



(12) 发明专利申请

(10) 申请公布号 CN 112232407 A

(43) 申请公布日 2021.01.15

(21) 申请号 202011101133.X

(22) 申请日 2020.10.15

(71) 申请人 杭州迪英加科技有限公司

地址 311121 浙江省杭州市余杭区仓前街道龙潭路7号未来研创园B座5楼B501-B508室

(72) 发明人 崔灿 惠文丽 杜家文 杨林

(74) 专利代理机构 杭州华知专利事务所(普通合伙) 33235

代理人 李姣姣

(51) Int. Cl.

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

权利要求书3页 说明书12页 附图3页

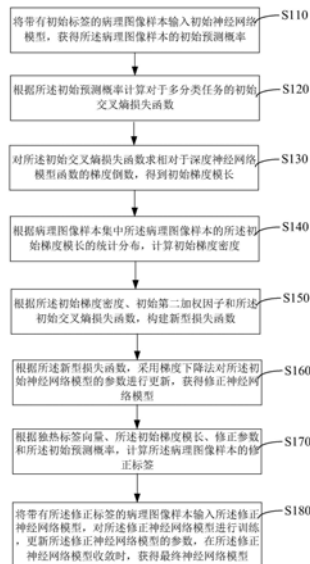
(54) 发明名称

病理图像样本的神经网络模型训练方法、装置

(57) 摘要

本发明涉及一种带有标签噪声的病理图像样本的神经网络模型训练方法、装置,该方法包括:将带有初始标签的病理图像样本输入到初始神经网络模型中,得到每个样本的初始预测概率;根据每个样本的初始标签和初始预测概率计算出初始交叉熵损失函数和初始梯度模长;使用梯度模长计算出每个样本对应梯度密度,使用梯度密度和高斯概率分布对初始交叉熵损失函数进行加权计算,调整易分样本和极度难分样本对模型训练的影响;使用梯度模长对样本的初始标签进行优化;使用梯度下降法和加权后的修正交叉熵损失函数对模型进行训练,更新模型的参数;在下次迭代使用优化后的标签作为样本的标签求损失值,重复迭代训练数次,直到模型收敛为止,可提高模型训练效果。

CN 112232407 A



1. 一种病理图像样本的神经网络模型训练方法,其特征在于,包括:

将带有初始标签的病理图像样本输入初始神经网络模型,获得所述病理图像样本的初始预测概率;其中,所述病理图像样本属于病理图像样本集;

根据所述初始预测概率计算对于多分类任务的初始交叉熵损失函数;

对所述初始交叉熵损失函数求相对于深度神经网络模型函数的梯度,得到初始梯度模长;

根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度;

根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数;

根据所述新型损失函数,采用梯度下降法对所述初始神经网络模型的参数进行更新,获得修正神经网络模型;

根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签;

将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型。

2. 根据权利要求1所述的方法,其特征在于,所述根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度,包括:

计算病理图像样本集中的每个所述病理图像样本的初始梯度模长的统计分布;

将所述初始梯度模长的取值均分为多个区域,并统计每个区域内的样本个数;

计算所述病理图像样本个数与所述区间的长度之比,得到所述病理图像样本在所述区间的初始梯度密度。

3. 根据权利要求2所述的方法,其特征在于,所述根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数,包括:

计算所述初始梯度密度的倒数和每个所述区域的样本个数之积,得到每个所述区域的所述病理图像样本的初始第一加权因子;

根据所述病理图像样本的初始梯度模长,计算所述病理图像样本的初始第二加权因子;

根据所述初始第一加权因子、初始第二加权因子、初始交叉熵损失函数和所述样本个数,构建新型损失函数。

4. 根据权利要求1所述的方法,其特征在于,在根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签之前,包括:

根据所述病理图像样本的初始标签的类别,定义每个初始标签的独热标签向量。

5. 根据权利要求1所述的方法,其特征在于,所述将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型,包括:

将带有所述修正标签的病理图像样本输入所述修正神经网络模型,获得所述病理图像样本的修正预测概率;其中,所述病理图像样本属于病理图像样本集;

根据所述修正预测概率计算对于多分类任务的修正交叉熵损失函数；

对所述修正交叉熵损失函数求相对于神经网络模型函数的梯度，得到修正梯度模长；

根据病理图像样本集中所述病理图像样本的所述修正梯度模长的统计分布，计算修正梯度密度；

根据所述修正梯度密度、修正第二加权因子和所述修正交叉熵损失函数，构建修正损失函数；

根据所述修正损失函数，采用梯度下降法对所述修正神经网络模型的参数进行更新，再次获得修正神经网络模型；

根据独热标签向量、所述修正梯度模长、修正参数和所述修正预测概率，再次计算所述病理图像样本的修正标签；

重复上述步骤，直到所述修正神经网络模型收敛，获得最终神经网络模型。

6. 根据权利要求5所述的方法，其特征在于，所述根据病理图像样本集中所述病理图像样本的所述修正梯度模长的统计分布，计算修正梯度密度，包括：

计算病理图像样本集中的每个所述病理图像样本的修正梯度模长的统计分布；

将所述修正梯度模长的取值均分为多个区域，并统计每个区域内的样本个数；

计算所述病理图像样本个数与所述区间的长度之比，得到所述病理图像样本在所述区间的修正梯度密度。

7. 根据权利要求6所述的方法，其特征在于，所述根据所述修正梯度密度、修正第二加权因子和所述修正交叉熵损失函数，构建修正损失函数，包括：

计算所述修正梯度密度的倒数和每个所述区域的样本个数之积，得到每个所述区域的所述病理图像样本的修正第一加权因子；

根据所述病理图像样本的修正梯度模长，计算所述病理图像样本的修正第二加权因子；

根据所述修正第一加权因子、修正第二加权因子、修正交叉熵损失函数和所述样本个数，构建新型损失函数。

8. 一种病理图像样本的神经网络模型训练装置，其特征在于，包括：

初始预测概率计算模块，用于将带有初始标签的病理图像样本输入初始神经网络模型，获得所述病理图像样本的初始预测概率；其中，所述病理图像样本属于病理图像样本集；

初始交叉熵损失函数计算模块，用于根据所述初始预测概率计算对于多分类任务的初始交叉熵损失函数；

初始梯度模长计算模块，用于对所述初始交叉熵损失函数求相对于神经网络模型函数的梯度，得到初始梯度模长；

初始梯度密度计算模块，用于根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布，计算初始梯度密度；

新型损失函数构建模块，用于根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数，构建新型损失函数；

参数更新模块，用于根据所述新型损失函数，采用梯度下降法对所述初始神经网络模

型的参数进行更新,获得修正神经网络模型;

修正标签计算模块,用于根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签;

模型训练模块,用于将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型。

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7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7中任一项所述的方法的步骤。

## 病理图像样本的神经网络模型训练方法、装置

### 技术领域

[0001] 本发明属于图像处理技术领域,特别涉及一种病理图像样本的神经网络模型训练方法、装置。

### 背景技术

[0002] 人工智能辅助病理影像分析是当前比较热门的研究领域之一,依靠人工智能辅助,可以节省大量的病理医生阅片时间,提升病理医生的工作效率并且缩小不同医生主观性判读造成的误差。目前主流的人工智能辅助病理影像分析大多基于深度学习的有监督学习方法。有监督学习首先需要人工标注大量的训练样本,然后使用这些标注的样本对模型进行训练,当模型训练完成后,即可以将需要分析的病理影像输入到模型中进行分析。

[0003] 在病理影像分析中,理想情况下,使用越多标注精良的样本对模型进行训练,模型可以取得越好的效果。但是在实际研发和生产过程中,想要取得更多的标注样本,往往需要许多位医生对数据进行标注。受限于医生们的水平层次不齐且判读标准有主观性,这样标注出来的训练集往往存在不少的错误标注的样本。例如同一个病变细胞,经常会出现不同医生将其划分到不同类别的病变级别上,甚至标准较为宽松的医生会将其标注为正常细胞。这些错误的标注又被称为标签噪声,存在标签噪声的数据集会极大的降低神经网络的训练效果。这些错误标注的样本会逐渐主导模型的训练方向。强行使模型学习这些带有标签噪声的样本不仅会学习到一些错误的标签,而且也会影响模型在本已可正确分类的数据上的准确性。如何提高传统病理图像样本的神经网络模型训练的准确率是亟待解决的问题。

### 发明内容

[0004] 基于此,有必要针对上述技术问题,提供一种能够改善标签噪声样本、提高神经网络模型训练精度的病理图像样本的神经网络模型训练方法、装置。

[0005] 为了实现上述目的,本发明通过以下技术方案实现:

[0006] 一种病理图像样本的神经网络模型训练方法,包括:

[0007] 将带有初始标签的病理图像样本输入初始神经网络模型,获得所述病理图像样本的初始预测概率;其中,所述病理图像样本属于病理图像样本集;

[0008] 根据所述初始预测概率计算对于多分类任务的初始交叉熵损失函数;

[0009] 对所述初始交叉熵损失函数求相对于深度神经网络模型函数的梯度,得到初始梯度模长;

[0010] 根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度;

[0011] 根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数;

[0012] 根据所述新型损失函数,采用梯度下降法对所述初始神经网络模型的参数进行更

新,获得修正神经网络模型;

[0013] 根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签;

[0014] 将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型。

[0015] 在其中一个实施例中,所述根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度,包括:计算病理图像样本集中的每个所述病理图像样本的初始梯度模长的统计分布;将所述初始梯度模长的取值均分为多个区域,并统计每个区域内的样本个数;计算所述病理图像样本个数与所述区间的长度之比,得到所述病理图像样本在所述区间的初始梯度密度。

[0016] 在其中一个实施例中,所述根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数,包括:计算所述初始梯度密度的倒数和每个所述区域的样本个数之积,得到每个所述区域的所述病理图像样本的初始第一加权因子;根据所述病理图像样本的初始梯度模长,计算所述病理图像样本的初始第二加权因子;根据所述初始第一加权因子、初始第二加权因子、初始交叉熵损失函数和所述样本个数,构建新型损失函数。

[0017] 在其中一个实施例中,在根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签之前,包括:根据所述病理图像样本的初始标签的类别,定义每个初始标签的独热标签向量。

[0018] 在其中一个实施例中,所述将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型,包括:将带有所述修正标签的病理图像样本输入所述修正神经网络模型,获得所述病理图像样本的修正预测概率;其中,所述病理图像样本属于病理图像样本集;根据所述修正预测概率计算对于多分类任务的修正交叉熵损失函数;对所述修正交叉熵损失函数求相对于深度神经网络模型函数的梯度,得到修正梯度模长;根据病理图像样本集中所述病理图像样本的所述修正梯度模长的统计分布,计算修正梯度密度;根据所述修正梯度密度、修正第二加权因子和所述修正交叉熵损失函数,构建修正损失函数;根据所述修正损失函数,采用梯度下降法对所述修正神经网络模型的参数进行更新,再次获得修正神经网络模型;根据独热标签向量、所述修正梯度模长、修正参数和所述修正预测概率,再次计算所述病理图像样本的修正标签;重复上述步骤,直到所述修正神经网络模型收敛,获得最终神经网络模型。

[0019] 在其中一个实施例中,所述根据病理图像样本集中所述病理图像样本的所述修正梯度模长的统计分布,计算修正梯度密度,包括:计算病理图像样本集中的每个所述病理图像样本的修正梯度模长的统计分布;将所述修正梯度模长的取值均分为多个区域,并统计每个区域内的样本个数;计算所述病理图像样本个数与所述区间的长度之比,得到所述病理图像样本在所述区间的修正梯度密度。

[0020] 在其中一个实施例中,所述根据所述修正梯度密度、修正第二加权因子和所述修正交叉熵损失函数,构建修正损失函数,包括:计算所述修正梯度密度的倒数和每个所述区

域的样本个数之积,得到每个所述区域的所述病理图像样本的修正第一加权因子;根据所述病理图像样本的修正梯度模长,计算所述病理图像样本的修正第二加权因子;根据所述修正第一加权因子、修正第二加权因子、修正交叉熵损失函数和所述样本个数,构建新型损失函数。

[0021] 一种病理图像样本的神经网络模型训练装置,包括:

[0022] 初始预测概率计算模块,用于将带有初始标签的病理图像样本输入初始神经网络模型,获得所述病理图像样本的初始预测概率;其中,所述病理图像样本属于病理图像样本集;

[0023] 初始交叉熵损失函数计算模块,用于根据所述初始预测概率计算对于多分类任务的初始交叉熵损失函数;

[0024] 初始梯度模长计算模块,用于对所述初始交叉熵损失函数求相对于深度神经网络模型函数的梯度,得到初始梯度模长;

[0025] 初始梯度密度计算模块,用于根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度;

[0026] 新型损失函数构建模块,用于根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数;

[0027] 参数更新模块,用于根据所述新型损失函数,采用梯度下降法对所述初始神经网络模型的参数进行更新,获得修正神经网络模型;

[0028] 修正标签计算模块,用于根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签;

[0029] 模型训练模块,用于将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型。

[0030] 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以下步骤:

[0031] 将带有初始标签的病理图像样本输入初始神经网络模型,获得所述病理图像样本的初始预测概率;其中,所述病理图像样本属于病理图像样本集;

[0032] 根据所述初始预测概率计算对于多分类任务的初始交叉熵损失函数;

[0033] 对所述初始交叉熵损失函数求相对于深度神经网络模型函数的梯度,得到初始梯度模长;

[0034] 根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度;

[0035] 根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数;

[0036] 根据所述新型损失函数,采用梯度下降法对所述初始神经网络模型的参数进行更新,获得修正神经网络模型;

[0037] 根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签;

[0038] 将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神

神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型。

[0039] 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以下步骤:

[0040] 将带有初始标签的病理图像样本输入初始神经网络模型,获得所述病理图像样本的初始预测概率;其中,所述病理图像样本属于病理图像样本集;

[0041] 根据所述初始预测概率计算对于多分类任务的初始交叉熵损失函数;

[0042] 对所述初始交叉熵损失函数求相对于深度神经网络模型函数的梯度,得到初始梯度模长;

[0043] 根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度;

[0044] 根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数;

[0045] 根据所述新型损失函数,采用梯度下降法对所述初始神经网络模型的参数进行更新,获得修正神经网络模型;

[0046] 根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签;

[0047] 将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型。

[0048] 本发明的有益效果如下:

[0049] 上述病理图像样本的神经网络模型训练方法、装置、计算机设备和存储介质,通过独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,不断对病理图像样本的标签进行修正,使得噪声标签对模型的训练影响逐渐减小,保证训练的最终神经网络模型能够准确的对病理图像进行分类。

## 附图说明

[0050] 图1为一个实施例中病理图像样本的神经网络模型训练方法的流程示意图;

[0051] 图2为一个实施例中梯度模长的统计分布图;

[0052] 图3为一个实施例中病理图像样本的神经网络模型训练装置的结构示意图;

[0053] 图4为一个实施例中实施例中计算机设备的内部结构图。

## 具体实施方式

[0054] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0055] 本发明提供的病理图像样本的神经网络模型训练方法,可以应用于带有噪声标签样本的病理图像样本训练的应用环境中。

[0056] 如图1所示,该方法包括:



[0057] S110,将带有初始标签的病理图像样本输入初始神经网络模型,获得所述病理图像样本的初始预测概率;其中,所述病理图像样本属于病理图像样本集。

[0058] 其中,病理图像是从人体或者其它动物身体的病变组织的图像,通过病理图像能够判断是否产生病变。病理图像样本是用于对病理图像分类的神经网络模型的样本,每个病理图像样本具有初始标签,每个不同的初始标签具有对应的值,例如,初始标签A对应的标签的值为1,初始标签B对应的标签的值为2,初始标签C对应的标签的值为3,初始标签的值可以根据具体情况进行定义。对于一个有M类初始标签的分类任务来说,首先定义带有标签噪声的病理图像样本集一个有N个病理图像样本,样本集可由  $D = \{x_i, y_i\}_{i=1}^N$  来表示,其中,  $x_i$  表示数据集中第i个病理图像样本,而  $y_i$  表示第i个病理图像样本的标签,  $y_i$  一共有M个类别的值,定义初始神经网络模型为  $f(x, \Theta)$ ,将病理图像样本  $x_i$  输入到初始神经网络模型得到初始预测概率  $\hat{y}_i$ 。由于涉及到多分类任务,初始预测概率可表示为  $\hat{y}_i = \text{soft max}(f(x_i, \Theta))$ ,其中, softmax函数的公式为:

$$[0059] \quad P_i = \frac{e^{a_i}}{\sum_{j=1}^M e^{a_j}}, i \in M;$$

[0060] 其中,  $a_j$  表示初始神经网络模型的原始输出。

[0061] S120,根据所述初始预测概率计算对于多分类任务的初始交叉熵损失函数。

[0062] 其中,每个初始预测概率对应一个初始交叉熵损失函数,对于多分类任务的初始交叉熵损失函数可表示为:

$$[0063] \quad l_{CE} = -\sum_{i=0}^N y_i \log(\hat{y}_i) = -\log(\hat{y}_k)$$

[0064] 其中,  $y_i$  表示第i个病理图像样本的标签,  $\hat{y}_i$  表示初始预测概率,  $\hat{y}_k$  表示第k类标签的值,  $k \in \{1, 2, \dots, M\}$ 。

[0065] S130,对所述初始交叉熵损失函数求相对于神经网络模型函数的梯度,得到初始梯度模长。

[0066] 其中,神经网络模型函数可表示为  $f(x_i, \Theta)$ 。初始梯度模长的取值范围为[0, 1]。

[0067] S140,根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度。

[0068] 其中,病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布如图2所示,在初始梯度模长的取值两端,样本所占比例高。梯度模长越小,表示该样本越是易分的,相反的,如果梯度模长约接近于1,说明该样本越可能是标签错误的样本。

[0069] S150,根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数。

[0070] 其中,根据初始梯度密度计算所述病理图像样本的初始第一加权因子,通过对病理图像样本集中所有的病理图像样本,计算初始第一加权因子、初始第二加权因子和所述初始交叉熵损失函数之积的和,并求平均值,得到新型损失函数,可以表示为:

$$[0071] \quad l = \frac{1}{N} \sum_{i=1}^N \alpha_i \beta_i l_{CE}(y_i, \hat{y}_i)$$

[0072] 其中,  $N$ 为样本个数,  $\alpha_i$ 为初始第一加权因子,  $\beta_i$ 为初始第二加权因子,  $l_{CE}$ 为初始交叉熵损失函数。

[0073] S160, 根据所述新型损失函数, 采用梯度下降法对所述初始神经网络模型的参数进行更新, 获得修正神经网络模型。

[0074] S170, 根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率, 计算所述病理图像样本的修正标签。

[0075] 其中, 独热标签向量根据初始标签的类别进行定义, 例如, 在  $M$  个类别的标签中, 初始标签的类别为 2, 则独热标签为  $y_{hot_i} = [0, 1, \dots, 0, 0]$ ,  $y_{hot_i}$  中有  $M$  个值, 当初始标签的类别为 3, 则独热标签  $y_{hot_i} = [0, 0, 1, \dots, 0, 0]$ ,  $y_{hot_i}$  中有  $M$  个值。修正标签的计算公式可为:

$$[0076] \quad y_{refine_i} = (1 - g_i^\gamma) * y_{hot_i} + g_i^\gamma * (\hat{y}_i - y_{hot_i})$$

[0077] 其中,  $y_{refine_i}$  表示修正标签,  $g_i$  表示梯度模长,  $\gamma$  为修正参数,  $g_i^\gamma$  用于控制修正的力度。对于梯度模长较大的样本, 则修正力度较大, 反之修正力度较小。对于  $g \rightarrow 0$  的样本, 其修正力度约为 0。使用此方法, 在迭代过程中, 逐步将带有标签噪声的样本拉回到了其正确的标签上。

[0078] S180, 将带有所述修正标签的病理图像样本输入所述修正神经网络模型, 对所述修正神经网络模型进行训练, 更新所述修正神经网络模型的参数, 在所述修正神经网络模型收敛时, 获得最终神经网络模型。

[0079] 理想情况下, 病理图像样本  $x_i$  样本集、样本标签集  $y_i$  的标注越精良、准确, 模型训练的效果越好。但是在实际研发和生产过程中, 想要获得更多的标注样本, 往往需要许多位医生对病理图像的数据进行标注, 人为主观标注出来的病理图像样本集存在不少的错误标注样本, 例如, 将样本病理细胞的标注等级划分错误, 甚至可能将错误标签样本标注为正常细胞样本。此处, 将将错误标注的病理图像样本定义为噪声标签样本。例如在宫颈液基细胞 (TCT) 检测项目中, 一张数字病理全场图中可能只存在 1% 的细胞为阳性病变细胞, 其余 99% 的细胞均为阴性细胞, 使用这种极度不均衡的样本来训练神经网络模型, 模型的训练方向会完全被占绝大多数的阴性样本主导, 即使模型将所有的细胞均预测为阴性, 这样也会产生 99% 的准确率, 然而人工智能辅助诊断的目的主要是检测阳性细胞, 所以这显然不是所被希望的。

[0080] 模型训练时, 噪声标签样本的数据集会极大的降低神经网络的训练效果。尤其是在使用 Focal Loss 的时候, Focal Loss 会压低易分样本对模型的影响, 使模型训练的重心在难分样本上, 而这些噪声标签样本会成为极度难分样本, 逐渐主导模型的训练方向。强行使模型学习这些带有标签噪声的样本不仅会学习到一些错误的标签参数, 而且也会影响模型训练在本已可正确分类的样本数据上的准确性。

[0081] 上述病理图像样本的神经网络模型训练方法, 通过独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率, 不断对病理图像样本的标签进行修正, 使得噪声标签对模型的训练影响逐渐减小, 保证训练的最终神经网络模型能够准确的对病理图像进行分类。

[0082] 在其中一个实施例中, 步骤 S130 具体过程为:

[0083] 对于多分类任务的初始交叉熵损失函数可表示为:

$$[0084] \quad l_{CE} = -\sum_{i=0}^N y_i \log(\hat{y}_i) = -\log(\hat{y}_k)$$

[0085] 其中,  $y_i$  表示第  $i$  个病理图像样本的标签,  $\hat{y}_i$  表示初始预测概率,  $\hat{y}_k$  表示第  $k$  类标签的值,  $k \in \{1, 2, \dots, M\}$ 。

[0086] 求  $l_{CE}$  对深度神经网络模型函数  $f(x_i, \Theta)$  的梯度(导数)可得:

$$\begin{aligned} [0087] \quad \frac{\partial l_{CE}}{\partial f(x_i, \Theta)} &= -\frac{\sum_{j=1}^M y_j \log(\hat{y}_j)}{\partial f(x_i, \Theta)} \\ &= -\sum_{j=1}^M y_j \frac{\partial \log(\hat{y}_j)}{\partial \hat{y}_j} \frac{\partial \hat{y}_j}{\partial f(x_i, \Theta)} \\ &= -\sum_{j=1}^M y_j \frac{1}{\hat{y}_j} \frac{\partial \hat{y}_j}{\partial f(x_i, \Theta)} \end{aligned}$$

[0088] 根据  $\hat{y}_i = \text{soft max}(f(x_i, \Theta))$  可以推算第  $j$  个输出对第  $i$  个输入的偏导数可以定义为:

$$\begin{aligned} [0089] \quad \frac{\partial \hat{y}_j}{\partial a_i} &= \frac{\partial \frac{e^{a_j}}{\sum_{k=1}^M e^{a_k}}}{\partial a_i} \\ &= \frac{\frac{\partial e^{a_j}}{\partial a_i} \cdot \sum_{k=1}^M e^{a_k} - \frac{\partial \sum_{k=1}^M e^{a_k}}{\partial a_i} \cdot e^{a_j}}{(\sum_{k=1}^M e^{a_k})^2} \\ &= \frac{\frac{\partial e^{a_j}}{\partial a_i} \cdot \sum_{k=1}^M e^{a_k} - e^{a_i} \cdot e^{a_j}}{(\sum_{k=1}^M e^{a_k})^2} \end{aligned}$$

[0090] 其中当  $i = j$  时可得:

$$[0091] \quad \frac{\partial e^{a_j}}{\partial a_i} = e^{a_j}$$

[0092] 于是有:

$$[0093] \quad \frac{\partial \hat{y}_j}{\partial a_i} = \frac{e^{a_j} \cdot \sum_{k=1}^M e^{a_k} - e^{a_i} \cdot e^{a_j}}{(\sum_{k=1}^M e^{a_k})^2} = \frac{\sum_{k=1}^M e^{a_k} - e^{a_i}}{\sum_{k=1}^M e^{a_k}} \frac{e^{a_j}}{\sum_{k=1}^M e^{a_k}} = (1 - \hat{y}_i) \hat{y}_j$$

[0094] 而当  $i \neq j$  时可得:

$$[0095] \quad \frac{\partial e^{a_j}}{\partial a_i} = 0$$

[0096] 于是有:

$$[0097] \quad \frac{\partial \hat{y}_j}{\partial a_i} = \frac{0 \cdot \sum_{k=1}^M e^{a_k} - e^{a_i} \cdot e^{a_j}}{(\sum_{k=1}^M e^{a_k})^2} = -\frac{e^{a_i}}{\sum_{k=1}^M e^{a_k}} \frac{e^{a_j}}{\sum_{k=1}^M e^{a_k}} = -\hat{y}_i \hat{y}_j$$

[0098] 将上述结果带入:

$$\begin{aligned}
\frac{\partial l_{CE}}{\partial f(x_i, \Theta)} &= -\sum_{j=1}^M y_j \frac{1}{\hat{y}_j} \frac{\partial \hat{y}_j}{\partial f(x_i, \Theta)} \\
&= -y_j \frac{1}{\hat{y}_j} (1 - \hat{y}_i) \hat{y}_j - \sum_{i \neq j}^M y_i \frac{1}{\hat{y}_i} (-\hat{y}_j \hat{y}_i) \\
[0099] \quad &= -y_j (1 - \hat{y}_i) + \sum_{i \neq j}^M y_i \hat{y}_j \\
&= -y_j + \sum_{i=1}^M y_i \hat{y}_j \\
&= \hat{y}_j - y_j
\end{aligned}$$

[0100] 由此可以定义一个初始梯度模长为  $g = |\hat{y} - y|$ ，由于  $\hat{y} \in (0,1)$  且  $y = 0$  或  $1$ ，故而  $g \in (0,1)$ 。

[0101] 在其中一个实施例中，所述根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布，计算初始梯度密度，包括：计算病理图像样本集中的每个所述病理图像样本的初始梯度模长的统计分布；将所述初始梯度模长的取值均分为多个区域，并统计每个区域内的样本个数；计算所述病理图像样本个数与所述区间的长度之比，得到所述病理图像样本在所述区间的初始梯度密度。

[0102] 其中，如图2所示，每个病理图像样本具有对应的初始梯度模长，初始梯度模长取值范围为0到1，梯度模长越小，表示该样本越是易分的，相反的，如果梯度模长约接近于1，说明该样本越可能是标签错误的样本。由图2可见，绝大多数的病理图像样本位于图标的左侧，且这些样本是易分样本，在绝对数量的优势之下，这些样本产生的梯度会对模型训练产生显著的影响；而位于图标右端也有大量的样本，这些样本的初始梯度模长较大，这些样本可能存在大量的标签噪声，本身就大的初始梯度模长在数量的加持下，会对模型训练产生非常负面的影响，基于上述观察，通过上述实施例中新型损失函数会尽量弱化样本数量对梯度的干扰且减弱初始梯度模长很小（易分样本）和初始梯度模长很大（噪声标签样本）对训练的影响。首先我们引入初始梯度密度的概念，将初始梯度模长的取值范围划分为  $\varepsilon$  个区域，接着分别统计落于各个区域内的样本个数，对于一个初始梯度模长为  $g$  的样本，它的初始梯度密度就定义为落于它的初始梯度模长所在区间的总样本数量除以区间的长度  $\frac{1}{\varepsilon}$ ，初

始梯度密度的公式可表示为： $GD(g) = R_{ind}(g) \varepsilon$ 。

[0103] 在其中一个实施例中，所述根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数，构建新型损失函数，包括：计算所述初始梯度密度的倒数和每个所述区域的样本个数之积，得到每个所述区域的所述病理图像样本的初始第一加权因子；根据所述病理图像样本的初始梯度模长，计算所述病理图像样本的初始第二加权因子；根据所述初始第一加权因子、初始第二加权因子、初始交叉熵损失函数和所述样本个数，构建新型损失函数。

[0104] 其中，初始第一加权因子的计算公式为： $\alpha_i = \frac{N}{GD(g_i)}$ ； $N$ 为样本个数， $GD(g_i)$ 为第  $i$  个病理图像样本对应的初始梯度密度， $g_i$ 为第  $i$  个病理图像样本对应的初始梯度模长。

[0105] 其中,初始第二加权因子的计算公式可为:  $\beta_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(g_i-\mu)^2}{2\sigma^2}}$ ; 其中,  $\beta_i$  为初始第二

加权因子,  $\sigma$  为标准差, 其计算公式为  $\sigma = \sqrt{\frac{\sum_{i=1}^N (g_i - \mu)^2}{N-1}}$ ,  $g_i$  为初始梯度模长,  $\mu$  为梯度学习中心,  $N$  为样本个数, 如果  $g_i$  越接近  $\mu$ , 则其梯度对模型训练影响越大, 相反的, 如果  $g_i$  越远离  $\mu$ , 则其对模型训练影响越小。

[0106] 其中, 新型损失函数的计算公式可为:

$$[0107] \quad l = \frac{1}{N} \sum_{i=1}^N \alpha_i \beta_i l_{CE}(y_i, \hat{y}_i)$$

[0108] 其中,  $N$  为样本个数,  $\alpha_i$  为初始第一加权因子,  $\beta_i$  为初始第二加权因子,  $l_{CE}$  为初始交叉熵损失函数。

[0109] 在其中一个实施例中, 在根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率, 计算所述病理图像样本的修正标签之前, 包括: 根据所述病理图像样本的初始标签的类别, 定义每个初始标签的独热标签向量。

[0110] 其中, 独热标签向量根据初始标签的类别进行定义, 例如, 在  $M$  个类别的标签中, 初始标签的类别为 2, 则独热标签为  $y_{hot_i} = [0, 1, \dots, 0, 0]$ ,  $y_{hot_i}$  中有  $M$  个值, 当初始标签的类别为 3, 则独热标签  $y_{hot_i} = [0, 0, 1, \dots, 0, 0]$ ,  $y_{hot_i}$  中有  $M$  个值。修正标签的计算公式可为:

$$[0111] \quad y_{refine_i} = (1 - g_i^\gamma) * y_{hot_i} + g_i^\gamma * (\hat{y}_i - y_{hot_i})$$

[0112] 其中,  $y_{refine_i}$  表示修正标签,  $g_i$  表示梯度模长,  $\gamma$  为修正参数,  $g_i^\gamma$  用于控制修正的力度。对于梯度模长较大的样本, 则修正力度较大, 反之修正力度较小。对于  $g \rightarrow 0$  的样本, 其修正力度约为 0。使用此方法, 在迭代过程中, 逐步将带有标签噪声的样本拉回到了其正确的标签上。

[0113] 在其中一个实施例中, 所述将带有所述修正标签的病理图像样本输入所述修正神经网络模型, 对所述修正神经网络模型进行训练, 更新所述修正神经网络模型的参数, 在所述修正神经网络模型收敛时, 获得最终神经网络模型, 包括: 将带有所述修正标签的病理图像样本输入所述修正神经网络模型, 获得所述病理图像样本的修正预测概率; 其中, 所述病理图像样本属于病理图像样本集; 根据所述修正预测概率计算对于多分类任务的修正交叉熵损失函数; 对所述修正交叉熵损失函数求相对于深度神经网络模型函数的梯度, 得到修正梯度模长; 根据病理图像样本集中所述病理图像样本的所述修正梯度模长的统计分布, 计算修正梯度密度; 根据所述修正梯度密度、修正第二加权因子和所述修正交叉熵损失函数, 构建修正损失函数; 根据所述修正损失函数, 采用梯度下降法对所述修正神经网络模型的参数进行更新, 再次获得修正神经网络模型; 根据独热标签向量、所述修正梯度模长、修正参数和所述修正预测概率, 再次计算所述病理图像样本的修正标签; 重复上述步骤, 直到所述修正神经网络模型收敛, 获得最终神经网络模型。

[0114] 其中, 修正交叉熵损失函数的计算过程与初始交叉熵损失函数的计算过程相同, 修正梯度模长、修正梯度密度的计算方法可以参照上述实施例中初始梯度模长、初始梯度密度的计算方法, 修正损失函数的构建方法可以参照上述实施例中新型损失函数的计算方

法。

[0115] 在其中一个实施例中,所述根据病理图像样本集中所述病理图像样本的所述修正梯度模长的统计分布,计算修正梯度密度,包括:计算病理图像样本集中的每个所述病理图像样本的修正梯度模长的统计分布;将所述修正梯度模长的取值均分为多个区域,并统计每个区域内的样本个数;计算所述病理图像样本个数与所述区间的长度之比,得到所述病理图像样本在所述区间的修正梯度密度。

[0116] 在其中一个实施例中,所述根据所述修正梯度密度、修正第二加权因子和所述修正交叉熵损失函数,构建修正损失函数,包括:计算所述修正梯度密度的倒数和每个所述区域的样本个数之积,得到每个所述区域的所述病理图像样本的修正第一加权因子;根据所述病理图像样本的修正梯度模长,计算所述病理图像样本的修正第二加权因子;根据所述修正第一加权因子、修正第二加权因子、修正交叉熵损失函数和所述样本个数,构建新型损失函数。

[0117] 其中,修正第一加权因子的计算过程可参加上述实施例中初始第一加权因子的计算过程。

[0118] 应该理解的是,虽然图1的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,图1中的至少一部分步骤可以包括多个步骤或者多个阶段,这些步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0119] 在其中一个实施例中,如图3所示,提供了一种病理图像样本的神经网络模型训练装置,包括:初始预测概率计算模块210、初始交叉熵损失函数计算模块220、初始梯度模长计算模块230、初始梯度密度计算模块240、新型损失函数构建模块250、参数更新模块260、修正标签计算模块270和模型训练模块280。

[0120] 初始预测概率计算模块210,用于将带有初始标签的病理图像样本输入初始神经网络模型,获得所述病理图像样本的初始预测概率;其中,所述病理图像样本属于病理图像样本集。

[0121] 初始交叉熵损失函数计算模块220,用于根据所述初始预测概率计算对于多分类任务的初始交叉熵损失函数。

[0122] 初始梯度模长计算模块230,用于对所述初始交叉熵损失函数求相对于深度神经网络模型函数的梯度,得到初始梯度模长。

[0123] 初始梯度密度计算模块240,用于根据病理图像样本集中所述病理图像样本的所述初始梯度模长的统计分布,计算初始梯度密度。

[0124] 新型损失函数构建模块250,用于根据所述初始梯度密度、初始第二加权因子和所述初始交叉熵损失函数,构建新型损失函数。

[0125] 参数更新模块260,用于根据所述新型损失函数,采用梯度下降法对所述初始神经网络模型的参数进行更新,获得修正神经网络模型。

[0126] 修正标签计算模块270,用于根据独热标签向量、所述初始梯度模长、修正参数和所述初始预测概率,计算所述病理图像样本的修正标签。

[0127] 模型训练模块280,用于将带有所述修正标签的病理图像样本输入所述修正神经网络模型,对所述修正神经网络模型进行训练,更新所述修正神经网络模型的参数,在所述修正神经网络模型收敛时,获得最终神经网络模型。

[0128] 在其中一个实施例中,所示初始梯度模长计算模块230包括:统计分布计算单元,用于计算病理图像样本集中的每个所述病理图像样本的初始梯度模长的统计分布;区域样本个数统计单元,用于将所述初始梯度模长的取值均分为多个区域,并统计每个区域内的样本个数;初始梯度密度计算单元,用于计算所述病理图像样本个数与所述区间的长度之比,得到所述病理图像样本在所述区间的初始梯度密度。

[0129] 在其中一个实施例中,新型损失函数构建模块250包括:初始第一加权因子计算单元,用于计算所述初始梯度密度的倒数和每个所述区域的样本个数之积,得到每个所述区域的所述病理图像样本的初始第一加权因子;初始第二加权因子计算单元,用于根据所述病理图像样本的初始梯度模长,计算所述病理图像样本的初始第二加权因子;新型损失函数构建单元,用于根据所述初始第一加权因子、初始第二加权因子、初始交叉熵损失函数和所述样本个数,构建新型损失函数。

[0130] 在其中一个实施例中,所述病理图像样本的神经网络模型训练装置,还包括:独热标签向量定义模块,用于根据所述病理图像样本的初始标签的类别,定义每个初始标签的独热标签向量。

[0131] 在其中一个实施例中,所述模型训练模块280包括:修正预测概率计算单元,用于将带有所述修正标签的病理图像样本输入所述修正神经网络模型,获得所述病理图像样本的修正预测概率;其中,所述病理图像样本属于病理图像样本集;修正交叉熵损失函数计算单元,用于根据所述修正预测概率计算对于多分类任务的修正交叉熵损失函数;修正梯度模长计算单元,用于对所述修正交叉熵损失函数求相对于深度神经网络模型函数的梯度,得到修正梯度模长;修正梯度密度计算单元,用于根据病理图像样本集中所述病理图像样本的所述修正梯度模长的统计分布,计算修正梯度密度;修正损失函数构建单元,用于根据所述修正梯度密度、修正第二加权因子和所述修正交叉熵损失函数,构建修正损失函数;所述参数更新模块,还用于根据所述修正损失函数,采用梯度下降法对所述修正神经网络模型的参数进行更新,再次获得修正神经网络模型;所述修正标签计算模块,还用于根据独热标签向量、所述修正梯度模长、修正参数和所述修正预测概率,再次计算所述病理图像样本的修正标签;循环单元,用于重复上述步骤,直到所述修正神经网络模型收敛,获得最终神经网络模型。

[0132] 在其中一个实施例中,所述修正梯度密度计算单元包括:统计分布计算子单元,用于计算病理图像样本集中的每个所述病理图像样本的修正梯度模长的统计分布;区域样本个数计算子单元,用于将所述修正梯度模长的取值均分为多个区域,并统计每个区域内的样本个数;修正梯度密度计算子单元,用于计算所述病理图像样本个数与所述区间的长度之比,得到所述病理图像样本在所述区间的修正梯度密度。

[0133] 在其中一个实施例中,修正损失函数构建单元包括:修正第一加权因子计算子单元,用于计算所述修正梯度密度的倒数和每个所述区域的样本个数之积,得到每个所述区域的所述病理图像样本的修正第一加权因子;修正第二加权因子计算子单元,用于根据所述病理图像样本的修正梯度模长,计算所述病理图像样本的修正第二加权因子;新型损失

函数构建子单元,用于根据所述修正第一加权因子、修正第二加权因子、修正交叉熵损失函数和所述样本个数,构建新型损失函数。

[0134] 关于病理图像样本的神经网络模型训练装置的具体限定可以参见上文中对于病理图像样本的神经网络模型训练方法的限定,在此不再赘述。上述病理图像样本的神经网络模型训练装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0135] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是服务器,其内部结构图可以如图4所示。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储病理图像样本数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种病理图像样本的神经网络模型训练方法。

[0136] 本领域技术人员可以理解,图4中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0137] 在一个实施例中,还提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现上述各方法实施例中的步骤。

[0138] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现上述各方法实施例中的步骤。

[0139] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory,ROM)、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(Static Random Access Memory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。

[0140] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0141] 以上实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。



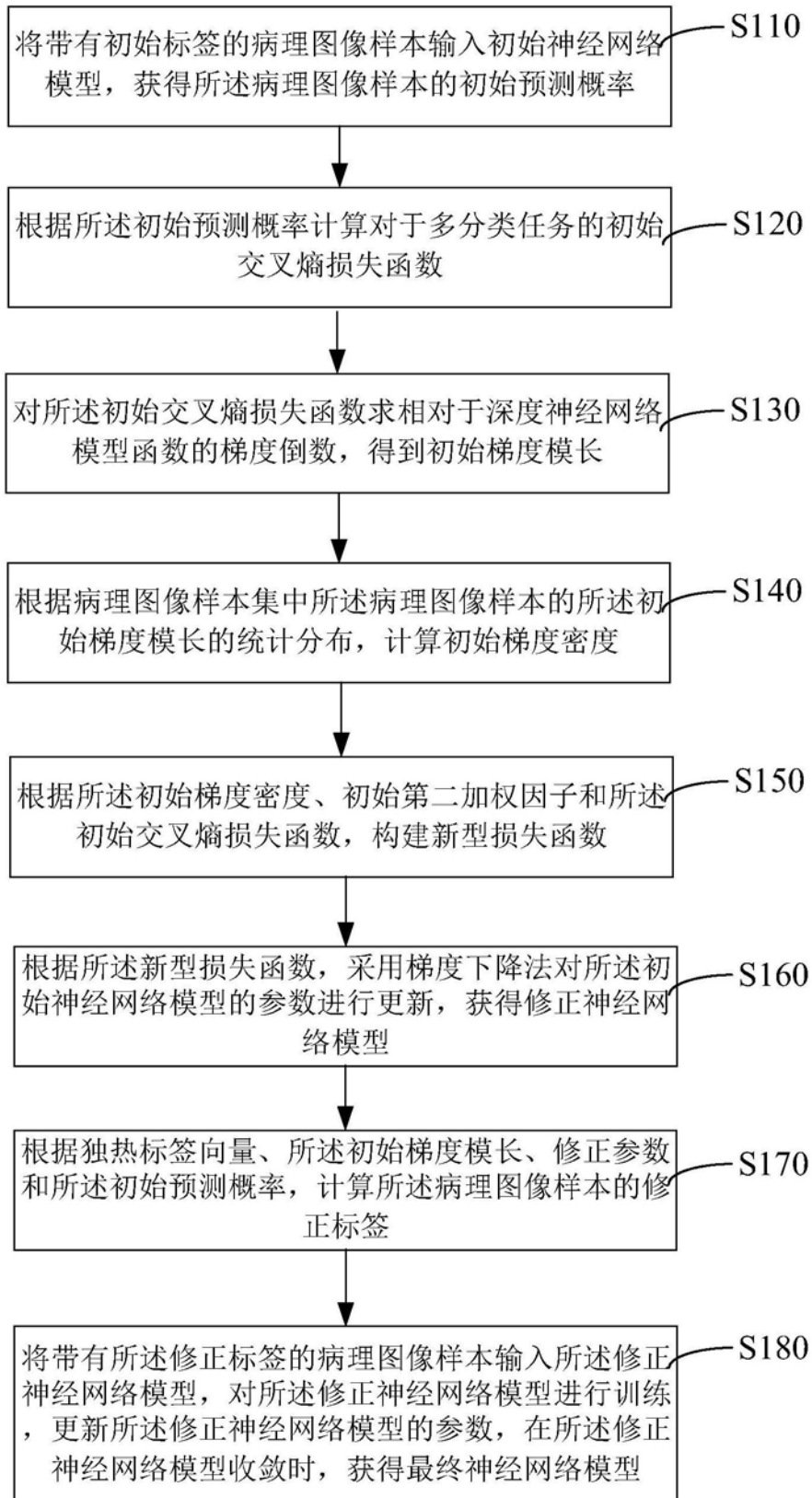


图1

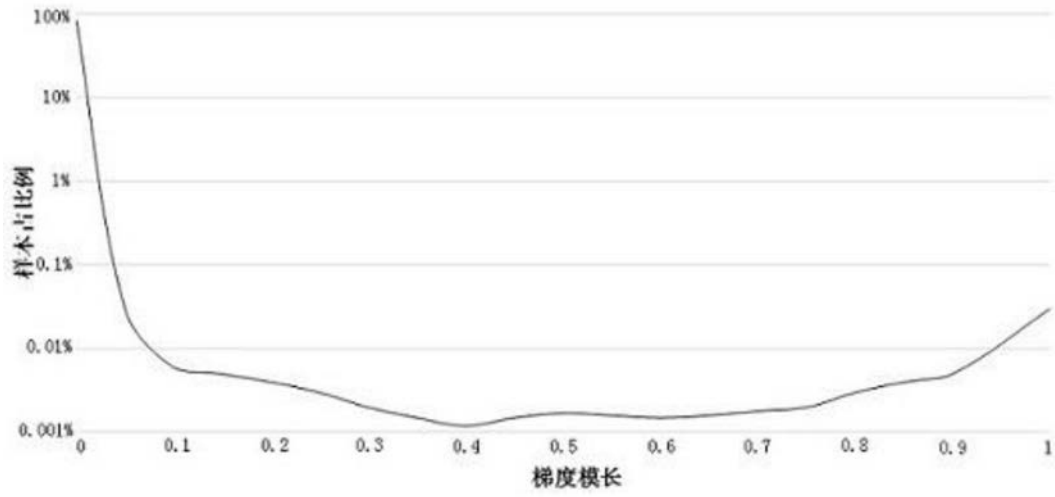


图2

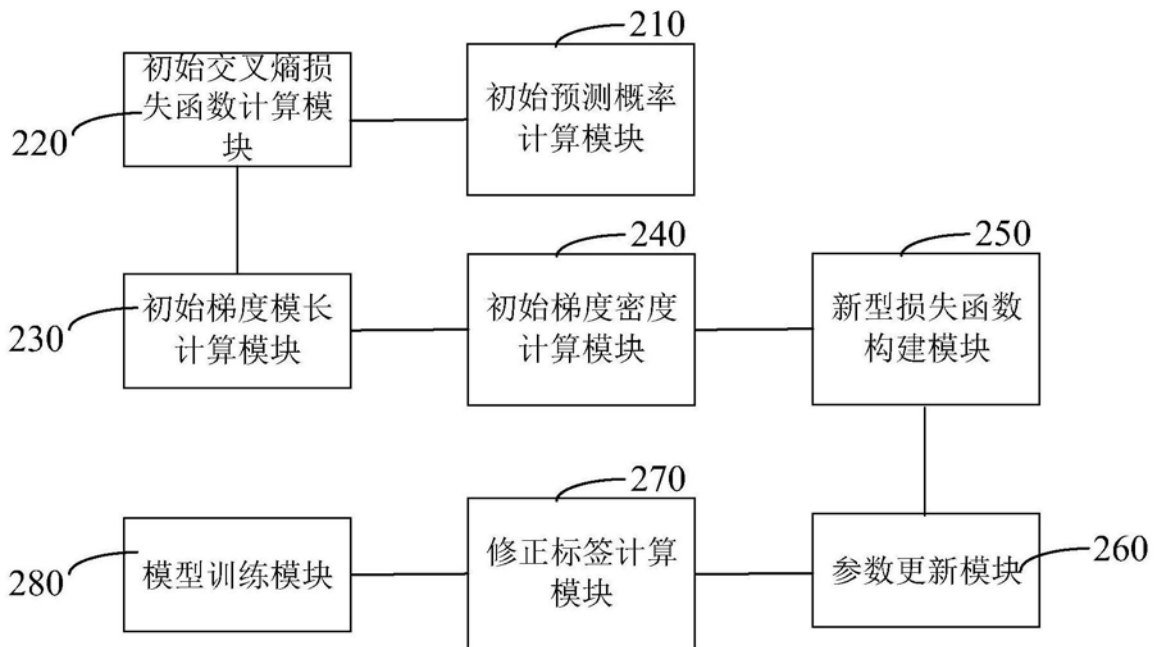


图3

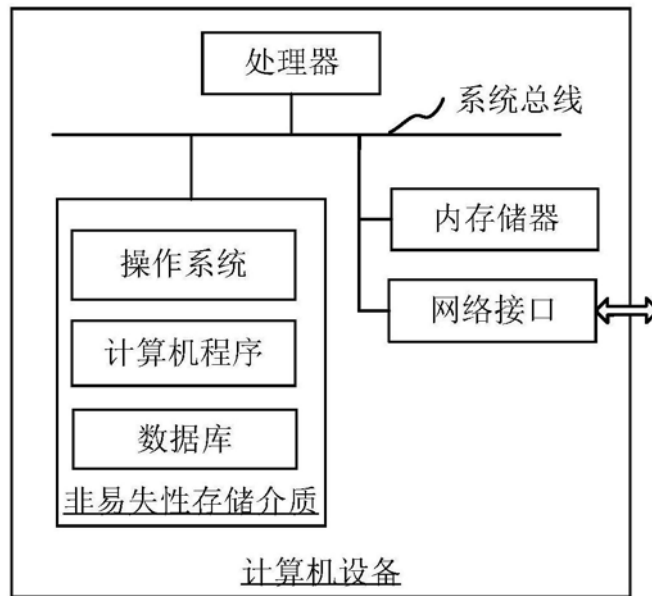


图4