



(12) 发明专利

(10) 授权公告号 CN 114898426 B

(45) 授权公告日 2024. 08. 09

(21) 申请号 202210413353.9

(22) 申请日 2022.04.20

(65) 同一申请的已公布的文献号

申请公布号 CN 114898426 A

(43) 申请公布日 2022.08.12

(73) 专利权人 国网智能电网研究院有限公司

地址 102209 北京市昌平区未来科技城滨河大道18号

专利权人 北京邮电大学

国网江苏省电力有限公司信息通信分公司

国家电网有限公司

(72) 发明人 蒋静 冯方向 许梦晗 朱力鹏

周爱华 潘森 乔俊峰

(74) 专利代理机构 北京三聚阳光知识产权代理

有限公司 11250

专利代理师 林韵英

(51) Int. Cl.

G06V 40/16 (2022.01)

G06V 10/74 (2022.01)

G06V 10/762 (2022.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

G06N 3/0464 (2023.01)

G06N 3/084 (2023.01)

G06N 20/00 (2019.01)

(56) 对比文件

Menghan Xu等.Multimodal tag synset induction.《Proc. SPIE 12294, 7th International Symposium on Advances in Electrical, Electronics, and Computer Engineering》.2022,第12294卷122945W.

审查员 梁清粼

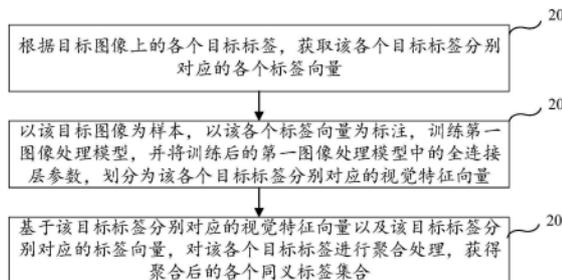
权利要求书3页 说明书13页 附图3页

(54) 发明名称

一种同义标签聚合方法、装置、设备及存储介质

(57) 摘要

本申请是关于一种同义标签聚合方法、装置、设备及存储介质,具体涉及计算机视觉技术领域。所述方法包括:根据目标图像上的各个目标标签,获取各个目标标签分别对应的各个标签向量;以目标图像为样本,以各个标签向量为标注,训练第一图像处理模型,并将训练后的第一图像处理模型中的全连接层参数,划分为各个目标标签分别对应的视觉特征向量;基于目标标签分别对应的视觉特征向量以及目标标签分别对应的标签向量,对所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。上述方案同时考虑到了目标标签的分布式表示以及目标标签在图像中含义,从而提高了同义的目标标签的聚合准确性。



1. 一种同义标签聚合方法,其特征在于,所述方法包括:

根据目标图像上的各个目标标签,获取所述各个目标标签分别对应的各个标签向量;

以所述目标图像为样本,以所述各个标签向量为标注,训练第一图像处理模型,并将训练后的第一图像处理模型中的全连接层参数,划分为所述各个目标标签分别对应的视觉特征向量;

基于所述目标标签分别对应的视觉特征向量以及所述目标标签分别对应的标签向量,对所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合;

其中,所述根据目标图像上的各个目标标签,获取所述各个目标标签分别对应的各个标签向量,包括:

获取所述目标图像上的各个目标标签;

通过词向量生成模型,对所述各个目标标签进行处理,获取所述各个目标标签分别对应的各个标签向量;所述全连接层参数为以目标数量为第一维度、以所述各个标签向量的数量为第二维度的矩阵;

所述将训练后的第一图像处理模型中的全连接层参数,划分为所述各个目标标签分别对应的视觉特征向量,包括:

将所述全连接层参数按照第二维度,划分为长度为目标数量,且分别与所述各个目标标签对应的视觉特征向量;

所述基于所述目标标签分别对应的视觉特征向量以及所述目标标签分别对应的标签向量,对所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合,包括:

针对任意两个目标标签,计算所述两个目标标签的标签向量之间的第一标签向量距离、所述两个目标标签分别对应的视觉向量之间的第一视觉向量距离以及所述两个目标标签的标签向量之间的第一正规化编辑距离;

根据所述第一标签向量距离、所述第一视觉向量距离以及所述第一正规化编辑距离,得到目标标签之间的标签距离;

根据所述标签距离,将所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。

2. 根据权利要求1所述的方法,其特征在于,所述以所述目标图像为样本,以所述各个标签向量为标注,训练第一图像处理模型,包括:

将所述目标图像通过第一图像处理模型进行处理,获得所述目标图像对应的各个预测结果;

根据所述目标图像的各个预测结果,与所述目标图像上的各个目标标签,通过损失函数进行处理,获得目标图像的损失函数值;

根据所述目标图像的损失函数值,对所述第一图像处理模型的参数进行反向传播更新。

3. 根据权利要求1或2所述的方法,其特征在于,所述基于所述目标标签分别对应的视觉特征向量以及所述目标标签分别对应的标签向量,对所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合,包括:

基于所述目标标签分别对应的视觉特征向量以及所述目标标签分别对应的标签向量,对所述目标标签执行第一聚合操作,获得各个第一类标签集合;

在所述各个第一类标签集合中选取标注标签集合;所述标注标签集合中的第一聚合标签中存在含义标注;所述含义标注用于区分不同语义的第一聚合标签;

以所述标注标签集合中的第一聚合标签为样本,以及第一聚合标签的含义标注为标注,对目标回归模型进行训练,获得训练后的目标回归模型;

根据所述训练后的目标回归模型,计算所述各个目标标签之间的标签距离,并根据所述各个目标标签之间的标签距离对所述各个目标标签进行第二聚合操作,获得聚合后的各个同义标签集合。

4. 根据权利要求3所述的方法,其特征在于,所述以所述标注标签集合中的第一聚合标签为样本,以及第一聚合标签的含义标注为标注,对目标回归模型进行训练,包括:

获取至少两个第一聚合标签分别对应的标签向量的第二标签向量距离、所述至少两个第一聚合标签分别对应的视觉特征向量之间的第二视觉向量距离以及所述至少两个第一聚合标签分别对应的标签向量之间的第二正规化编辑距离;

将所述第二标签向量距离、第二视觉向量距离以及所述第二正规化编辑距离,根据模型权重加权后通过激活函数处理,获得所述至少两个第一聚合标签之间的距离函数值;

根据所述距离函数值,以及所述至少两个第一聚合标签的含义标注,对所述目标回归模型中的模型权重进行更新。

5. 一种同义标签聚合装置,其特征在于,所述装置包括:

标签向量获取模块,用于根据目标图像上的各个目标标签,获取所述各个目标标签分别对应的各个标签向量;

模型训练模块,用于以所述目标图像为样本,以所述各个标签向量为标注,训练第一图像处理模型,并将训练后的第一图像处理模型中的全连接层参数,划分为所述各个目标标签分别对应的视觉特征向量;

标签聚合模块,用于基于所述目标标签分别对应的视觉特征向量以及所述目标标签分别对应的标签向量,对所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合;

所述标签向量获取模块,还用于通过词向量生成模型,对所述各个目标标签进行处理,获取所述各个目标标签分别对应的各个标签向量;所述全连接层参数为以目标数量为第一维度、以所述各个标签向量的数量为第二维度的矩阵;

所述模型训练模块,还用于将所述全连接层参数按照第二维度,划分为长度为目标数量,且分别与所述各个目标标签对应的视觉特征向量;

所述标签聚合模块,还用于针对任意两个目标标签,计算所述两个目标标签的标签向量之间的第一标签向量距离、所述两个目标标签分别对应的视觉向量之间的第一视觉向量距离以及所述两个目标标签的标签向量之间的第一正规化编辑距离;

根据所述第一标签向量距离、所述第一视觉向量距离以及所述第一正规化编辑距离,得到目标标签之间的标签距离;

根据所述标签距离,将所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。

6. 一种计算机设备,其特征在于,所述计算机设备中包含处理器和存储器,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、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如权利要求1至4任一所述的同义标签聚合

方法。

7.一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,所述至少一条指令由处理器加载并执行以实现如权利要求1至4任一所述的同义标签聚合方法。

一种同义标签聚合方法、装置、设备及存储介质

技术领域

[0001] 本发明涉及计算机视觉技术领域,具体涉及一种同义标签聚合方法、装置、设备及存储介质。

背景技术

[0002] 在互联网中,由于不同的用户经常使用不同的词语来描述相同或相似的内容,因此会出现大量的同义标签。

[0003] 由于在机器学习过程中,图像标注模型需要建立图像和标签之间的关系,而同义标签会导致内容相似甚至相同的图片被标注成不同的标签。由于图像标注模型需要建立图像和标签之间的关系,而同义标签会导致内容相似甚至相同的图片被标注成不同的标签,显然这种一对多的噪声对图像和标签关系的建立有负面影响,因此可以将同义标签进行自动合并,以减小此类噪声。识别同类标签的关键是建模词语之间的语义相似度,如可以将标签转化为分布式表示,并通过分布式表示之间的相似性来衡量标签之间的关联程度。

[0004] 但上述方案中,仅通过分布式表示之间的相似度来衡量标签之间的关联程度的准确性较低。

发明内容

[0005] 本申请提供了一种同义标签聚合方法、装置、设备及存储介质,提高了同义标签聚合的准确性,该技术方案如下。

[0006] 一方面,提供了一种同义标签聚合方法,所述方法包括:

[0007] 根据目标图像上的各个目标标签,获取所述各个目标标签分别对应的各个标签向量;

[0008] 以所述目标图像为样本,以所述各个标签向量为标注,训练第一图像处理模型,并将训练后的第一图像处理模型中的全连接层参数,划分为所述各个目标标签分别对应的视觉特征向量;

[0009] 基于所述目标标签分别对应的视觉特征向量、以及所述目标标签分别对应的标签向量,对所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。

[0010] 再一方面,提供了一种同义标签聚合装置,所述装置包括:

[0011] 标签向量获取模块,用于根据目标图像上的各个目标标签,获取所述各个目标标签分别对应的各个标签向量;

[0012] 模型训练模块,用于以所述目标图像为样本,以所述各个标签向量为标注,训练第一图像处理模型,并将训练后的第一图像处理模型中的全连接层参数,划分为所述各个目标标签分别对应的视觉特征向量;

[0013] 标签聚合模块,用于基于所述目标标签分别对应的视觉特征向量以及所述目标标签分别对应的标签向量,对所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。

- [0014] 在一种可能的实现方式中,所述标签向量获取模块,还用于,
- [0015] 获取所述目标图像上的各个目标标签;
- [0016] 通过词向量生成模型,对所述各个目标标签进行处理,获取所述各个目标标签分别对应的各个标签向量。
- [0017] 在一种可能的实现方式中,所述全连接层参数为以目标数量为第一维度、以所述各个标签向量的数量为第二维度的矩阵;
- [0018] 所述模型训练模块,还用于,
- [0019] 将所述全连接层参数按照第二维度,划分为长度为目标数量,且分别与所述各个目标标签对应的视觉特征向量。
- [0020] 在一种可能的实现方式中,所述模型训练模块,还用于,
- [0021] 将所述目标图像通过第一图像处理模型进行处理,获得所述目标图像对应的各个预测结果;
- [0022] 根据所述目标图像的各个预测结果,与所述目标图像上的各个目标标签,通过损失函数进行处理,获得目标图像的损失函数值;
- [0023] 根据所述目标图像的损失函数值,对所述第一图像处理模型的参数进行反向传播更新。
- [0024] 在一种可能的实现方式中,所述标签聚合模块,还用于,
- [0025] 针对任意两个目标标签,计算所述两个目标标签的标签向量之间的第一标签向量距离、所述两个目标标签分别对应的视觉向量之间的第一视觉向量距离以及所述两个目标标签的标签向量之间的第一正规化编辑距离;
- [0026] 将所述第一标签向量距离、所述第一视觉向量距离以及所述第一正规化编辑距离按照目标权重加权求和,获取所述目标标签之间的距离相似度;
- [0027] 根据所述目标标签之间的距离相似度,将所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。
- [0028] 在一种可能的实现方式中,所述标签聚合模块,还用于,
- [0029] 基于所述目标标签分别对应的视觉特征向量以及所述目标标签分别对应的标签向量,对所述目标标签执行第一聚合操作,获得各个第一类标签集合;
- [0030] 在所述各个第一类标签集合中选取标注标签集合;所述标注标签集合中的第一聚合标签中存在含义标注;所述含义标注用于区分不同语义的第一聚合标签;
- [0031] 以所述标注标签集合中的第一聚合标签为样本,以及第一聚合标签的含义标注为标注,对目标回归模型进行训练,获得训练后的目标回归模型;
- [0032] 根据所述训练后的目标回归模型,计算所述各个目标标签之间的标签距离,并根据所述各个目标标签之间的标签距离对所述各个目标标签进行第二聚合操作,获得聚合后的各个同义标签集合。
- [0033] 在一种可能的实现方式中,所述标签聚合模块,还用于,
- [0034] 获取至少两个第一聚合标签分别对应的标签向量的第二标签向量距离、所述至少两个第一聚合标签分别对应的视觉特征向量之间的第二视觉向量距离以及所述至少两个第一聚合标签分别对应的标签向量之间的第二正规化编辑距离;
- [0035] 将所述第二标签向量距离、第二视觉向量距离以及所述第二正规化编辑距离,根

据所述模型权重加权后通过激活函数处理,获得所述至少两个第一聚合标签之间的距离函数值;

[0036] 根据所述距离函数值,以及所述至少两个第一聚合标签的含义标注,对所述目标回归模型中的模型权重进行更新。

[0037] 再一方面,提供了一种计算机设备,所述计算机设备中包含处理器和存储器,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、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同义标签聚合方法。

[0038] 又一方面,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,所述至少一条指令由处理器加载并执行以实现上述的同义标签聚合方法。

[0039] 再一方面,提供了一种计算机程序产品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的同义标签聚合方法。

[0040] 本申请提供的技术方案可以包括以下有益效果:

[0041] 当获取到用于执行机器学习的目标图像后,计算机设备可以先根据目标图像上的各个目标标签,获取各个标签向量作为各个目标标签的分布式表示;计算机设备再以各个标签向量为标注,以目标图像为样本,训练第一图像处理模型,并根据第一图像处理模型中的全连接层参数划分为各个目标标签分别对应的视觉特征向量,再通过视觉特征向量以及标签向量,对各个目标标签进行聚合处理,从而使得各个目标标签在聚合为同义标签集合时,同时考虑到了目标标签的分布式表示以及目标标签在图像中含义,从而提高了同义的目标标签的聚合准确性。

附图说明

[0042] 为了更清楚地说明本申请具体实施方式或现有技术中的技术方案,下面将对具体实施方式或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本申请的一些实施方式,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。

[0043] 图1是根据一示例性实施例示出的一种图像处理系统的结构示意图。

[0044] 图2是根据一示例性实施例示出的同义标签聚合方法的方法流程图。

[0045] 图3是根据一示例性实施例示出的同义标签聚合方法的方法流程图。

[0046] 图4是根据一示例性实施例示出的同义标签聚合装置的结构方框图。

[0047] 图5是根据本申请一示例性实施例提供的一种计算机设备示意图。

具体实施方式

[0048] 下面将结合附图对本申请的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,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,都属于本申请保护的范围。

[0049] 应理解,在本申请的实施例中提到的“指示”可以是直接指示,也可以是间接指示,还可以是表示具有关联关系。举例说明,A指示B,可以表示A直接指示B,例如B可以通过A获

取;也可以表示A间接指示B,例如A指示C,B可以通过C获取;还可以表示A和B之间具有关联关系。

[0050] 在本申请实施例的描述中,术语“对应”可表示两者之间具有直接对应或间接对应的关系,也可以表示两者之间具有关联关系,也可以是指示与被指示、配置与被配置等关系。

[0051] 本申请实施例中,“预定义”可以通过在设备(例如,包括终端设备和网络设备)中预先保存相应的代码、表格或其他可用于指示相关信息的方式来实现,本申请对于其具体的实现方式不做限定。

[0052] 在对本申请所示的各个实施例进行说明之前,首先对本申请涉及到的几个概念进行介绍。

[0053] 1) AI(Artificial Intelligence,人工智能)

[0054] 人工智能(Artificial Intelligence),英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学的一个分支,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,该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能从诞生以来,理论和技术日益成熟,应用领域也不断扩大,可以设想,未来人工智能带来的科技产品,将会是人类智慧的“容器”。人工智能可以对人的意识、思维的信息过程的模拟。人工智能不是人的智能,但能像人那样思考、也可能超过人的智能。

[0055] 用来研究人工智能的主要物质基础以及能够实现人工智能技术平台的机器就是计算机。除了计算机科学以外,人工智能还涉及信息论、控制论、自动化、仿生学、生物学、心理学、数理逻辑、语言学、医学和哲学等多门学科。人工智能学科研究的主要内容包括:知识表示、自动推理和搜索方法、机器学习和知识获取、知识处理系统、自然语言理解、计算机视觉、智能机器人、自动程序设计等方面。

[0056] 2) 机器学习(Machine Learning,ML)

[0057] 机器学习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。

[0058] 图1是根据一示例性实施例示出的一种图像处理系统的结构示意图。可选的,该图像处理系统中包含服务器110以及终端120。其中,终端120与服务器110之间通过通信网络进行数据通信,该通信网络可以是有线网络也可以是无网络。

[0059] 可选的,该服务器110中包括用于执行图像处理的机器学习模型,该机器学习模型可以在服务器110中通过训练图像数据训练出的机器学习模型,或者该机器学习模型(例如模型训练设备)也可以是在其他计算机设备通过训练图像数据训练出的机器学习模型。在模型训练设备上,当通过训练图像数据训练出用于执行图像处理的机器学习模型后,可以将机器学习模型的结构以及机器学习模型参数信息发送给服务器110,以便服务器110构建出用于机器学习模型以用于执行图像处理。

[0060] 由于不同的用户经常使用不同的词语来描述相同或相似的内容,因此训练图像中

会出现大量的同义标签,可选的,当服务器110作为模型训练设备接收到训练图像数据后,可以在通过训练图像数据对机器学习模型进行训练之前,先对训练图像数据进行优化,识别训练图像数据中的同义标签,并将同义标签聚合为各个同义标签集合,从而使得后续训练图像数据在识别过程中,可以将不同的标签赋予相同的语义标注,尽可能降低机器学习模型的训练过程由于不同的用户经常使用不同的词语描述相同或相似内容所造成的训练误差。

[0061] 可选的,图像处理过程可以在终端120上执行,即终端120可以接收到模型训练设备或服务器110所发送的机器学习模型的结构信息以及机器学习模型的结构信息,在终端120上构建出对应的机器学习模型。当终端120接收到图像后,可以通过应用程序,调用机器学习模型,对该图像进行图像处理(如图像分割或人脸识别),并将图像处理得到的各个数据(例如人脸图像等)发送并保存在服务器110中。

[0062] 可选的,该终端120可以是具有指令输入组件的终端设备,该指令输入组件可以包括触摸显示屏、鼠标、键盘等根据用户操作生成指令信息的组件,该用户可以通过对指令输入组件执行指定操作,控制终端120执行指定操作(如获取图像、对图像执行图像处理等操作)。

[0063] 可选的,该终端120可以是智能手机、平板电脑,膝上便携式笔记本电脑等移动终端,也可以是台式电脑、投影式电脑等终端,或是具有数据处理组件的智能终端,本申请实施例对此不设限制。

[0064] 服务器110可以实现为一台服务器,也可以实现为一组服务器构成的服务器集群,其可以是物理服务器,也可以实现为云服务器。在一种可能的实现方式中,服务器110是终端120中应用程序的后台服务器。

[0065] 可选的,上述服务器可以是独立的物理服务器,也可以是由多个物理服务器构成的服务器集群或者是分布式系统,还可以是提供云服务、云数据库、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、CDN(Content Delivery Network,内容分发网络)、以及大数据和人工智能平台等技术云计算服务的云服务器。

[0066] 可选的,该系统还可以包括管理设备,该管理设备用于对该系统进行管理(如管理各个模块与服务器之间的连接状态等),该管理设备与服务器之间通过通信网络相连。可选的,该通信网络是有线网络或无线网络。

[0067] 可选的,上述的无线网络或有线网络使用标准通信技术和/或协议。网络通常为因特网,但也可以是其他任何网络,包括但不限于局域网、城域网、广域网、移动、有限或无线网络、专用网络或者虚拟专用网络的任何组合。在一些实施例中,使用包括超文本标记语言、可扩展标记语言等的技术和/或格式来代表通过网络交换的数据。此外还可以使用诸如安全套接字层、传输层安全、虚拟专用网络、网际协议安全等常规加密技术来加密所有或者一些链路。在另一些实施例中,还可以使用定制和/或专用数据通信技术取代或者补充上述数据通信技术。

[0068] 图2是根据一示例性实施例示出的同义标签聚合方法的方法流程图。该方法由计算机设备执行,该计算机设备可以是如图1所示的图像处理系统中的服务器,如图2所示,该同义标签聚合方法可以包括如下步骤:

[0069] 步骤201,根据目标图像上的各个目标标签,获取该各个目标标签分别对应的各个

标签向量。

[0070] 在本申请实施例的一种可能的实现方式中,该目标图像可以是用于对图像处理模型进行训练的训练图像数据,因此该目标图像上存在有各个目标标签,该目标标签可以是人工预先标注的,用于作为图像处理模型的训练标注数据。

[0071] 在本申请实施例的一种可能的实现方式中,计算机设备在对图像处理模型进行训练之前,可以先获取到对应的训练数据集,该目标图像可以是该训练数据集中的任一个。

[0072] 在本申请实施例的一种可能的实现方式中,该由于目标图像上的各个目标标签均为文本形式,因此计算机设备可以直接对目标标签进行处理,从而获得目标标签对应的词向量(也就是标签向量),例如,该计算机设备可以通过SENNa、word2vec、GloVe和fastText等词向量模型获取到各个目标标签对应的标签向量。

[0073] 步骤202,以该目标图像为样本,以该各个标签向量为标注,训练第一图像处理模型,并将训练后的第一图像处理模型中的全连接层参数,划分为该各个目标标签分别对应的视觉特征向量。

[0074] 当获取到目标图像中的各个目标标签对应的标签向量后,计算机设备可以以目标图像为样本,以各个标签向量为标注,对第一图像处理模型进行训练,也就是说计算机设备将目标图像作为样本输入第一图像模型进行处理,并通过全连接层参数,得到各个预测标签向量,再与目标图像上的各个标签向量进行对比,根据误差(如损失函数)对第一图像处理模型进行更新。

[0075] 此时更新后的第一图像处理模型中的全连接层用于生成各个预测标签向量。而由于更新后的第一图像处理模型最后生成的预测标签向量是与实际的标签向量逐渐逼近的,因此更新后的第一图像处理模型中的全连接层参数也在一定程度上与标签向量有关。

[0076] 例如在一种可能的实现方式中,当全连接层的层数为 $2048*N$,此时全连接层对第一图像处理模型的卷积层所生成的特征图进行处理后,可以生成 N 个预测标签向量,此时 N 个预测标签向量分别与 N 个目标标签所对应的标签向量进行对比,从而通过损失函数进行处理,使得每次迭代训练后的预测标签向量愈发趋近于 N 个目标标签所对应的标签向量,此时全连接层的 $2048*N$ 的参数,也可以按照输出预测标签向量的顺序,拆分为 N 个视觉特征向量,此时 N 个视觉特征向量即可以分别表征从目标图像中提取出的,与 N 个目标标签分别对应的图像特点。

[0077] 步骤203,基于该目标标签分别对应的视觉特征向量以及该目标标签分别对应的标签向量,对该各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。

[0078] 在一种可能的实现方式中,针对任意两个目标标签,计算该两个目标标签的标签向量之间的第一标签向量距离、该至少两个目标标签分别对应的视觉向量之间的第一视觉向量距离以及该至少两个目标标签的标签向量之间的第一正规化编辑距离;

[0079] 将该第一标签向量距离、该第一视觉向量距离以及该第一正规化编辑距离按照目标权重加权求和,获取该目标标签之间的距离相似度;

[0080] 根据该目标标签之间的距离相似度,将该各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。

[0081] 其中,上述两个目标标签的标签向量之间的第一标签向量距离,为两个目标标签的标签向量的欧式距离;上述两个目标标签分别对应的视觉特征向量之间的第一视觉向量

距离,为视觉特征向量之间的欧式距离;上述第一正规化编辑距离可以通过如下公式计算:

$$[0082] \quad NED(t_i, t_j) = \frac{\text{EditDistance}(t_i, t_j)}{\max(\text{length}(t_i), \text{length}(t_j))}$$

[0083] $NED(t_i, t_j)$ 为目标向量 t_i 与 t_j 之间的第一正规化编辑距离, $\text{EditDistance}(t_i, t_j)$ 为目标向量 t_i 与 t_j 的最小编辑距离,也就是目标向量 t_i 编辑为目标向量 t_j 所需要使用的最小编辑次数, $\max(\text{length}(t_i), \text{length}(t_j))$ 为目标向量 t_i 与 t_j 的最大向量长度。

[0084] 可选的,在将各个目标标签进行聚合处理时,对于目标标签(以目标标签中的任一个目标标签A为例),计算机设备可以计算出该目标标签A与其他目标标签的距离相似度,并将与该目标标签A的距离相似度小于相似度阈值的目标标签获取为目标标签A同义的标签,并聚合为目标标签A的同义标签集合。

[0085] 此时计算机设备可以迭代如下步骤直至不存在可以聚合的目标标签:

[0086] 在各个目标标签中选取未聚合的剩余标签,并选取任一个剩余标签B,计算该剩余标签B与其他剩余标签的距离相似度,并将与该剩余标签B相似度小于相似度阈值的剩余标签获取为剩余标签B的同义标签集合。

[0087] 当计算机设备完成上述迭代过程后,则可以得到聚合后的各个同义标签集合。

[0088] 在一种可能的实现方式中,在将各个目标标签进行聚合处理时,对于目标标签(以目标标签中的任一个目标标签C为例),计算机设备可以计算出该目标标签C与其他目标标签的距离相似度,并将与该目标标签C距离相似度小于相似度阈值且最小的目标标签D,获取为与目标标签C同义的标签,并将目标标签C与目标标签D获取为候选标签集合。

[0089] 同理,当将各个目标标签聚合为各个初始标签集合后,计算机设备再分别计算出各个初始标签集合之间的所有标签之间的距离相似度的平均值,作为标签集合距离,并将标签集合距离最小的候选标签集合进行合并处理,计算机设备迭代上述过程,直至标签集数量达到数量限制,并将最后剩下的各个候选标签集合确定为聚合后的同义标签集合。

[0090] 在一种可能的实现方式中,对于目标候选标签集合,当与该目标候选标签集合距离最小的候选标签集合二者之间任一集合中的目标标签数大于数量阈值时,将该目标候选标签集合与该候选标签集合距离最小的候选标签集合确定为聚合后的同义标签集合。也就是说,当两个候选标签集合合并后包含的标签数大于L,就不合并这两个标签集,而去检查下一对距离最近的标签集。

[0091] 综上所述,当获取到用于执行机器学习的目标图像后,计算机设备可以先根据目标图像上的各个目标标签,获取各个标签向量作为各个目标标签的分布式表示;计算机设备再以各个标签向量为标注,以目标图像为样本,训练第一图像处理模型,并根据第一图像处理模型中的全连接层参数划分为各个目标标签分别对应的视觉特征向量,再通过视觉特征向量以及标签向量,对各个目标标签进行聚合处理,从而使得各个目标标签在聚合为同义标签集合时,同时考虑到了目标标签的分布式表示以及目标标签在图像中含义,从而提高了同义的目标标签的聚合准确性。

[0092] 图3是根据一示例性实施例示出的同义标签聚合方法的方法流程图。该方法由计算机设备执行,该计算机设备可以是如图1所示的图像处理系统中的服务器,如图3所示,该同义标签聚合方法可以包括如下步骤:

[0093] 步骤301,根据目标图像上的各个目标标签,获取该各个目标标签分别对应的各个

标签向量。

[0094] 在一种可能的实现方式中,获取该目标图像上的各个目标标签;

[0095] 通过词向量生成模型,对该各个目标标签进行处理,获取该各个目标标签分别对应的各个标签向量。

[0096] 即在本申请实施例的一种可能的实现方式中,用fastText模型来获得每个标签的文本表示,且该fastText模型是根据训练图像上存在的训练标签所组成的句子训练得到的。

[0097] 和另一个更为广泛使用的词嵌入模型word2vec相比,fastText考虑了词子串的特征,因此能够更好的捕捉到词语间形态上的相似。由于很多同义标签都有相同或相似的子串,所以fastText模型学习到的标签表示更适合解决同义标签挖掘的问题。由于没有公开可用的中文标签嵌入,这里需要重新训练fastText模型。在原始方法中,词嵌入模型需要在由句子组成的语料上训练,而本申请中所考虑的场景的数据并不是由句子组成。但是,标注在同一图片上的所有标签在一个上下文中,可以被当做成一个句子。这样,就有大量的“句子”来训练fastText模型。

[0098] 步骤302,以该目标图像为样本,以该各个标签向量为标注,训练第一图像处理模型,并将训练后的第一图像处理模型中的全连接层参数,划分为该各个目标标签分别对应的视觉特征向量。

[0099] 在一种可能的实现方式中,将该目标图像通过第一图像处理模型进行处理,获得该目标图像对应的各个预测结果;

[0100] 根据该目标图像的各个预测结果,与该目标图像上的各个目标标签,通过损失函数进行处理,获得目标图像的损失函数值;

[0101] 根据该目标图像的损失函数值,对该第一图像处理模型的参数进行反向传播更新。

[0102] 在一种可能的实现方式中,该全连接层参数为以目标数量为第一维度、以该各个标签向量的数量为第二维度的矩阵;将该全连接层参数按照第二维度,划分为长度为目标数量,且分别与该各个目标标签对应的视觉特征向量。

[0103] 即在本申请实施例中,计算机设备可以利用图像+标签数据集训练基于卷积神经网络的图像多标签分类模型获取标签的视觉表示。由于一张图片对应多个标签,因此需要使用多标签的损失函数来学习卷积神经网络模型。

[0104] 假定训练集为 $\{(x_n, y_n)\}_{n=1, \dots, N}$, 其中 $x_n \in \mathcal{R}^D$ 为第 n 个样本中的图片输入, $y_n \in \{0, 1\}^K$ 为相应的多标签目标。从图片像素到ResNet-50的倒数第二层(pool15)的映射函数记为 $f(x_n; \theta)$, 其中 θ 为ResNet-50中除了最后一层的其它层的参数。最后一层为全连接层, 参数记做 W , 其为一个 $2048 \times K$ 的矩阵, 2048 是ResNet-50中pool15层的神经元数目。这样, softmax交叉熵损失函数定义为

$$[0105] \quad -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K \frac{y_{nk}}{\sum_{k'=1}^K y_{nk'}} \log \frac{\exp(w_k^T f(x_n; \theta))}{\sum_{k'=1}^K \exp(w_{k'}^T f(x_n; \theta))}$$

[0106] 当使用随机梯度下降算法AdaGrad训练完模型后, 图像 I 的特征向量就可以通过计

算 $f(I; \theta)$ 得到,而标签的视觉特征则是由其对应的图像特征组合得到。之前的工作常使用两种组合图像特征方式:计算所有图像特征的平均值的CNN-Mean和计算所有图像特征的最大值CNN-Max。经过实验对比,建议使用完成训练的卷积神经网络的最后一层权重作为标签的视觉特征,即 k 个标签的视觉特征为 W_{*k} 。

[0107] 步骤303,基于该目标标签分别对应的视觉特征向量以及该目标标签分别对应的标签向量,对该目标标签执行第一聚合操作,获得各个第一类标签集合。

[0108] 在本申请实施例中,计算机设备可以采用自底向上的合成聚类算法将标签聚类成同义标签集。首先定义标签之间的距离为标签的文本表示之间的欧式距离、标签的视觉表示之间的欧式距离以及标签之间的正规化编辑距离的平均值,然后将每一个标签看作是一个单独的标签集,最后每次合并距离最近的两个标签集合,上述第一聚合操作可以参考如图2所示实施例示出的聚合操作,但在图2所示实施例中,通过聚合操作可以直接得到同义标签集合,而在本申请实施例中,通过第一聚合操作得到的第一类标签集合并非直接作为同义标签集合,还需要通过后续步骤进行处理。

[0109] 步骤304,在该各个第一类标签集合中选取标注标签集合。

[0110] 该标注标签集合中的第一聚合标签中存在含义标注;该含义标注用于区分不同语义的第一聚合标签。

[0111] 可选的,计算机设备在该第一类标签集合中可以选取出标注标签集合,此时标注标签集合中的第一聚合标签(也就是标注标签集合通过聚合后得到的各个目标标签)是存在含义标注的标签,也就是说,在目标图像中的目标标签被聚合在标注标签集合中,可以通过人工标注的方式,提前对目标图像上的目标标签进行进一步标注,例如在(多个)目标图像中,对于手机,但此时由于人工标注的语言的标准不一致,此时存在手机的目标图像中可以包含“手机”标签,也可以包含“移动终端”标签,但此时“手机”标签与“移动终端”标签显然表示的含义一致,但却有不同的文字表示,因此可以通过人工在“手机”标签与“移动终端”标签上设定相同的标注,从而确定“手机”标签与“移动终端”标签具有相同的语义。

[0112] 可选的,在本申请实施例的一种可能的实现方式中,在获取到各个第一类标签集合后,可以通过人工对各个第一类标签集合中的部分或全部标签集合进行标注,再通过计算机设备在各个第一类标签集合中选取出标注标签集合。

[0113] 可选的,可以在第一类标签集合中选取20%的集合进行人工标注,得到部分标注标签集合,减小消耗的人力资源。

[0114] 步骤305,以该标注标签集合中的第一聚合标签为样本,以及第一聚合标签的含义标注为标注,对目标回归模型进行训练,获得训练后的目标回归模型。

[0115] 在一种可能的实现方式中获取该至少两个第一聚合标签分别对应的标签向量的第二标签向量距离、该至少两个第一聚合标签分别对应的视觉特征向量之间的第二视觉向量距离以及该至少两个第一聚合标签分别对应的标签向量之间的第二正规化编辑距离;

[0116] 将该第二标签向量距离、第二视觉向量距离以及该第二正规化编辑距离,根据该模型权重加权后通过激活函数处理,获得该至少两个第一聚合标签之间的距离函数值;

[0117] 根据该距离函数值,以及该至少两个第一聚合标签的含义标注,对该目标回归模型中的模型权重进行更新。

[0118] 可选的,利用标注同义标签集(也就是标注标签集合),通过有监督分类算法获得

标签之间的距离。例如在本申请实施例中计算机设备可以使用一个经典的逻辑斯蒂回归模型来判定两个标签是否为同义词。模型的输入为标签之间的文本表示距离、视觉表示距离和编辑距离,输出为两个标签为同义词的概率。给定两个标签 t_i 和 t_j , $D(t_{ti}, t_{tj})$ 为标签之间文本表示的欧式距离(也就是第二标签向量距离), $D(t_{vi}, t_{vj})$ 为标签之间视觉表示的欧式距离(第二视觉向量距离), $NED(t_i, t_j)$ 为标签之间的正则化编辑距离, w_T 、 w_V 和 w_e 分别是标签的文本表示距离、视觉表示距离和正则化编辑距离在最终距离中的权重。则两个标签的距离函数定义为:

$$[0119] \quad y = \sigma(w_T D(t_{ti}, t_{tj}) + w_V D(t_{vi}, t_{vj}) + w_e NED(t_i, t_j) + b)$$

[0120] 其中, b 是偏置项, $\sigma(x) = 1 / (1 + \exp(-x))$ 是sigmoid函数。模型的优化目标是当两个标签是同义标签时, y 尽可能大,而当两个标签不是同义标签时, y 尽可能小。所有训练数据中的标签对都被用来训练这个逻辑斯蒂回归模型。模型经过训练,即可以得到三个距离之间的权重,最终得到更精准的任意两个标签之间的距离函数。和平均分配各个来源对最终距离的权重相比,有监督方法学习到的距离能够更加合理的考虑每一种距离的贡献。

[0121] 步骤306,根据该训练后的目标回归模型,计算该各个目标标签之间的标签距离,并根据该各个目标标签之间的标签距离对该各个目标标签进行第二聚合操作,获得聚合后的各个同义标签集合。

[0122] 在本申请实施例中,当通过步骤305,即利用有监督方法学习到的标签之间的距离,再次采用自底向上的合成聚类算法(如步骤203所示,此处不再赘述)将标签聚类成同义标签集。并且聚类过程中 L 值和聚成的标签集的数目的选取可以根据训练集上的统计特征得到。

[0123] 当聚合后得到各个同义标签集合后,则可以在通过图像训练集合对图像处理模型进行训练之前,先通过同义标签集合对图像训练集合中的各个标签进行处理,将图像训练集合中的相似的标签划分为相同的语义(如用同一个标签来代替),从而尽可能减小了由于用户的标注用语不同导致的机器学习过程的噪声。

[0124] 综上所述,当获取到用于执行机器学习的目标图像后,计算机设备可以先根据目标图像上的各个目标标签,获取各个标签向量作为各个目标标签的分布式表示;计算机设备再以各个标签向量为标注,以目标图像为样本,训练第一图像处理模型,并根据第一图像处理模型中的全连接层参数划分为各个目标标签分别对应的视觉特征向量,再通过视觉特征向量以及标签向量,对各个目标标签进行聚合处理,从而使得各个目标标签在聚合为同义标签集合时,同时考虑到了目标标签的分布式表示以及目标标签在图像中含义,从而提高了同义的目标标签的聚合准确性。

[0125] 图4是根据一示例性实施例示出的同义标签聚合装置的结构方框图。

[0126] 所述装置包括:

[0127] 标签向量获取模块401,用于根据目标图像上的各个目标标签,获取所述各个目标标签分别对应的各个标签向量;

[0128] 模型训练模块402,用于以所述目标图像为样本,以所述各个标签向量为标注,训练第一图像处理模型,并将训练后的第一图像处理模型中的全连接层参数,划分为所述各个目标标签分别对应的视觉特征向量;

[0129] 标签聚合模块403,用于基于所述目标标签分别对应的视觉特征向量以及所述目

标签分别对应的标签向量,对所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。

[0130] 在一种可能的实现方式中,所述标签向量获取模块,还用于,

[0131] 获取所述目标图像上的各个目标标签;

[0132] 通过词向量生成模型,对所述各个目标标签进行处理,获取所述各个目标标签分别对应的各个标签向量。

[0133] 在一种可能的实现方式中,所述全连接层参数为以目标数量为第一维度、以所述各个标签向量的数量为第二维度的矩阵;

[0134] 所述模型训练模块,还用于,

[0135] 将所述全连接层参数按照第二维度,划分为长度为目标数量,且分别与所述各个目标标签对应的视觉特征向量。

[0136] 在一种可能的实现方式中,所述模型训练模块,还用于,

[0137] 将所述目标图像通过第一图像处理模型进行处理,获得所述目标图像对应的各个预测结果;

[0138] 根据所述目标图像的各个预测结果,与所述目标图像上的各个目标标签,通过损失函数进行处理,获得目标图像的损失函数值;

[0139] 根据所述目标图像的损失函数值,对所述第一图像处理模型的参数进行反向传播更新。

[0140] 在一种可能的实现方式中,所述标签聚合模块,还用于,

[0141] 针对任意两个目标标签,计算所述两个目标标签的标签向量之间的第一标签向量距离、所述两个目标标签分别对应的视觉向量之间的第一视觉向量距离以及所述两个目标标签的标签向量之间的第一正规化编辑距离;

[0142] 将所述第一标签向量距离、所述第一视觉向量距离以及所述第一正规化编辑距离按照目标权重加权求和,获取所述目标标签之间的距离相似度;

[0143] 根据所述目标标签之间的距离相似度,将所述各个目标标签进行聚合处理,获得聚合后的各个同义标签集合。

[0144] 在一种可能的实现方式中,所述标签聚合模块,还用于,

[0145] 基于所述目标标签分别对应的视觉特征向量以及所述目标标签分别对应的标签向量,对所述目标标签执行第一聚合操作,获得各个第一类标签集合;

[0146] 在所述各个第一类标签集合中选取标注标签集合;所述标注标签集合中的第一聚合标签中存在含义标注;所述含义标注用于区分不同语义的第一聚合标签;

[0147] 以所述标注标签集合中的第一聚合标签为样本,以及第一聚合标签的含义标注为标注,对目标回归模型进行训练,获得训练后的目标回归模型;

[0148] 根据所述训练后的目标回归模型,计算所述各个目标标签之间的标签距离,并根据所述各个目标标签之间的标签距离对所述各个目标标签进行第二聚合操作,获得聚合后的各个同义标签集合。

[0149] 在一种可能的实现方式中,所述标签聚合模块,还用于,

[0150] 获取至少两个第一聚合标签分别对应的标签向量的第二标签向量距离、所述至少两个第一聚合标签分别对应的视觉特征向量之间的第二视觉向量距离以及所述至少两个

第一聚合标签分别对应的标签向量之间的第二正规化编辑距离；

[0151] 将所述第二标签向量距离、第二视觉向量距离以及所述第二正规化编辑距离,根据所述模型权重加权后通过激活函数处理,获得所述至少两个第一聚合标签之间的距离函数值;

[0152] 根据所述距离函数值,以及所述至少两个第一聚合标签的含义标注,对所述目标回归模型中的模型权重进行更新。

[0153] 综上所述,当获取到用于执行机器学习的目标图像后,计算机设备可以先根据目标图像上的各个目标标签,获取各个标签向量作为各个目标标签的分布式表示;计算机设备再以各个标签向量为标注,以目标图像为样本,训练第一图像处理模型,并根据第一图像处理模型中的全连接层参数划分为各个目标标签分别对应的视觉特征向量,再通过视觉特征向量以及标签向量,对各个目标标签进行聚合处理,从而使得各个目标标签在聚合为同义标签集合时,同时考虑到了目标标签的分布式表示以及目标标签在图像中含义,从而提高了同义的目标标签的聚合准确性。

[0154] 请参阅图5,其是根据本申请一示例性实施例提供的一种计算机设备示意图,所述计算机设备包括存储器和处理器,所述存储器用于存储计算机程序,所述计算机程序被所述处理器执行时,实现上述方法。

[0155] 其中,处理器可以为中央处理器(Central Processing Unit,CPU)。处理器还可以为其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等芯片,或者上述各类芯片的组合。

[0156] 存储器作为一种非暂态计算机可读存储介质,可用于存储非暂态软件程序、非暂态计算机可执行程序以及模块,如本发明实施方式中的方法对应的程序指令/模块。处理器通过运行存储在存储器中的非暂态软件程序、指令以及模块,从而执行处理器的各种功能应用以及数据处理,即实现上述方法实施方式中的方法。

[0157] 存储器可以包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需要的应用程序;存储数据区可存储处理器所创建的数据等。此外,存储器可以包括高速随机存取存储器,还可以包括非暂态存储器,例如至少一个磁盘存储器件、闪存器件、或其他非暂态固态存储器件。在一些实施方式中,存储器可选包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至处理器。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0158] 在一示例性实施例中,还提供了一种计算机可读存储介质,用于存储有至少一条计算机程序,所述至少一条计算机程序由处理器加载并执行以实现上述方法中的全部或部分步骤。例如,该计算机可读存储介质可以是只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、只读光盘(Compact Disc Read-Only Memory,CD-ROM)、磁带、软盘和光数据存储设备等。

[0159] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识

或惯用技术手段。说明书和实施例仅被视为示例性的,本申请的真正范围和精神由下面的权利要求指出。

[0160] 应当理解的是,本申请并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求来限制。

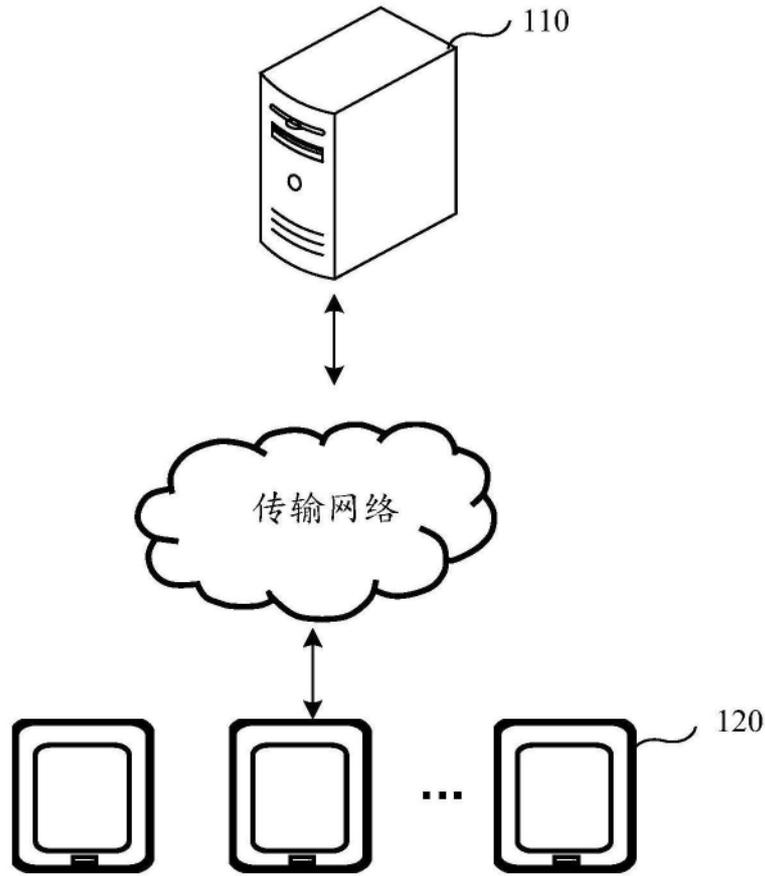


图1

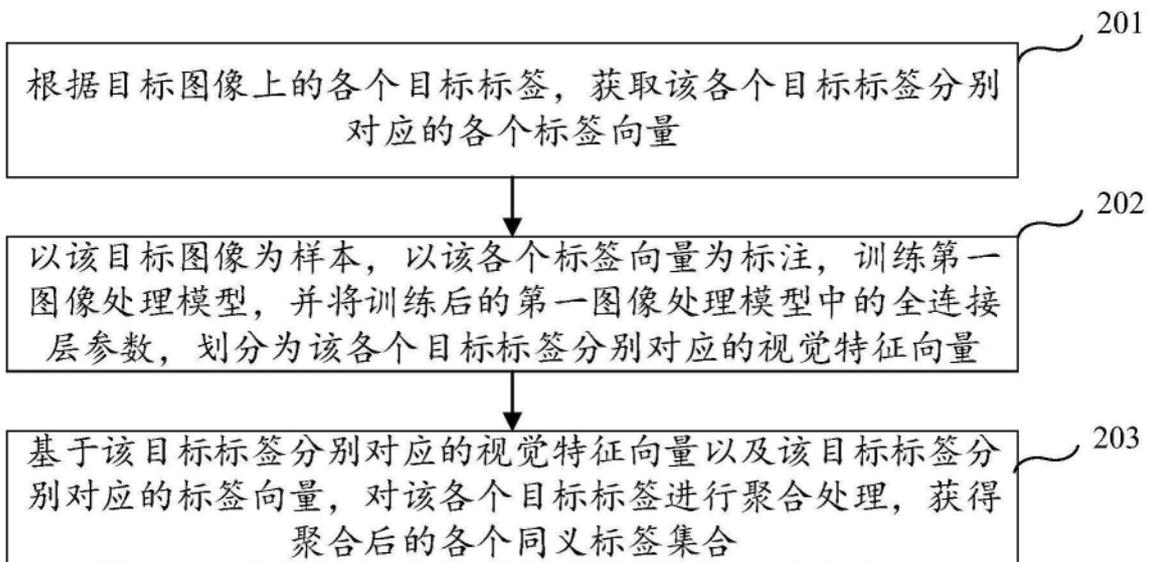


图2

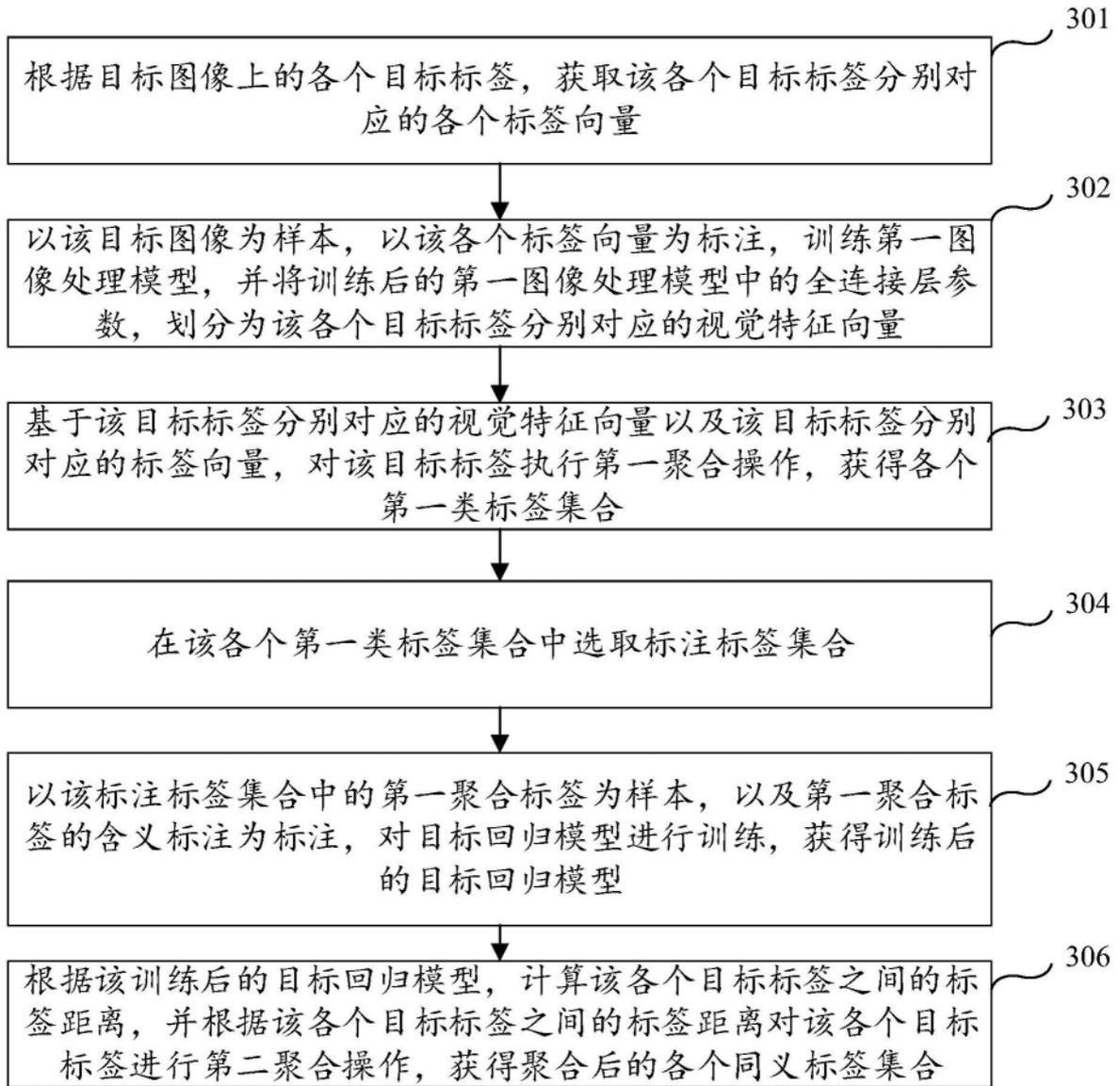


图3



图4

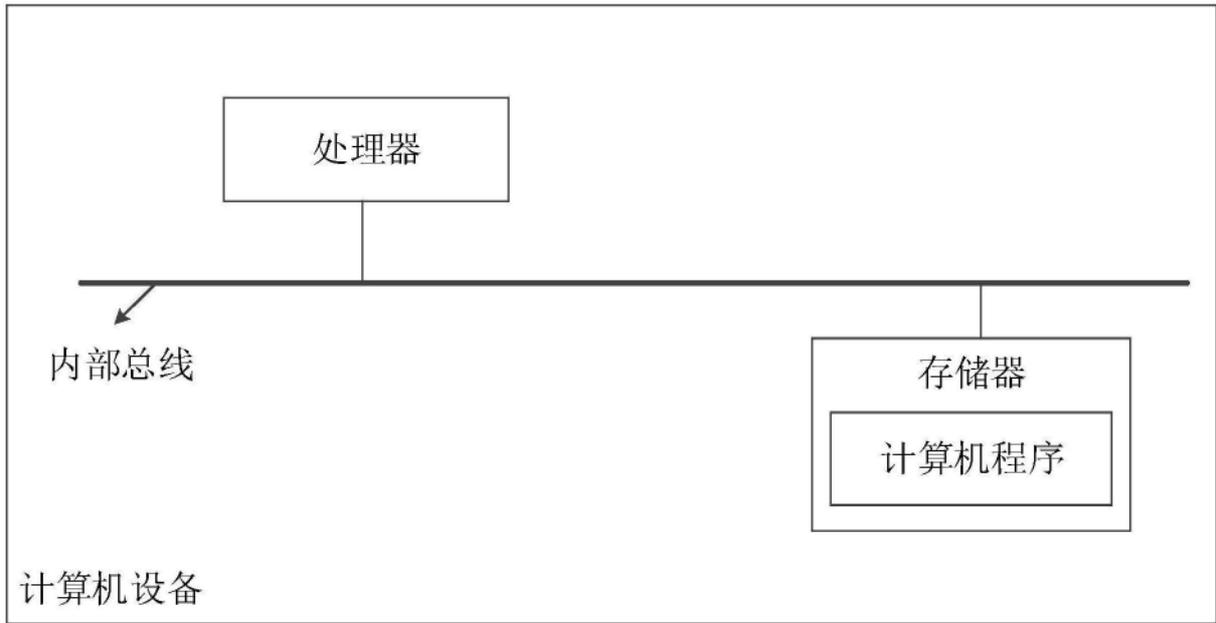


图5