类模型输出的预测结果,样本预测概率是分类模型输出的样本病理图像与多个样本病理图像标签分别对应的预测结果,样本预测概率与样本病理图像库对应,且从属于病理图像结果。

[0074] 步骤240,基于样本病理图像库对应的样本预测概率和样本标签确定损失值,对分类模型进行训练,得到病理图像分类模型。

[0075] 可选地,在基于样本预测概率和样本标签确定损失值时,包括如下方式中的至少一种:

[0076] 第一种,获取分类模型的病理图像结果中与样本病理图像库对应的样本预测概率,样本预测概率中包括样本病理图像库的样本病理图像标签分别对应的预测概率;将样本病理图像库对应样本预测概率的标签预测结果与样本标签输入损失函数,得到样本病理图像对应的损失值。

[0077] 示意性的,样本病理图像A对应的样本标签为 $a_1$ ,样本病理图像库对应样本病理图像标签集,样本病理图像标签集中包括三个样本病理图像标签,分别为 $a_2, b, c$ ,将样本病理图像A输入分类模型中,分类模型输出与三个病理图像标签 $a_2, b, c$ 分别对应的样本预测概率为0.7、0.1、0.2,则确定样本病理图像A对应样本病理图像库的标签预测结果为 $a_2$ ,则将样本病理图像A对应的样本标签为 $a_1$ 和标签预测结果为 $a_2$ 输入损失函数,得到样本病理图

像A对应的损失值。

[0078] 第二种,当样本标签与样本病理图像库中某个病理图像标签相同,即样本匹配成功,该样本匹配结果为1;基于样本标签与样本病理图像库中某个病理图像标签不同,即样本匹配失败,该样本匹配结果为0。基于样本病理图像对应样本病理图像库的样本病理图像标签的样本匹配结果,以及分类模型的病理图像结果中与样本病理图像库对应的样本预测概率计算样本病理图像对应的损失值,样本预测概率中包括样本病理图像库的样本病理图像标签分别对应的预测概率。

[0079] 示意性的,样本病理图像A对应的样本标签为a,样本病理图像库对应样本病理图像标签集,样本病理图像标签集中包括三个样本病理图像标签,分别为a,b,c,将样本标签与样本病理图像标签集中的病理图像标签进行匹配,得到的样本匹配结果分别为“1、0、0”;将样本病理图像A输入分类模型中,分类模型输出与三个病理图像标签a,b,c分别对应的样本预测概率,为“0.7、0.1、0.2”。对“1、0、0”和“0.7、0.1、0.2”进行距离计算,得到样本病理图像对应的损失值。

[0080] 值得注意的是,本申请实施例中,以通过样本标签和标签预测结果计算损失值为例进行说明。

[0081] 另外,上述样本预测概率以分类模型输出软标签为例进行说明,在一些实施例中,分类模型所输出的样本预测概率也可以实现为硬标签,也即,将样本病理图像A输入分类模型中,分类模型输出与三个病理图像标签a,b,c分别对应的样本预测概率为“1、0、0”,本实施例对此不加以限定。

[0082] 可选地,损失值是通过损失函数计算得到的,损失函数为预先设定好的函数,将样本标签与基于样本病理图像库得到的标签预测结果代入上述损失函数中,得到损失值,并基于损失值对分类模型进行训练,得到病理图像分类模型。其中,病理图像分类模型为对分类模型进行训练后得到的。

[0083] 在一个可选的实施例中,病理图像分类模型用于对目标病理图像进行分类,得到与至少两个病理图像库分别对应的预测结果。

[0084] 可选地,在得到病理图像分类模型后,可以通过病理图像分类模型对目标病理图像进行分类。其中,目标病理图像既可以是至少两个病理图像库中的病理图像,也可以是至少两个病理图像库之外的病理图像,即:病理图像分类模型的应用情形并不限于构建分类模型时所采用的至少两个病理图像库。示意性的,病理图像分类模型可以对与至少两个病理图像库种类相同的病理图像库中的病理图像进行较为精确的分类预测,通过将目标病理图像输入病理图像分类模型中,得到该目标病理图像与至少两个病理图像库分别对应的预测结果,例如:确定该目标病理图像在至少两个病理图像库中对应的病理图像标签。

[0085] 可选地,在获取目标病理图像前,获取参考病理图像,参考病理图像为未经过预处理前的目标病理图像。根据参考病理图像的病理图像类型,对参考病理图像进行预处理后,得到更适合模型识别的目标病理图像,可选地,对参考病理图像的预处理包括裁剪处理、锐化处理、去噪处理等。示意性的,以裁剪处理为例,参考病理图像为一幅细胞核病理图像,该细胞核病理图像为对目标采集点进行采集后得到的病理图像,之后,以目标采集点对应的区域对该细胞核病理图像进行获取,得到至少两个目标细胞核病理图像,即获取得到目标病理图像。以上仅为示意性的举例,本申请实施例对此不加以限定。

[0086] 综上所述,从病理图像库中获取样本病理图像,将样本病理图像对应的样本标签与病理图像库对应的病理图像标签集进行匹配,从而从病理图像库中确定样本病理图像所属的样本病理图像库,之后,将样本病理图像输入基于上述病理图像库构建的分类模型中,输出得到样本病理图像库对应的样本预测概率,从而基于样本标签和样本预测概率确定损失值,实现对分类模型的训练过程,得到病理图像分类模型。在应用时,将待分类的目标病理图像输入上述病理图像分类模型中,同时输出与病理图像库相关的多个病理图像标签作为目标病理图像对应的预测结果,实现目标病理图像的分类。通过上述方法,可以避免在对目标病理图像进行分类时,需要将目标病理图像输入多个病理图像库分别对应的多个分类模型中才能得到预测结果的问题,通过本申请中的病理图像分类模型,可以加强病理图像库之间的关联性,更好地实现集中化、专业化、标准化的病理图像分类,也便于更好地管理和使用病理图像。

[0087] 在一个可选的实施例中,对分类模型进行训练得到病理图像分类模型的过程是通过损失值实现的。示意性的,如图3所示,上述图2所示出的实施例中的步骤240还可以实现为如下步骤310至步骤340。

[0088] 步骤310,获取分类模型的病理图像结果中与样本病理图像库对应的样本预测概率。

[0089] 可选地,将样本病理图像输入分类模型后,分类模型输出病理图像结果,例如:病理图像结果以概率形式输出。病理图像结果中既包括与样本病理图像库对应的样本预测概率,也包括与除样本病理图像库以外的其他病理图像库对应的病理图像预测概率。

[0090] 其中,样本预测概率中包括样本病理图像库的样本病理图像标签分别对应的预测概率,即:样本预测概率是分类模型输出的预测概率,该预测概率为样本病理图像与多个样本病理图像标签分别对应的预测概率。

[0091] 步骤320,将样本病理图像库对应样本预测概率的标签预测结果与样本标签输入损失函数,得到样本病理图像对应的损失值。

[0092] 可选地,通过损失函数可以计算得到的损失值,损失函数为预先设定的函数。例如:损失函数的表达式如下所示。

$$[0093] \quad L_0 = - \sum_{i \in B} \sum_{j=1}^D d_{i,j} \sum_{m=1}^{S_j} y_{i,j}(m) \log \left( \frac{\exp(y'_{i,j}(m))}{\sum_m \exp(y'_{i,j}(m))} \right)$$

[0094] 其中, $L_0$ 表示损失函数; $i$ 表示的是第*i*个病理图像, $B$ 表示病理图像采样组; $j$ 表示的是第*j*个病理图像库, $D$ 表示的是全部病理图像库; $d_{i,j}$ 是一个0/1指示函数,为1则代表病理图像*i*有病理图像库*j*对应的病理图像标签集的病理图像标签,为0则代表该病理图像*i*没有病理图像库*j*对应的病理图像标签集的病理图像标签; $m$ 表示在第*j*个病理图像库里的第*m*个标签, $S_j$ 表示病理图像标签集一共有*S<sub>j</sub>*个病理图像标签; $y_{i,j}(m)$ 表示第*j*个病理图像库中第*m*个标签的匹配概率; $y'_{i,j}(m)$ 表示模型输出的第*j*个标签集中第*m*个标签的预测概率。 $B$ 病理图像采样组既可以是上述至少两个病理图像库,也可以是从上述至少两个病理图像库中随机获取得到的病理图像的集合。

[0095] 可选地,上述损失函数表达式是带有0/1指示函数的交叉熵损失函数,指示函数用于在对分类模型的训练过程中,确定样本病理图像所属的样本病理图像库,即用于指示样

本标签是否为某个病理图像库对应的病理图像标签。示意性的,  $i$  表示样本病理图像, 当指示函数为1时, 表示样本病理图像有病理图像库  $j$  对应的病理图像标签集中的某个病理图像标签, 将该病理图像标签视为真实标签。可选地, 应用上述损失函数时, 当指示函数为1时, 将样本病理图像库对应样本预测概率的标签预测结果参与损失函数计算; 当指示函数为0时, 该样本病理图像与该病理图像库对应的病理图像标签集中的病理图像标签不匹配, 其他病理图像库对应的病理图像预测概率不参与损失函数计算。

[0096] 可选地, 表示样本病理图像;  $j$  表示样本病理图像库;  $y_{i,j}(m)$  为样本标签。基于分类模型以及样本病理图像库得到样本预测概率的标签预测结果  $y'_{i,j}(m)$ , 将样本标签  $y_{i,j}(m)$  和样本预测概率的标签预测结果  $y'_{i,j}(m)$  代入上述损失函数中, 计算得到损失值, 损失值用于指示样本标签与标签预测结果之间的差异程度, 可选地, 样本标签与标签预测结果之间的差异程度通过概率表示。

[0097] 基于上述损失函数, 确定样本病理图像对应的损失值, 可选地, 从病理图像采样组中获取病理图像进行预测时, 当被选择的样本病理图像为多个病理图像, 对每一个样本病理图像采用上述方法进行损失值计算; 或者, 当选择一次样本病理图像后, 需要对分类模型进行多次训练, 既可以继续采用上次选取的样本病理图像对分类模型进行训练, 也可以从至少两个病理图像库中重新选择其余样本病理图像对分类模型进行训练, 以上仅为示意性的举例, 本申请实施例对此不加以限定。

[0098] 步骤330, 基于样本标签对应的损失值, 对分类模型的模型参数进行调整, 得到候选分类模型。

[0099] 示意性的, 以降低损失值为目标对分类模型的模型参数进行调整, 如: 采用梯度下降法降低损失值; 或者, 采用反向传播算法降低损失值等。

[0100] 可选地, 基于一个样本标签对应的损失值, 可以对分类模型的模型参数进行至少一次调整, 当存在多个样本病理图像对应的样本标签时, 需要对分类模型的模型参数进行多次调整。对分类模型的模型参数进行调整的目的在于得到训练好的病理图像分类模型, 示意性的, 在对分类模型的模型参数进行调整从而得到病理图像分类模型的过程中, 模型参数被调整但尚未达到病理图像分类模型条件的模型可以称为候选分类模型, 即: 候选分类模型为对分类模型的模型参数进行调整后得到的模型, 因为尚未训练完毕, 故候选分类模型为中间状态的模型。

[0101] 步骤340, 响应于基于损失值对候选分类模型的训练达到训练目标, 得到病理图像分类模型。

[0102] 示意性的, 从至少两个病理图像库中随机获取多个样本病理图像, 以每个样本病理图像对分类模型进行一次训练为前提, 对分类模型进行训练模型。例如: 对每一个样本病理图像进行标签匹配、概率预测以及损失计算后, 得到每一个样本病理图像对应的损失值; 之后, 以第一样本病理图像对应的第一损失值对分类模型进行初次调整后, 得到候选分类模型; 之后, 以第二个样本病理图像对应的第二损失值对候选分类模型进行训练等。可选地, 同一个样本病理图像也可以对分类模型进行一次或者多次训练, 以上仅为示意性的举例, 本申请实施例对此不加以限定。

[0103] 可选地, 在损失值对候选分类模型进行训练的过程中, 会因为对候选分类模型的训练达到训练目标而得到病理图像分类模型, 示意性的, 训练目标至少包括如下一种情况。

[0104] 1、响应于损失值达到收敛状态,将最近一次迭代训练得到的候选分类模型作为病理图像分类模型。

[0105] 示意性的,损失值达到收敛状态用于指示通过损失函数得到的损失值的数值不再变化或者变化幅度小于预设阈值。例如:第n个样本病理图像对应的损失值为0.1,第n+1个样本病理图像对应的损失值也为0.1,可以视为该损失值达到收敛状态,将第n个样本病理图像或者第n+1个样本病理图像对应的损失值调整的候选分类模型作为病理图像分类模型,实现对分类模型的训练过程。

[0106] 2、响应于损失值的获取次数达到次数阈值,将最近一次迭代训练得到的候选分类模型作为病理图像分类模型。

[0107] 示意性的,一次获取可以得到一个损失值,预先设定用于训练分类模型的损失值的获取次数,当一个样本病理图像对应一个损失值时,损失值的获取次数即为样本病理图像个数;或者,当一个样本病理图像对应多个损失值时,损失的获取次数即为损失值的个数。例如:预先设定一次获取可以得到一个损失值,损失值获取的次数阈值为10次,即当达到获取次数阈值时,将最近一次损失值调整的候选分类模型作为病理图像分类模型,或者将损失值10次调整过程中最小损失值调整的候选分类模型作为病理图像分类模型,实现对分类模型的训练过程。

[0108] 以上仅为示意性的举例,本申请实施例对此不加以限定。

[0109] 综上所述,将样本标签与病理图像库对应的病理图像标签集进行匹配,从而确定样本病理图像对应的样本病理图像库,之后,将样本病理图像输入分类模型中,输出得到样本病理图像库对应的样本预测概率,从而利用样本标签和样本预测概率确定的损失值,对分类模型进行训练得到病理图像分类模型。在应用时,将目标病理图像输入病理图像分类模型中,同时输出多个病理图像标签作为目标病理图像对应的预测结果,实现目标病理图像的分类。通过本申请中的病理图像分类模型,可以加强病理图像库之间的关联性,更好地实现集中化、专业化、标准化的病理图像分类,也便于更好地管理和使用病理图像。

[0110] 在本申请实施例中,对基于损失值训练分类模型得到病理图像分类模型的过程进行阐述,基于将样本病理图像输入分类模型得到的样本预测概率的标签预测结果以及样本病理图像对应的样本标签,可以知悉样本病理图像的真实标签与经过待训练模型得到的样本病理图像的预测标签之间的差异程度,即损失值,损失值可以通过上述损失函数计算得到。基于损失值对分类模型的模型参数进行调整,直至达到训练目标得到病理图像分类模型。通过上述方法,可以利用损失值提高分类模型的鲁棒性,得到能够更好预测病理图像分类情况的病理图像分类模型。

[0111] 在一个可选的实施例中,在得到病理图像分类模型后,可以通过病理图像分类模型对目标病理图像进行病理图像分类,得到病理图像分类结果。示意性的,如图4所示,上述图2所示出的实施例中的步骤240之后还包括如下步骤410至步骤430。

[0112] 步骤410,将目标病理图像输入病理图像分类模型中,确定目标病理图像与病理图像标签总集中的病理图像标签分别对应的标签预测概率。

[0113] 在一个可选的实施例中,将目标病理图像输入病理图像分类模型;通过病理图像分类模型预测目标病理图像与病理图像标签总集中的病理图像标签分别对应的预测情况,得到标签预测概率。



[0114] 示意性的,病理图像分类模型是通过样本病理图像对分类模型训练之后得到的模型,目标病理图像包括文本、图像、音频等多种形式。目标病理图像可以通过提取、截取、合成等方式进行获取,在得到目标病理图像后,可以将目标病理图像输入病理图像分类模型中;或者,对目标病理图像进行预处理后(例如:对文本进行分词;对图像进行截取;对音频进行解码、编码等),将处理后的目标病理图像输入病理图像分类模型中。

[0115] 病理图像标签总集为至少两个病理图像库分别对应的病理图像标签集的集合。示意性的,每一个病理图像库对应一个病理图像标签集,一个病理图像标签集中包括至少一个病理图像标签。病理图像标签总集包括了至少两个病理图像库对应的所有病理图像标签。可选地,目标病理图像对应标注有目标标签,在将目标病理图像输入病理图像分类模型后,分类模型输出与病理图像标签数量相同的多个标签预测概率。示意性的,病理图像标签总集中有 $x$ 个病理图像标签,将一个目标病理图像输入病理图像分类模型后,该病理图像分类模型输出 $x$ 个标签预测概率。

[0116] 步骤420,基于至少两个病理图像库之间的划分标准,对病理图像标签总集进行划分,得到目标病理图像在至少两个病理图像库中分别对应的标签预测概率。

[0117] 示意性的,至少两个病理图像库之间存在一定的划分标准,例如:至少两个病理图像库的来源不同;或者,至少两个病理图像库中存储的病理图像种类不同;或者,至少两个病理图像库中病理图像之间的联系方式不同等。可选地,在获取至少两个病理图像库时,病理图像库之间的划分标准随之确定,如:Y病理图像库来源于Y研究机构,Z病理图像库来源于Z研究机构,病理图像库的来源不同即为Y病理图像库与Z病理图像库之间的划分标准。

[0118] 可选地,不同的病理图像库对应不同的病理图像标签集,病理图像标签总集为病理图像标签集的集合,根据至少两个病理图像库之间的划分标准对病理图像标签总集进行划分的过程,是对病理图像标签总集进行划分确定病理图像标签集的过程。一个病理图像标签集中包括多个病理图像标签,将目标病理图像输入病理图像分类模型后,病理图像分类模型输出多个标签预测概率,基于对病理图像库的划分标准,一个病理图像库中对应的多个标签预测概率随之被划分为一组。示意性的,以 $P$ 个病理图像库为基础对分类模型进行训练, $P$ 个病理图像库中分别对应的多个标签预测概率随之被划分为 $P$ 组。

[0119] 步骤430,基于目标病理图像在至少两个病理图像库中分别对应的标签预测概率,确定目标病理图像对应的病理图像分类结果。

[0120] 根据标签预测概率可以知道目标病理图像对应的目标标签(目标病理图像的真实标签)与多个病理图像标签之间的差异程度,可选地,在得到目标病理图像在至少两个病理图像库中分别对应的标签预测概率后,根据标签预测概率可以确定目标病理图像对应的病理图像分类结果,病理图像分类结果为目标病理图像对应至少两个病理图像库中的病理图像标签。

[0121] 示意性的,响应于目标病理图像在至少两个病理图像库中分别对应的标签预测概率达到预测概率标准,确定目标病理图像分别对应至少两个病理图像库的病理图像分类结果。其中,根据预测概率标准的不同,病理图像分类结果包括如下至少一种情况。

[0122] 1、预测概率标准为取最大的标签预测概率。

[0123] 示意性的,将目标病理图像输入病理图像分类模型后得到 $d$ 个标签预测概率,将 $d$ 个标签预测概率进行数值比较,取最大的标签预测概率对应的病理图像标签,将该病理图



像标签作为病理图像分类结果,确定目标病理图像的病理图像分类情况。

[0124] 2、预测概率标准为达到预先设定好的概率阈值。

[0125] 示意性的,将目标病理图像输入病理图像分类模型后得到e个标签预测概率,预先设定概率阈值为0.8,将e个标签预测概率中超过0.8的标签预测概率对应的病理图像标签作为病理图像分类结果,确定目标病理图像的病理图像分类情况。可选的,基于概率阈值确定病理图像分类结果时,可能存在目标病理图像对应一个病理图像库中0个病理图像标签、1个病理图像标签或者多个病理图像标签的情况,基于此种情况,既可以将以上对应的病理图像标签均进行保留,得到多个病理图像分类结果,也可以从中确定唯一一个病理图像标签作为唯一的病理图像分类结果,本申请实施例对此不加以限定。

[0126] 综上所述,将样本标签与病理图像库对应的病理图像标签集进行匹配,从而确定样本病理图像对应的样本病理图像库,之后,将样本病理图像输入分类模型中,输出得到样本病理图像库对应的样本预测概率,从而利用样本标签和样本预测概率确定的损失值,对分类模型进行训练得到病理图像分类模型。在应用时,将目标病理图像输入病理图像分类模型中,同时输出多个病理图像标签作为目标病理图像对应的预测结果,实现目标病理图像的分类。通过本申请中的病理图像分类模型,可以加强病理图像库之间的关联性,更好地实现集中化、专业化、标准化的病理图像分类,也便于更好地管理和使用病理图像。

[0127] 在本申请实施例中,利用上述病理图像分类方法得到的病理图像分类模型,对待分类的目标病理图像进行分类,确定目标病理图像对应的目标标签与每一个病理图像标签分别对应的标签预测概率,基于病理图像库之间的划分标准,对标签预测概率进行划分,之后,基于不同的病理图像库对应的标签预测概率,确定目标病理图像对应的病理图像分类结果,达到将目标病理图像输入一个病理图像分类模型同时得到多个标签的目的,实现对目标病理图像进行多标签分类的过程。

[0128] 在一个可选的实施例中,通过上述病理图像分类模型,可以对病理图像进行多标签分类,即每一个目标病理图像可以对应预测一个或者多个分类标签,从而实现基于多病理图像集的多标签分类的过程。示意性的,将病理图像分类方法应用于细胞核分类中,病理图像分类方法实现为训练过程和应用过程。

[0129] 一、训练过程

[0130] 可选地,将未经训练用于分类的模型称为分类模型;将对细胞核进行多标签分类的、训练完毕的分类模型称为细胞核分类模型,即:通过病理图像库对分类模型进行训练后得到细胞核分类模型。在训练得到细胞核分类模型时,首先确定参与训练分类模型的细胞核病理图像库以及细胞核病理图像库对应的病理图像标签集。可选地,分类模型是以五个数据库为基础进行构建的,每个数据库中存储有病理图像,故可以将数据库视为病理图像库。五个数据库分别为:NuCLS数据库(Nucleus Classification,核分类),BreCaHAD数据库(Breast Cancer Histopathological Annotation and Diagnosis,乳腺癌的组织病理学注释与诊断),CoNSEP数据库(Colon segmentation and phenotype,结直肠核分割和表型),MoNuSAC数据库(Multi-organ Nuclei Segmentation and Classification,多器官细胞核的分割与分类),panNuke数据库(pan-cancer histology dataset for Nuclei instance segmentation and classification,癌组织细胞数据库)。示意性的,每一个数据库(可视为病理图像库)对应一个数据标签集(可视为病理图像标签集),每个数据标签集

中包括多个数据标签(可视为病理图像标签),不同的数据标签集中的数据标签可能相同,也可能不同。示意性的,上述5个数据库分别对应一个数据标签集,数据标签集如下所示,其中,阿拉伯数字编号可用于指示不同的数据标签。

[0131] NuCLS数据标签集:0、肿瘤细胞;1、成纤维细胞;2、淋巴细胞;3、浆细胞;4、巨噬细胞;5、有丝分裂像;6、血管内皮;7、肌上皮细胞;8、凋亡体;9、中性粒细胞;10、导管上皮细胞;11、嗜酸性细胞

[0132] BreCaHAD数据标签集:0、有丝分裂体;1、非有丝分裂体;2、细胞凋亡体;3、肿瘤细胞;4、非肿瘤细胞;5、管腔细胞;6、非管腔细胞

[0133] CoNSEP数据标签集:0、其他细胞;1、炎症细胞;2、健康上皮细胞;3、恶性上皮细胞;4、成纤维细胞;5、结缔细胞;6、内皮细胞

[0134] MoNuSAC数据标签集:0、上皮细胞;1、淋巴细胞;2、中性粒细胞;3、巨噬细胞

[0135] PanNuke数据标签集:0、肿瘤性细胞;1、炎症细胞;2、结缔细胞;3、死亡细胞;4、上皮细胞

[0136] 上述5个数据标签集共对应35个数据标签,其中,NuCLS数据标签集对应12个数据标签;BreCaHAD数据标签集对应7个数据标签;CoNSEP数据标签集对应7个数据标签;MoNuSAC数据标签集对应4个数据标签;PanNuke数据标签集对应5个数据标签。

[0137] 在对分类模型进行训练时,从上述5个数据库中任意选择一个样本数据输入分类模型中,分类模型会将样本数据与35个数据标签分别进行匹配,在分类模型的最后一层,共计输出35个概率,用于指示样本数据与35个数据标签的差异情况。但是在训练过程中,关注样本数据本身真实的标签(即:样本标签),取样本数据标签对应的样本预测概率的交叉熵作为损失函数,并且忽略出样本预测概率以外其他概率。示意性的,分类模型输出的概率是以上述数据集和数据标签的次序依次输出的,若样本数据为从NuCLS数据集出来获取到的数据,当在进行损失函数计算时,只对输出前12个样本预测概率做交叉熵损失函数计算,得到损失值。

[0138] 可选地,将上述5个数据集集中的数据通过分类模型后确定不同数据对应的损失值,采用得到的损失值对分类模型进行训练,即可以实现用数据集集中的数据训练分类模型的过程。损失函数是训练分类模型的函数,通过损失函数确定损失值,可以更好地训练分类模型前的编码部分,提升分类模型预测的准确度,进而得到准确度较高、且可以进行多标签分类的细胞核分类模型。

[0139] 二、应用过程

[0140] HE染色(Hematoxylin-Eosin staining,苏木精-伊红染色)是显微镜医学分析最常用的染色方式之一,由两种染色组成,其中,苏木精染液为碱性,主要使细胞核内的染色质与胞质内的核酸着紫蓝色;伊红染液为酸性,主要是细胞质和细胞外基质中的成分着红色。示意性的,如图5所示,是使用HE染色方法对不同细胞进行染色得到的成像结果。

[0141] 可选地,在基于训练过程得到细胞核分类模型后,对目标细胞核进行细胞核多标签分类,即:利用多个数据库对应的不同类别标签,对同一个细胞核进行多个分类。示意性的,如图6所示,基于细胞核分类模型对细胞核进行多标签分类的应用过程包括如下步骤。

[0142] 步骤610,细胞检测分割。

[0143] 示意性的,如图7所示,在对细胞核进行分析时,对细胞核进行HE染色,使得细胞核

便于观察。之后,获取具有细胞核染色的病理图像710,将该病理图像710输入通用的图像检测分割模型,例如:Hovernet模型(悬停网模型,用于多组织组织学图像中细胞核的同时分割和分类);或者,采用通用图像的检测分割算法,例如:Mask R-CNN(Mask Region-Convolution Neural Network,掩模区域卷积神经网络)和Fast R-CNN(Fast Region-Convolution Neural Network,快速区域卷积神经网络)。示意性的,将病理图像710输入Hovernet模型中得到检测图像720,从而得到待检测细胞的位置信息。可选地,以检测图像720中的待检测细胞为中心,切割出一个尺寸为 $64 \times 64$ 大小的细胞图像730,将细胞图像730放大到尺寸为 $224 \times 224$ 后,输入训练得到的细胞核分类模型中进行分类处理。

[0144] 其中,将检测图像720切割为 $64 \times 64$ 大小的细胞图像730是为了较全面而精确地呈现待检测细胞的细胞状态,将细胞图像730放大到 $224 \times 224$ 是因为尺寸为 $224 \times 224$ 的图像更便于细胞核分类模型进行分类预测。

[0145] 步骤620,细胞多标签分类。

[0146] 示意性的,将目标细胞图像输入训练得到的细胞核分类模型中,可以实现对细胞的多标签分类过程。其中,目标细胞图像为待进行分类预测的细胞图像。通过细胞核分类模型,可以得到目标细胞图像对应的35个预测概率,将35个预测概率以数据库划分标准进行划分,即:NuCLS数据库对应12个预测概率;BreCaHAD数据库对应7个预测概率;CoNSEP数据库对应7个预测概率;MoNuSAC数据库对应4个预测概率;PanNuke数据库对应5个预测概率。可选地,将每个数据库对应的预测概率最大数据标签作为目标细胞图像对应的预测标签,因此一个目标细胞图像对应应有五个预测标签。

[0147] 可选地,预测标签可以通过对应的数据标签集和对应的数据标签进行表示,示意性的,预测标签对应的数据标签集以数据标签集的名称表示;预测标签对应的数据标签通过将该数据标签的阿拉伯数字编号作为下标的形式表示(其中,阿拉伯数字编号对应上述数据标签集中不同的数据标签对应的阿拉伯数字编号)。例如:目标细胞图像对应的预测标签为 $NuCLS_0$ ,即表示:该预测标签对应的数据标签集为NuCLS数据标签集,该预测标签对应的数据标签为NuCLS数据标签集中的数据标签0对应的标签结果——肿瘤细胞。同理,若目标细胞图像对应的预测标签为 $BreCaHAD_3$ ,即表示:BreCaHAD数据标签集下的肿瘤细胞;若目标细胞图像对应的预测标签为 $CoNSEP_3$ ,即表示:CoNSEP数据标签集下的恶性上皮细胞;若目标细胞图像对应的预测标签为 $MoNuSAC_0$ ,即表示:MoNuSAC数据标签集下的上皮细胞;若目标细胞图像对应的预测标签为 $panNuke_0$ ,即表示:panNuke数据标签集下的肿瘤细胞。

[0148] 可选地,将上述数据分类方法应用于终端或者服务器,实现实时对提取得到的病理图片中的细胞核进行自动多标签分类过程。当采用上述五个数据库时,对每个细胞核会输出五个标签,示意性的,输入与输出结果如图8所示。

[0149] 病理图片810对应病理标签811,将病理图片810输入数据分类模型中,会得到病理图片810与上述5个数据集分别对应的标签预测情况。例如:病理图片810与NuCLS数据标签集对应的NuCLS标签预测821;病理图片810与BreCaHAD数据标签集对应的BreCaHAD标签预测831;病理图片810与CoNSEP数据标签集对应的CoNSEP标签预测841;病理图片810与MoNuSAC数据标签集对应的MoNuSAC标签预测851;病理图片810与panNuke数据标签集对应的panNuke标签预测861。

[0150] 综上所述,将样本标签与病理图像库对应的病理图像标签集进行匹配,从而确定

样本病理图像对应的样本病理图像库,之后,将样本病理图像输入分类模型中,输出得到样本病理图像库对应的样本预测概率,从而利用样本标签和样本预测概率确定的损失值,对分类模型进行训练得到病理图像分类模型。在应用时,将目标病理图像输入病理图像分类模型中,同时输出多个病理图像标签作为目标病理图像对应的预测结果,实现目标病理图像的分类。通过本申请中的病理图像分类模型,可以加强病理图像库之间的关联性,更好地实现集中化、专业化、标准化的病理图像分类,也便于更好地管理和使用病理图像。

[0151] 在本申请实施例中,基于上述病理图像分类方法,将目标细胞图像对应的目标标签和每个参与训练的数据集对应的数据标签集中的数据标签进行匹配,通过匹配得到的预测概率,从每一个数据集对应的数据标签集中确定一个预测概率最高的数据标签作为目标细胞图像对应的分类预测结果,即可得到一个目标细胞图像对应的多个数据标签,实现对一个目标细胞图像进行多标签分类的目的。

[0152] 图9是本申请一个示例性实施例提供的病理图像分类装置的结构框图,如图9所示,该装置包括如下部分:

[0153] 获取模块910,用于从至少两个病理图像库中获取样本病理图像,所述样本病理图像标注有样本标签,所述至少两个病理图像库分别对应有病理图像标签集;

[0154] 匹配模块920,用于将所述样本标签与所述至少两个病理图像库分别对应的病理图像标签集进行匹配,从所述至少两个病理图像库中确定与所述样本标签匹配的样本病理图像库;

[0155] 确定模块930,用于将所述样本病理图像输入分类模型,得到样本病理图像库对应的样本预测概率,所述分类模型是基于所述至少两个病理图像库构建的待训练模型;

[0156] 训练模块940,用于基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值,对所述分类模型进行训练,得到病理图像分类模型,所述病理图像分类模型用于对目标病理图像进行分类,得到与所述至少两个病理图像库分别对应的预测结果。

[0157] 如图10所示,在一个可选的实施例中,所述样本病理图像库对应的样本病理图像标签集中包括至少一个样本病理图像标签;

[0158] 所述训练模块940包括:

[0159] 获取单元941,用于获取所述分类模型的病理图像结果中与所述样本病理图像库对应的样本预测概率,所述样本预测概率中包括所述样本病理图像库的样本病理图像标签分别对应的预测概率;

[0160] 输入单元942,用于将所述样本病理图像库对应样本预测概率的标签预测结果与所述样本标签输入损失函数,得到所述样本病理图像对应的损失值。

[0161] 在一个可选的实施例中,病理图像标签集中包括至少一个病理图像标签;

[0162] 所述匹配模型920用于将所述样本标签与所述至少两个病理图像库分别对应的病理图像标签集中的病理图像标签进行匹配;响应于所述病理图像库对应的病理图像标签集中包括所述样本标签,将所述病理图像库确定为所述样本病理图像库。

[0163] 在一个可选的实施例中,所述装置还包括:

[0164] 概率确定模块950,用于将所述目标病理图像输入所述病理图像分类模型中,确定所述目标病理图像与病理图像标签总集中的病理图像标签分别对应的标签预测概率,所述病理图像标签总集为所述至少两个病理图像库分别对应的病理图像标签集的集合;

[0165] 划分模块960,用于基于所述至少两个病理图像库之间的划分标准,对所述病理图像标签总集进行划分,得到所述目标病理图像在所述至少两个病理图像库中分别对应的标签预测概率;

[0166] 结果确定模块970,用于基于所述目标病理图像在所述至少两个病理图像库中分别对应的标签预测概率,确定所述目标病理图像对应的病理图像分类结果。

[0167] 在一个可选的实施例中,所述概率确定模块950还用于将所述目标病理图像输入所述病理图像分类模型;通过所述病理图像分类模型预测所述目标病理图像与所述病理图像标签总集中的病理图像标签分别对应的预测情况,得到所述标签预测概率。

[0168] 在一个可选的实施例中,所述结果确定模块970还用于响应于所述目标病理图像在所述至少两个病理图像库中分别对应的标签预测概率达到预测概率标准,确定所述目标病理图像分别对应所述至少两个病理图像库的病理图像分类结果。

[0169] 在一个可选的实施例中,所述训练模块940还用于基于所述样本病理图像库对应的样本预测概率和所述样本标签确定损失值;基于所述样本标签对应的损失值,对所述分类模型的模型参数进行调整,得到候选分类模型;响应于基于所述损失值对所述候选分类模型的训练达到训练目标,获取所述病理图像分类模型。

[0170] 在一个可选的实施例中,所述训练模块940还用于响应于所述损失值达到收敛状态,将最近一次迭代训练得到的所述候选分类模型作为所述病理图像分类模型;或者,响应于所述损失值的获取次数达到次数阈值,将最近一次迭代训练得到的候选分类模型作为所述病理图像分类模型。

[0171] 在一个可选的实施例中,所述获取模块910还用于从所述至少两个病理图像库中随机选择至少一个病理图像作为所述样本病理图像;或者,从所述至少两个病理图像库中轮询选择病理图像作为所述样本病理图像。

[0172] 需要说明的是:上述实施例提供的病理图像分类装置,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将设备的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。另外,上述实施例提供的病理图像分类装置与病理图像分类方法实施例属于同一构思,其具体实现过程详见方法实施例,这里不再赘述。

[0173] 图11示出了本申请一个示例性实施例提供的服务器的结构示意图。该服务器1100包括中央处理单元(Central Processing Unit,CPU)1101、包括随机存取存储器(Random Access Memory,RAM)1102和只读存储器(Read Only Memory,ROM)1103的系统存储器1104,以及连接系统存储器1104和中央处理单元1101的系统总线1105。服务器1100还包括用于存储操作系统1113、应用程序1114和其他程序模块1115的大容量存储设备1106。

[0174] 大容量存储设备1106通过连接到系统总线1105的大容量存储控制器(未示出)连接到中央处理单元1101。大容量存储设备1106及其相关联的计算机可读介质为服务器1100提供非易失性存储。也就是说,大容量存储设备1106可以包括诸如硬盘或者紧凑型光盘只读存储器(Compact Disc Read Only Memory,CD-ROM)驱动器之类的计算机可读介质(未示出)。

[0175] 不失一般性,计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其他数据等信息的任何

方法或技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括RAM、ROM、可擦除可编程只读存储器(Erasable Programmable Read Only Memory, EPROM)、带电可擦可编程只读存储器(Electrically Erasable Programmable Read Only Memory, EEPROM)、闪存或其他固态存储技术, CD-ROM、数字通用光盘(Digital Versatile Disc, DVD)或其他光学存储、磁带盒、磁带、磁盘存储或其他磁性存储设备。当然, 本领域技术人员可知计算机存储介质不局限于上述几种。上述的系统存储器1104和大容量存储设备1106可以统称为存储器。

[0176] 根据本申请的各种实施例, 服务器1100还可以通过诸如因特网等网络连接到网络上的远程计算机运行。也即服务器1100可以通过连接在系统总线1105上的网络接口单元1111连接到网络1112, 或者说, 也可以使用网络接口单元1111来连接到其他类型的网络或远程计算机系统(未示出)。

[0177] 上述存储器还包括一个或者一个以上的程序, 一个或者一个以上程序存储于存储器中, 被配置由CPU执行。

[0178] 本申请的实施例还提供了一种计算机设备, 该计算机设备包括处理器和存储器, 该存储器中存储有至少一条指令、至少一段程序、代码集或指令集, 至少一条指令、至少一段程序、代码集或指令集由处理器加载并执行以实现上述各方法实施例提供的病理图像分类方法。

[0179] 本申请的实施例还提供了一种计算机可读存储介质, 该计算机可读存储介质上存储有至少一条指令、至少一段程序、代码集或指令集, 至少一条指令、至少一段程序、代码集或指令集由处理器加载并执行, 以实现上述各方法实施例提供的病理图像分类方法。

[0180] 本申请的实施例还提供了一种计算机程序产品或计算机程序, 该计算机程序产品或计算机程序包括计算机指令, 该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令, 处理器执行该计算机指令, 使得该计算机设备执行上述实施例中任一所述的病理图像分类方法。

[0181] 可选地, 该计算机可读存储介质可以包括: 只读存储器(ROM, Read Only Memory)、随机存取记忆体(RAM, Random Access Memory)、固态硬盘(SSD, Solid State Drives)或光盘等。其中, 随机存取记忆体可以包括电阻式随机存取记忆体(ReRAM, Resistance Random Access Memory)和动态随机存取存储器(DRAM, Dynamic Random Access Memory)。上述本申请实施例序号仅仅为了描述, 不代表实施例的优劣。

[0182] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成, 也可以通过程序来指令相关的硬件完成, 所述的程序可以存储于一种计算机可读存储介质中, 上述提到的存储介质可以是只读存储器, 磁盘或光盘等。

[0183] 以上所述仅为本申请的可选实施例, 并不用以限制本申请, 凡在本申请的精神和原则之内, 所作的任何修改、等同替换、改进等, 均应包含在本申请的保护范围之内。

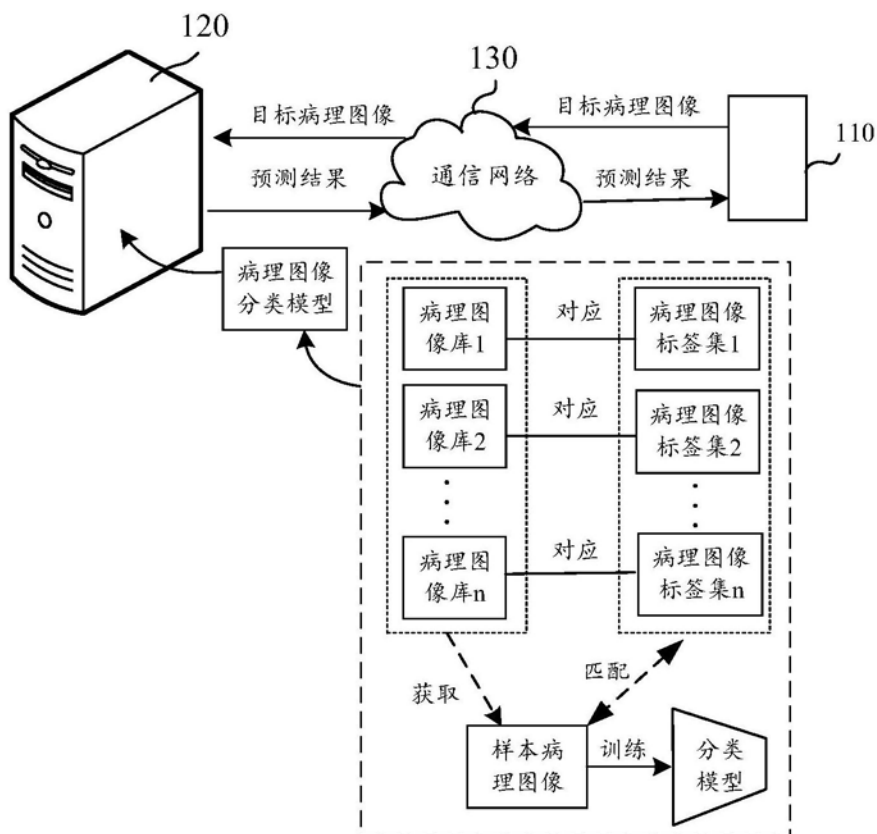


图1

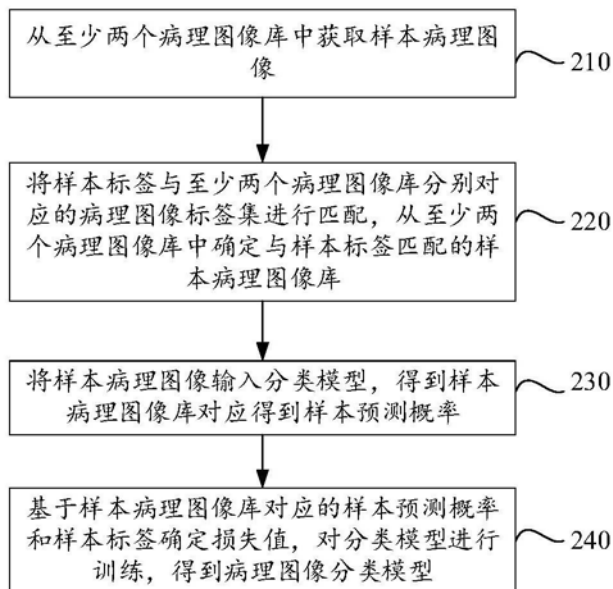


图2

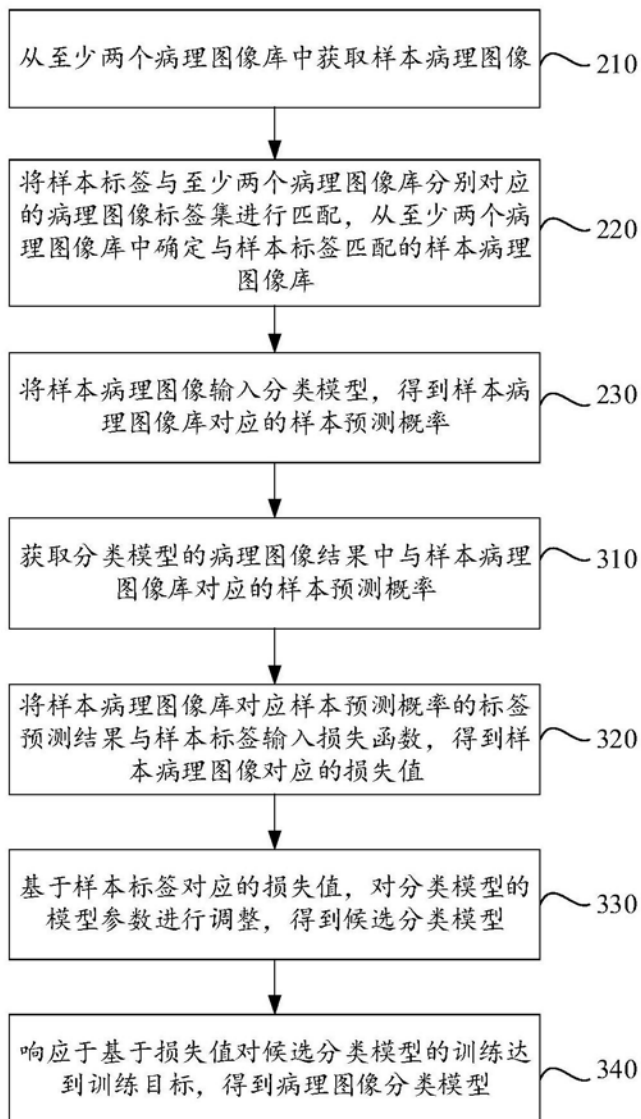


图3



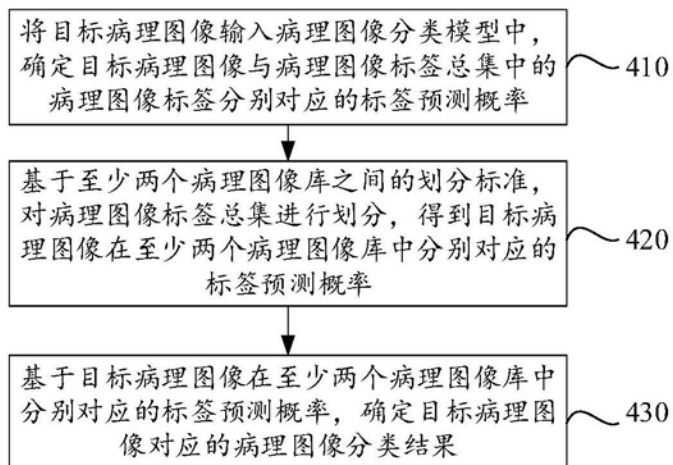


图4

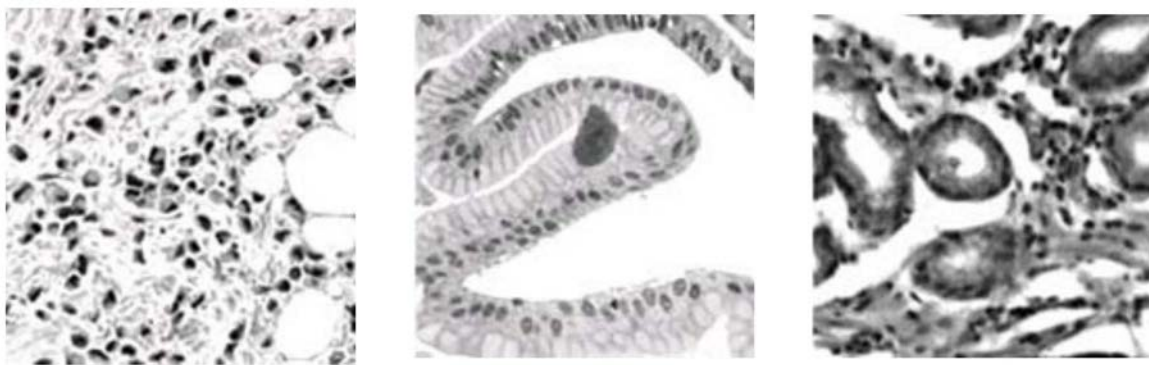


图5

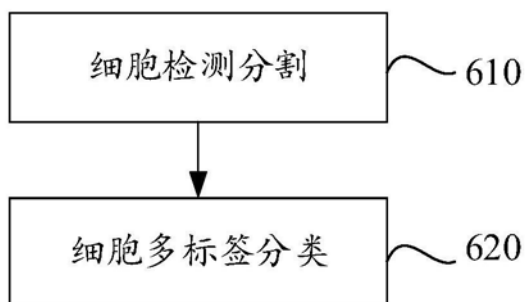


图6

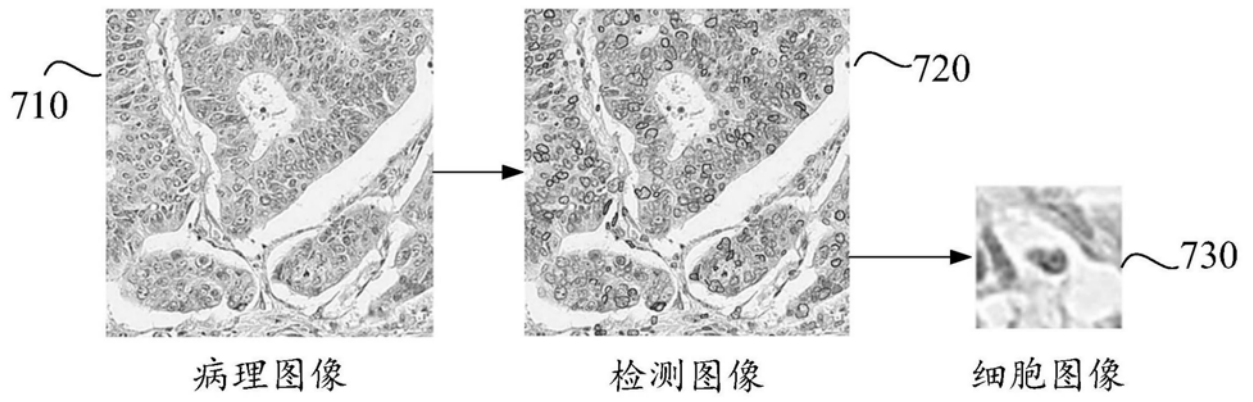


图7

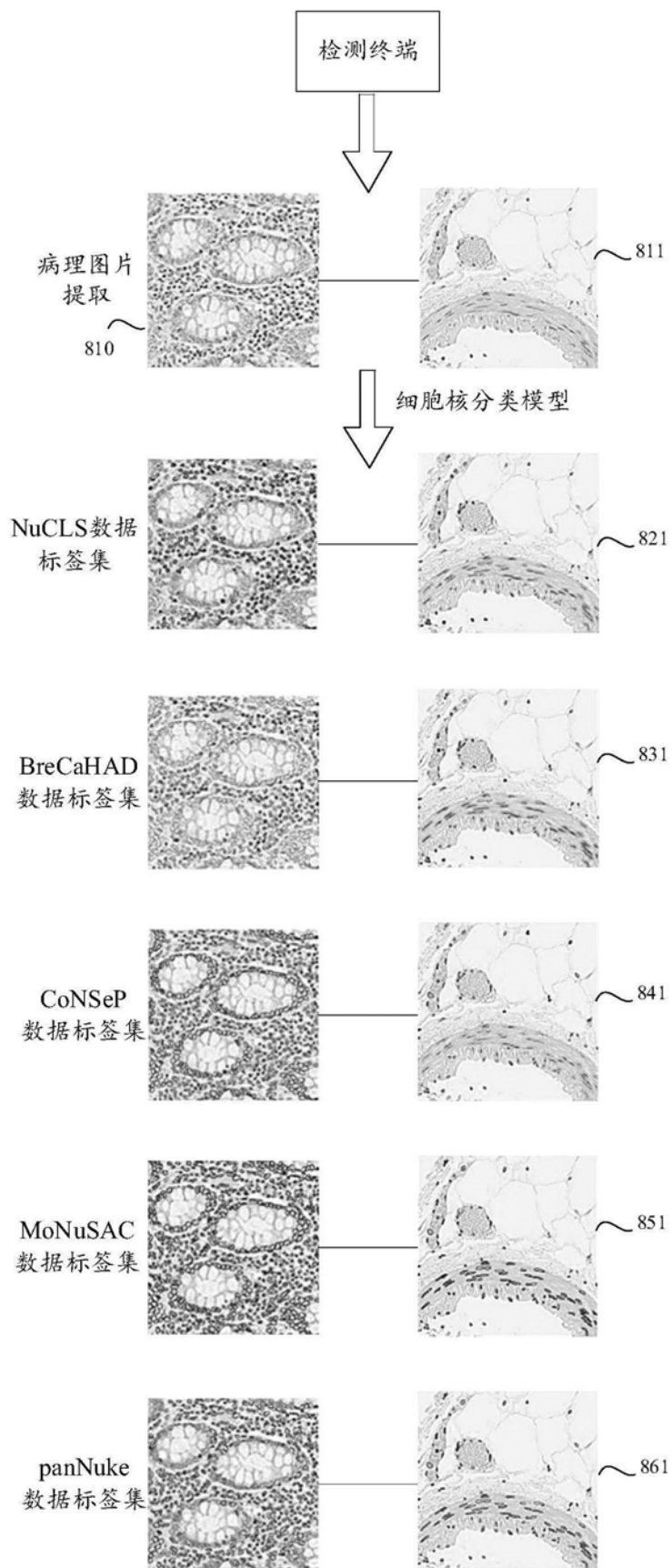


图8

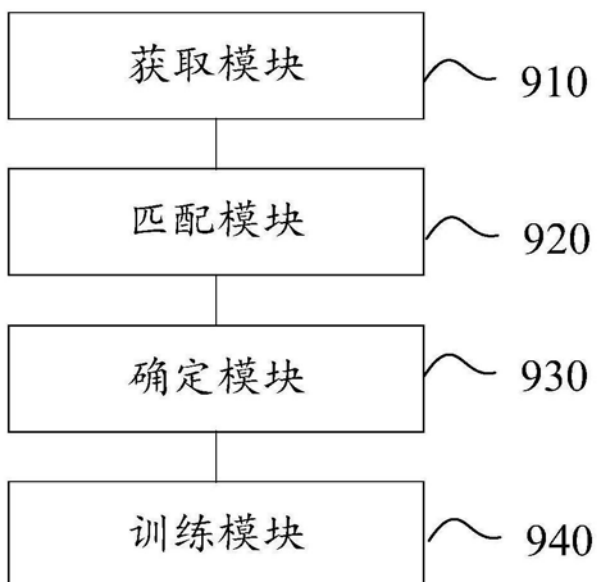


图9

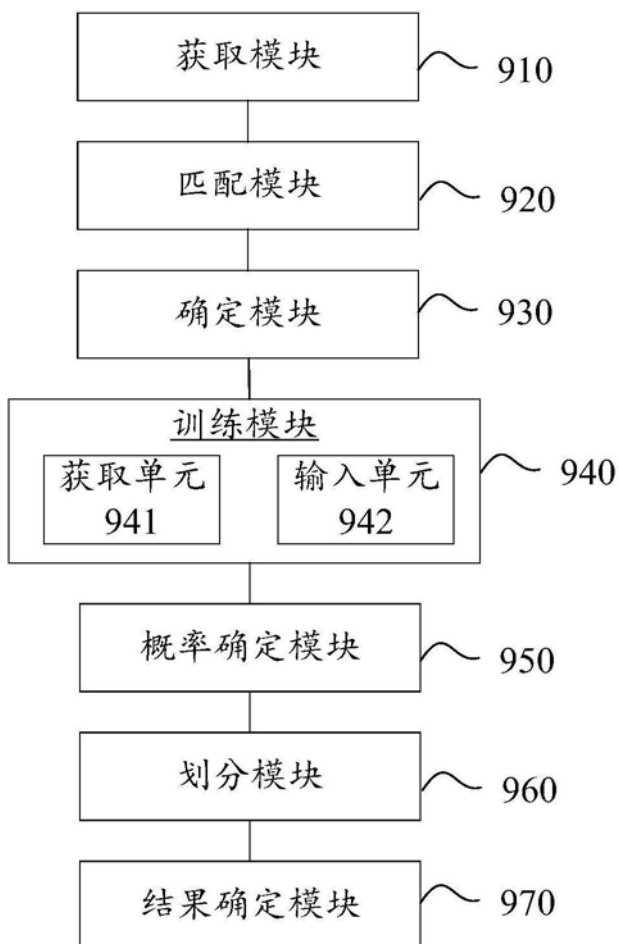


图10

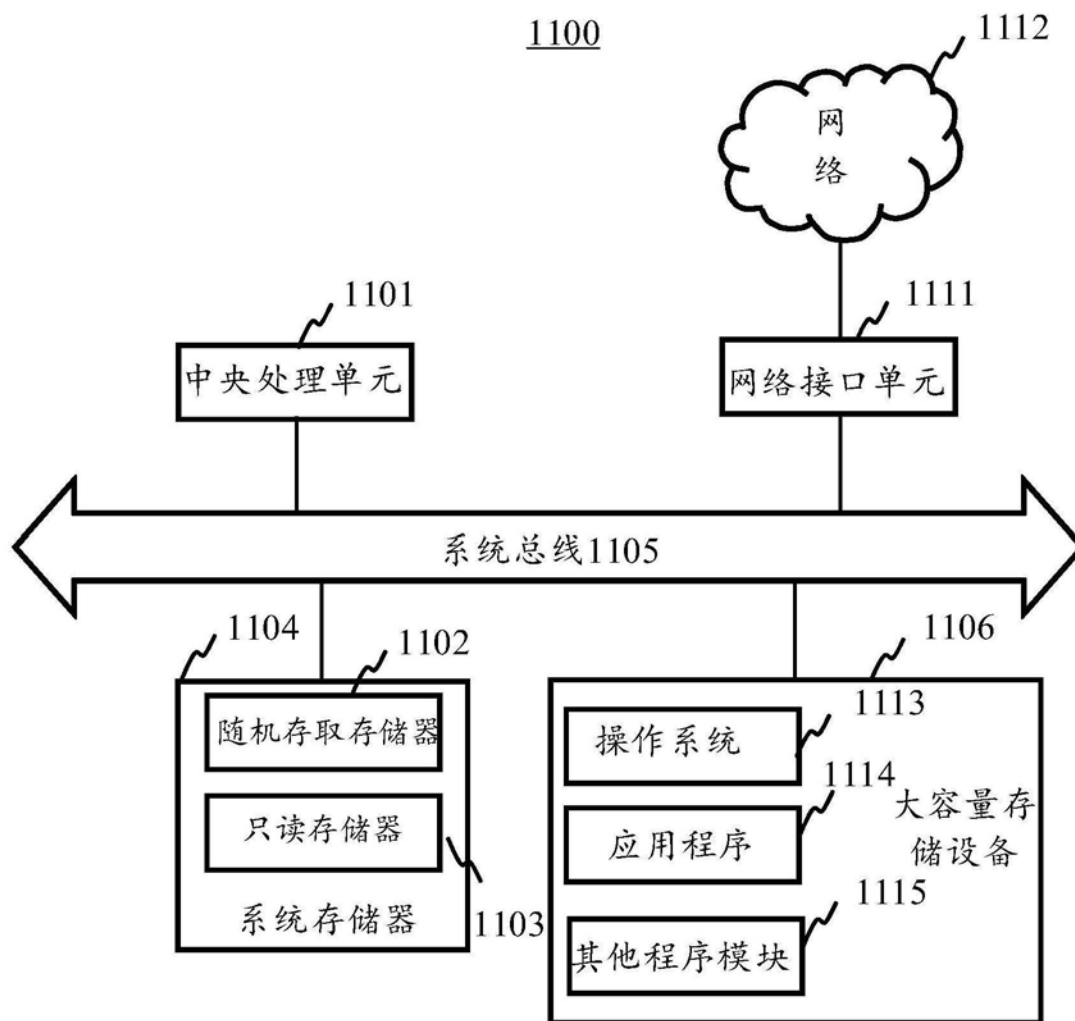


图11