



(12) 发明专利申请

(10) 申请公布号 CN 112200031 A

(43) 申请公布日 2021. 01. 08

(21) 申请号 202011033394.2

(22) 申请日 2020.09.27

(71) 申请人 上海眼控科技股份有限公司
地址 200030 上海市徐汇区中山南二路107号1幢20层I单元

(72) 发明人 赵佳男

(74) 专利代理机构 上海百一领御专利代理事务所(普通合伙) 31243
代理人 王奎宇 朱永海

(51) Int. Cl.
G06K 9/00 (2006.01)
G06N 3/04 (2006.01)
G06N 3/08 (2006.01)

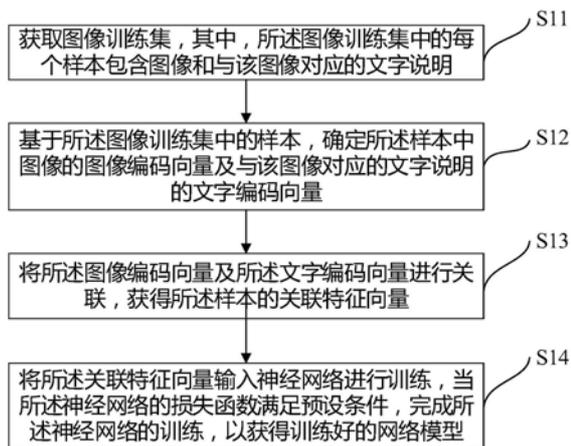
权利要求书2页 说明书9页 附图2页

(54) 发明名称

一种用于生成图像对应文字说明的网络模型训练方法与设备

(57) 摘要

与现有技术相比,本申请通过一种用于生成图像对应文字说明的网络模型训练方法与设备,首先获取图像训练集,接着基于所述图像训练集中的样本,确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量,然后将所述图像编码向量及所述文字编码向量进行关联,获得所述样本的关联特征向量,最后将所述关联特征向量输入网络模型进行训练,当所述网络模型的损失函数满足预设条件,完成所述网络模型的训练,以获得训练好的网络模型。通过该方法,获得一个可用于对图像自动标注文字说明的网络模型,大大降低人工标注文字说明的人力成本。将该网络模型应用于行人重识别任务,可提升行人重识别的识别性能,具有实际应用价值。



1. 一种用于生成图像对应文字说明的网络模型训练方法,其特征在于,所述方法包括:
获取图像训练集,其中,所述图像训练集中的每个样本包含图像和与该图像对应的文字说明;

基于所述图像训练集中的样本,确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量;

将所述图像编码向量及所述文字编码向量进行关联,获得所述样本的关联特征向量;

将所述关联特征向量输入神经网络进行训练,当所述神经网络的损失函数满足预设条件,完成所述神经网络的训练,以获得训练好的网络模型。

2. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

获取待生成文字说明的目标图像;

基于所述目标图像,确定所述目标图像的图像编码向量;

将所述图像编码向量输入所述训练好的网络模型,以获得所述目标图像对应的文字说明。

3. 根据权利要求2所述的方法,其特征在于,在所述基于所述目标图像,确定所述目标图像的图像编码向量之前,所述方法还包括:

将所述目标图像的图片风格与所述图像训练集中的图片风格进行统一。

4. 根据权利要求2或3所述的方法,其特征在于,所述将所述图像编码向量输入所述训练好的网络模型,以获得所述目标图像对应的文字说明包括:

将所述图像编码向量输入所述训练好的网络模型,提取所述目标图像的属性特征信息;

基于所述目标图像的属性特征信息,生成零散的文字描述信息;

对所述零散的文字描述信息进行整合,以获得所述目标图像对应的文字说明。

5. 一种用于生成图像对应文字说明的网络模型训练设备,其特征在于,所述设备包括:

第一装置,用于获取图像训练集,其中,所述图像训练集中的每个样本包含图像和与该图像对应的文字说明;

第二装置,用于基于所述图像训练集中的样本,确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量;

第三装置,用于将所述图像编码向量及所述文字编码向量进行关联,获得所述样本的关联特征向量;

第四装置,用于将所述关联特征向量输入神经网络进行训练,当所述神经网络的损失函数满足预设条件,完成所述神经网络的训练,以获得训练好的网络模型。

6. 根据权利要求5所述的设备,其特征在于,所述第二装置包括:

VGG16网络模块,用于基于所述图像训练集中样本的图像来确定所述图像的图像编码向量;

RNN网络模块,用于基于所述图像训练集中样本的图像对应的文字说明来确定所述图像对应的文字说明的文字编码向量。

7. 根据权利要求5或6所述的设备,其特征在于,所述设备还包括:

第五装置,用于获取待生成文字说明的目标图像;

第六装置,用于基于所述目标图像,确定所述目标图像的图像编码向量;

第七装置,用于将所述图像编码向量输入所述训练好的网络模型,以获得所述目标图像对应的文字说明。

8.一种计算机可读介质,其特征在于,
其上存储有计算机可读指令,所述计算机可读指令可被处理器执行以实现如权利要求1至4中任一项所述的方法。

9.一种设备,其特征在于,该设备包括:
一个或多个处理器;以及
存储有计算机可读指令的存储器,所述计算机可读指令在被执行时使所述处理器执行如权利要求1至4中任一项所述方法的操作。

一种用于生成图像对应文字说明的网络模型训练方法与设备

技术领域

[0001] 本申请涉及计算机视觉处理技术领域,尤其涉及一种用于生成图像对应文字说明的技术。

背景技术

[0002] 计算机视觉领域中的行人重识别技术是对不同的监控设备获取的图像中的同一目标行人进行识别、匹配,在智能安防等领域的研究和应用具有重要的意义。在现实场景中,由于受到人体姿态、拍摄角度变化以及照明条件等因素的影响,如何准确地识别并匹配目标行人成为行人重识别技术中十分具有挑战性的问题。

[0003] 近年来,随着深度学习相关算法在计算机视觉领域的成功应用,以及大规模数据库的不断出现,很多行人重识别方法都在提取图像本身提供的特征之外,利用一些例如人体姿态、人体各部位属性、对图像的文字说明等辅助性信息,来提升行人重识别的效果。

[0004] 其中,由于对图像的文字说明可以提供具体而全面的信息,在语义上比可视属性更丰富,而且不同图像对同一目标行人的位置说明通常更一致。因此,在行人重识别方法中利用对图像的文字说明,可以减少行人重识别方法中行人外观差异的问题,提升识别、匹配准确度。

[0005] 但是,对行人重识别方法的众多图像,特别是采用深度学习对行人重识别方法的图像数据集进行文字标注需要投入很高的人力成本、效率低,这就使得在行人重识别方法中应用对图像的文字说明受到了限制。

发明内容

[0006] 本申请的目的是提供一种用于生成图像对应文字说明的网络模型训练的方法与设备,用以解决现有技术中基于深度学习的行人重识别的图像数据集的人工文字标注投入人力成本高企、效率低的技术问题。

[0007] 根据本申请的一个方面,提供了一种用于生成图像对应文字说明的网络模型训练方法,其中,所述方法包括:

[0008] 获取图像训练集,其中,所述图像训练集中的每个样本包含图像和与该图像对应的文字说明;

[0009] 基于所述图像训练集中的样本,确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量;

[0010] 将所述图像编码向量及所述文字编码向量进行关联,获得所述样本的关联特征向量;

[0011] 将所述关联特征向量输入神经网络进行训练,当所述神经网络的损失函数满足预设条件,完成所述神经网络的训练,以获得训练好的网络模型。

[0012] 可选地,其中,所述方法还包括:

[0013] 获取待生成文字说明的目标图像;

- [0014] 基于所述目标图像,确定所述目标图像的图像编码向量;
- [0015] 将所述图像编码向量输入所述训练好的网络模型,以获得所述目标图像对应的文字说明。
- [0016] 可选地,其中,在所述基于所述目标图像,确定所述目标图像的图像编码向量之前,所述方法还包括:
- [0017] 将所述目标图像的图片风格与所述图像训练集中的图片风格进行统一。
- [0018] 可选地,其中,所述将所述图像编码向量输入所述训练好的网络模型,以获得所述目标图像对应的文字说明包括:
- [0019] 将所述图像编码向量输入所述训练好的网络模型,提取所述目标图像的属性特征信息;
- [0020] 基于所述目标图像的属性特征信息,生成零散的文字描述信息;
- [0021] 对所述零散的文字描述信息进行整合,以获得所述目标图像对应的文字说明。
- [0022] 根据本申请的另一方面,还提供了一种用于生成图像对应文字说明的网络模型训练设备,其中,所述设备包括:
- [0023] 第一装置,用于获取图像训练集,其中,所述图像训练集中的每个样本包含图像和与该图像对应的文字说明;
- [0024] 第二装置,用于基于所述图像训练集中的样本,确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量;
- [0025] 第三装置,用于将所述图像编码向量及所述文字编码向量进行关联,获得所述样本的关联特征向量;
- [0026] 第四装置,用于将所述关联特征向量输入神经网络进行训练,当所述神经网络的损失函数满足预设条件,完成所述神经网络的训练,以获得训练好的网络模型。
- [0027] 与现有技术相比,本申请通过一种用于生成图像对应文字说明的网络模型训练方法与设备,首先获取图像训练集,接着基于所述图像训练集中的样本,确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量,然后将所述图像编码向量及所述文字编码向量进行关联,获得所述样本的关联特征向量,最后将所述关联特征向量输入网络模型进行训练,当所述网络模型的损失函数满足预设条件,完成所述网络模型的训练,以获得训练好的网络模型。通过该方法,获得一个可用于对图像自动标注文字说明的网络模型,大大降低人工标注文字说明的人力成本。将该网络模型应用于行人重识别任务,可提升行人重识别的识别性能,具有实际应用价值。

附图说明

- [0028] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本发明的其它特征、目的和优点将会变得更明显:
- [0029] 图1示出根据本申请一个方面的一种用于生成图像对应文字说明的网络模型训练方法流程图;
- [0030] 图2示出根据本申请一个方面的一种用于生成图像对应文字说明的网络模型结构示意图;
- [0031] 图3示出根据本申请一个方面的一种用于生成图像对应文字说明的网络模型的

CNN模块卷积层结构示意图；

[0032] 图4示出根据本申请一个方面的一种用于生成图像对应文字说明的网络模型训练设备示意图；

[0033] 附图中相同或相似的附图标记代表相同或相似的部件。

具体实施方式

[0034] 下面结合附图对本发明作进一步详细描述。

[0035] 在本申请一个典型的配置中，系统各模块和可信方均包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0036] 内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM)和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0037] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非暂存电脑可读媒体(transitory media)，如调制的数据信号和载波。

[0038] 为更进一步阐述本申请所采取的技术手段及取得的效果，下面结合附图及优选实施例，对本申请的技术方案，进行清楚和完整的描述。

[0039] 图1示出本申请一个方面的一种用于生成图像对应文字说明的网络模型训练方法流程图，其中，一个实施例的方法包括：

[0040] S11获取图像训练集，其中，所述图像训练集中的每个样本包含图像和与该图像对应的文字说明；

[0041] S12基于所述图像训练集中的样本，确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量；

[0042] S13将所述图像编码向量及所述文字编码向量进行关联，获得所述样本的关联特征向量；

[0043] S14将所述关联特征向量输入神经网络进行训练，当所述神经网络的损失函数满足预设条件，完成所述神经网络的训练，以获得训练好的网络模型。

[0044] 在本申请中，所述方法通过设备1执行，所述设备1为计算机设备和/或云，所述计算机设备包括但不限于个人计算机、笔记本电脑、工业计算机、网络主机、单个网络服务器、多个网络服务器集；所述云由基于云计算(Cloud Computing)的大量计算机或网络服务器构成，其中，云计算是分布式计算的一种，由一群松散耦合的计算机集组成的一个虚拟超级计算机。

[0045] 在此，所述计算机设备和/或云仅为举例，其他现有的或者今后可能出现的设备和/或资源共享平台如适用于本申请也应包含在本申请的保护范围内，在此，以引用的方式

包含于此。

[0046] 在该实施例中,在所述步骤S11中,设备1获取到图像训练集,其中,所述图像训练集中的每个样本包含图像和与该图像对应的文字说明,所述文字说明通常以与图像一一对应的标注文件形式保存。所述图像训练集可以是现有的行人重识别数据集,所述现有的行人重识别数据集中的每个样本包含图像和与该图像对应的文字说明,例如在行人重识别领域较为著名的CUHK-PEDES数据集,也可以是获取到的适用于行人重识别任务的图像数据集,然后进行预处理,即遍历对图像数据集中的每一个图像进行人工标注文字说明,再将图像及其对应的文字说明作为一个样本,所有样本组成图像训练集。

[0047] 继续在该实施例中,在所述步骤S12中,设备1获取到图像训练集后,对图像训练集中的每个样本进行处理,其中,将样本中的图像进行编码,以获得图像的图像编码向量。

[0048] 例如,将样本中的图像输入训练好的VGG16模块结合ReLU激活函数的神经网络模型进行图像编码,获得图像对应的图像编码向量,图像编码向量的长度可根据实际需要设置,比如,设置为512维。其中,训练好的VGG16结合ReLU激活函数的神经网络模型可基于实际条件进行训练后获得,例如,在现有的ImageNet图像数据集上进行VGG16结合ReLU激活函数的神经网络预训练,目的是给予网络参数一个较好的初始化,使得训练过程可以更快的收敛,而且在网络中还可加入弃权(dropout)操作,以减少神经网络在训练中可能出现的过拟合现象。

[0049] 设备1获取到图像训练集后,对图像训练集中的每个样本进行处理,除了将样本中的图像进行编码,以获得图像的图像编码向量之外,还将样本中图像对应的文字说明按字/词进行字/词向编码,以获得图像对应的文字说明的文字编码向量。

[0050] 例如,将样本中的图像对应的文字说明输入基础的RNN神经网络模型进行文字编码,获得图像对应的文字说明的文字编码向量,文字编码向量的长度可根据实际需要设置,但应与图像的图像编码向量的长度相同,比如,设置为512维。

[0051] 继续在该实施例中,在所述步骤S13中,设备1将获得图像的图像编码向量及图像对应的文字说明的文字编码向量进行关联,获得对应样本的关联特征向量。

[0052] 其中,所述将获得图像的图像编码向量及图像对应的文字说明的文字编码向量进行关联可以是将获得图像的图像编码向量及图像对应的文字说明的文字编码向量进行拼接,例如,将512维的图像编码向量及512维的文字编码向量进行拼接,获得对应样本的1024维关联特征向量;也可以是将获得图像的图像编码向量及图像对应的文字说明的文字编码向量进行对应通道相加,例如,将512维的图像编码向量及512维的文字编码向量进行对应通道相加,获得对应样本的512维关联特征向量;还可以是将获得图像的图像编码向量及图像对应的文字说明的文字编码向量基于加权映射(attention)方法进行关联,例如,将512维的图像编码向量及512维的文字编码向量基于加权映射方法进行关联,获得对应样本的512维关联特征向量。

[0053] 继续在该实施例中,在所述步骤S14中,设备1将关联特征向量输入神经网络进行训练,当所述神经网络的损失函数满足预设条件,完成所述神经网络的训练,以获得训练好的网络模型。通过神经网络构建图像训练集的样本中图像及图像对应的文字说明之间的关联联系。

[0054] 其中,所述预设条件包括神经网络的损失函数的数值符合预先设定的阈值,或者

在预设的训练循环次数(epoch)内,虽然神经网络的损失函数的数值损失未符合预先设定的阈值,但不再下降。

[0055] 其中,所述训练好的网络模型用于对行人重识别任务的图像数据集中的图像自动标注图像对应的文字说明。

[0056] 一个实施例的一种用于生成图像对应文字说明的网络模型结构如图2所示。例如,设备1获取图像训练集,其中的一个样本的图像对应标注的文字说明是A woman in red shirt,将该图像输入训练好的VGG16结合ReLU激活函数的神经网络模型进行图像编码,得到512维的图像编码向量;获取该图像对应的文字说明标注,将文字说明输入基础的RNN神经网络模型进行文字编码,得到图像对应的文字说明的512维的文字编码向量;然后将512维的图像编码向量及512维的文字编码向量进行拼接成1024维的关联特征向量,输入由CNN模块和分类层构成的神经网络进行训练,其中,CNN模块由3层卷积层和GLU激活函数组成,每层卷积层的结构如图3所示,每层卷积层中还加入了权重归一化、残差连接和弃权操作,以使得神经网络训练的效果更好,CNN模块输出512维向量,作为分类层的输入,分类层对输入的512维向量进行下采样以得到256维向量,再经过归一化的softmax函数,对特征向量进行分类。

[0057] 当所述神经网络的损失函数的数值满足预设条件,则完成所述神经网络的训练,得到训练好的网络模型,以用于对行人重识别任务的图像数据集中的图像自动标注图像对应的文字说明。

[0058] 例如,神经网络采用交叉熵损失函数(cross-entropy loss function)进行训练,并采用基于梯度下降的反向传播算法对神经网络进行优化,其中,优化器可采用RMSProp,所述交叉熵损失函数的数值可基于如下公式计算:

$$[0059] \quad L_{ID} = - \sum_k^K y_k \log \left(\frac{e^{\theta_k \cdot f}}{\sum_j^K e^{\theta_j \cdot f}} \right)$$

[0060] 其中,

[0061] L_{ID} 为交叉熵损失函数;

[0062] K 为分类的类别总数;

[0063] y_k 为第 k 个类别的概率;

[0064] θ 为优化初始参数;

[0065] f 为神经网络提取的特征。

[0066] 可选地,其中,所述一种用于生成图像对应文字说明的网络模型训练方法还包括:

[0067] S15(未示出)获取待生成文字说明的目标图像;

[0068] S16(未示出)基于所述目标图像,确定所述目标图像的图像编码向量;

[0069] S17(未示出)将所述图像编码向量输入所述训练好的网络模型,以获得所述目标图像对应的文字说明。

[0070] 在一个实施例中,在步骤S15中,设备1获取行人重识别任务的待识别的目标图像,其中,所述待识别的目标图像没有对应的文字说明。

[0071] 继续在该实施例中,在步骤S16中,设备1将获得的所述待识别的目标图像进行图像编码,获得所述待识别的目标图像对应的图像编码向量。例如,将所述待识别的目标图像

输入训练好的VGG16结合ReLU激活函数的神经网络模型进行图像编码,获得图像对应的图像编码向量,图像编码向量的长度可根据实际需要设置,比如,设置为512维。其中,训练好的VGG16结合ReLU激活函数的神经网络模型可基于实际条件进行训练后获得,例如,在现有的ImageNet图像数据集上进行VGG16结合ReLU激活函数的神经网络预训练,目的是给予网络参数一个较好的初始化,使得训练过程可以更快的收敛,而且在网络中还可加入弃权(dropout)操作,以减少神经网络在训练中可能出现的过拟合现象。

[0072] 如果所述待识别的目标图像的图片风格与所述图像训练集的样本中的图像的图片风格存在较大差异,比如图像拍摄的光线明亮度、曝光度、色彩饱和度、色调等差异较大,会导致基于所述图像训练集(例如CUHK-PEDES)得到的训练好的网络模型对所述待识别的目标图像的文字说明预测的泛化能力较差。

[0073] 可选地,其中,在所述步骤S16之前,所述方法还包括:

[0074] S18(未示出)将所述目标图像的图片风格与所述图像训练集中的图片风格进行统一。

[0075] 为了使得基于所述图像训练集得到的训练好的网络模型对所述待识别的目标图像的文字说明的预测泛化能力较好,在对所述待识别的目标图像进行图像编码之前,采用风格迁移技术,将所述目标图像的图片风格与所述图像训练集中的图片风格进行统一,以消除不同图像的图片风格差异,提升训练好的网络模型预测的泛化能力。例如,可采用SPGAN(Similarity Preserving cycle-consistent Generative Adversarial Network,)方法将所述待识别的目标图像的图片风格迁移为所述图像训练集的样本图像的图片风格,比如,可将用于行人重识别任务的Duke-MTMC数据集中的图像作为目标图像,采用SPGAN方法,将Duke-MTMC数据集中的图像的图片风格迁移转换为CUHK-PEDES数据集的图片风格。也可采用其它图片风格迁移方法,比如CycleGAN, Pix2Pix等图片风格迁移方法,在此,不作限定,其他任何图片风格迁移方法如适用于本申请,都处于本申请保护范围。

[0076] 继续在该实施例中,在步骤S17中,设备1将得到的所述待识别的目标图像对应的图像编码向量输入训练好的网络模型,以获得所述待识别的目标图像对应的文字说明,进一步地,将获得所述待识别的目标图像对应的文字说明保存为所述待识别的目标图像对应的标注文件。

[0077] 可选地,其中,所述步骤S17包括:

[0078] 将所述图像编码向量输入所述训练好的网络模型,提取所述目标图像的属性特征信息;

[0079] 基于所述目标图像的属性特征信息,生成零散的文字描述信息;

[0080] 对所述零散的文字描述信息进行整合,以获得所述目标图像对应的文字说明。

[0081] 例如,将图片风格迁移后的用于行人重识别任务的Duke-MTMC数据集中的图像编码成图像编码向量后输入到训练好的网络模型中,网络模型从中提取属性特征,根据属性特征预测目标图像的解释性零散的字(词)文字描述,结合训练学习到所述图像训练集中样本图像对应的文字说明的语言风格(语法、规则),把零散的字(词)文字描述加工整合,组合成对应的文字说明。

[0082] 进一步地,采用所述方法遍历所述目标图像所在图像数据集的每一张图像,可以高效准确地自动为所述目标图像所在图像数据集标注文字说明,得到带有文字说明标注的

图像数据集。比如采用所述方法遍历Duke-MTMC数据集中的每张图像,可以高效准确地获得Duke-MTMC数据集的文字说明标注,得到带有文字说明标注的Duke-MTMC数据集。

[0083] 进一步地,可将获得的带有文字说明的图像数据集(比如带有文字说明的Duke-MTMC数据集)用于行人重识别任务中,为现有利用文字说明来辅助行人重识别任务的神经网络模型提供视觉图像特征的同时,还提供了具有丰富语义信息的文字说明,从而使基于行人重识别任务获得更好的性能。

[0084] 图4示出根据本申请另一个方面的一种用于生成图像对应文字说明的网络模型训练设备示意图,其中,所述设备包括:

[0085] 第一装置41,用于获取图像训练集,其中,所述图像训练集中的每个样本包含图像和与该图像对应的文字说明;

[0086] 第二装置42,用于基于所述图像训练集中的样本,确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量;

[0087] 第三装置43,用于将所述图像编码向量及所述文字编码向量进行关联,获得所述样本的关联特征向量;

[0088] 第四装置44,用于将所述关联特征向量输入神经网络进行训练,当所述神经网络的损失函数满足预设条件,完成所述神经网络的训练,以获得训练好的网络模型。

[0089] 其中,所述第一装置41获取到图像训练集,其中,所述图像训练集中的每个样本包含图像和与该图像对应的文字说明,所述文字说明通常以与图像一一对应的标注文件形式保存。所述图像训练集可以是现有的行人重识别数据集,所述现有的行人重识别数据集中的每个样本包含图像和与该图像对应的文字说明,例如在行人重识别领域较为著名的CUHK-PEDES数据集,也可以是获取到的适用于行人重识别任务的图像数据集,然后进行预处理,即遍历对图像数据集中的每一个图像进行人工标注文字说明,再将图像及其对应的文字说明作为一个样本,所有样本组成图像训练集。

[0090] 其中,所述第二装置42对图像训练集中的每个样本进行处理,其中,将样本中的图像进行编码,以获得图像的图像编码向量。

[0091] 可选地,其中,所述第二装置42包括:

[0092] VGG16网络模块,用于基于所述图像训练集中样本的图像来确定所述图像的图像编码向量;

[0093] RNN网络模块,用于基于所述图像训练集中样本的图像对应的文字说明来确定所述图像对应的文字说明的文字编码向量。

[0094] 其中,VGG16网络模块包括VGG16模块及ReLU激活函数,VGG16网络模块将输入的图像训练集样本中的图像进行图像编码,获得图像对应的图像编码向量。图像编码向量的长度可根据实际需要设置,比如,设置为512维。

[0095] 其中,RNN网络模块将输入的图像训练集样本中的图像对应的文字说明按字/词进行字/词向编码,以获得图像对应的文字说明的文字编码向量。文字编码向量的长度可根据实际需要设置,但应与图像的图像编码向量的长度相同,比如,设置为512维。

[0096] 其中,所述第三装置43将获得图像的图像编码向量及图像对应的文字说明的文字编码向量进行关联,获得对应样本的关联特征向量。

[0097] 其中,所述将获得图像的图像编码向量及图像对应的文字说明的文字编码向量进

行关联可以是获得图像的图像编码向量及图像对应的文字说明的文字编码向量进行拼接,例如,将512维的图像编码向量及512维的文字编码向量进行拼接,获得对应样本的1024维关联特征向量;也可以是将获得图像的图像编码向量及图像对应的文字说明的文字编码向量进行对应通道相加,例如,将512维的图像编码向量及512维的文字编码向量进行对应通道相加,获得对应样本的512维关联特征向量;还可以是将获得图像的图像编码向量及图像对应的文字说明的文字编码向量基于加权映射(attention)方法进行关联,例如,将512维的图像编码向量及512维的文字编码向量基于加权映射方法进行关联,获得对应样本的512维关联特征向量。

[0098] 其中,所述第四装置44将关联特征向量输入神经网络进行训练,当所述神经网络的损失函数满足预设条件,完成所述神经网络的训练,以获得训练好的网络模型。通过神经网络构建图像训练集的样本中图像及图像对应的文字说明之间的关联联系,可将所述训练好的网络模型用于对行人重识别任务的图像数据集中的图像自动标注图像对应的文字说明。

[0099] 可选地,所述设备还包括:

[0100] 第五装置45(未示出),用于获取待生成文字说明的目标图像;

[0101] 第六装置46(未示出),用于基于所述目标图像,确定所述目标图像的图像编码向量;

[0102] 第七装置47(未示出),用于将所述图像编码向量输入所述训练好的网络模型,以获得所述目标图像对应的文字说明。

[0103] 其中,所述第五装置45获取行人重识别任务的待识别的目标图像,其中,所述待识别的目标图像没有对应的文字说明。

[0104] 其中,所述第六装置46将获得的所述待识别的目标图像进行图像编码,获得所述待识别的目标图像对应的图像编码向量。其中,图像编码向量的长度可根据实际需要设置,比如,设置为512维。

[0105] 其中,所述第七装置47将所述待识别的目标图像对应的图像编码向量输入训练好的网络模型,以获得所述待识别的目标图像对应的文字说明,进一步地,将获得所述待识别的目标图像对应的文字说明保存为所述待识别的目标图像对应的标注文件。

[0106] 进一步地,遍历所述目标图像所在图像数据集中的每一张图像,可以高效准确地自动为所述目标图像所在图像数据集标注文字说明,得到带有文字说明标注的图像数据集。比如采用所述方法遍历Duke-MTMC数据集中的每张图像,可以高效准确地获得Duke-MTMC数据集的文字说明标注,得到带有文字说明标注的Duke-MTMC数据集。

[0107] 可选地,所述设备还可包括:

[0108] 第八装置48(未示出),用于在所述目标图像进行编码之前,将所述目标图像的图片风格与所述图像训练集中的图片风格进行统一。

[0109] 以消除不同图像的图片风格差异,提升训练好的网络模型预测的泛化能力。例如,可采用SPGAN(Similarity Preserving cycle-consistent Generative Adversarial Network,)方法将所述待识别的目标图像的图片风格迁移为所述图像训练集的样本图像的图片风格,比如,可将用于行人重识别任务的Duke-MTMC数据集中的图像作为目标图像,采用SPGAN方法,将Duke-MTMC数据集中的图像的图片风格迁移转换为CUHK-PEDES数据集的图

片风格。也可采用其它图片风格迁移方法,比如CycleGAN, Pix2Pix等图片风格迁移方法,在此,不作限定,其他任何图片风格迁移方法如适用于本申请,都处于本申请保护范围。

[0110] 进一步地,所述设备还可以包括:

[0111] 第九装置49(未示出),用于将获得的带有文字说明标注的图像数据集用于行人重识别任务。

[0112] 例如,所述第九装置49将带有文字说明的Duke-MTMC数据集用于行人重识别任务,为现有利用文字说明来辅助行人重识别任务的神经网络模型提供视觉图像特征的同时,还提供了具有丰富语义信息的文字说明,从而使基于行人重识别任务获得更好的性能。

[0113] 根据本申请的又一方面,还提供了一种计算机可读介质,所述计算机可读介质存储有计算机可读指令,所述计算机可读指令可被处理器执行以实现前述方法。

[0114] 根据本申请的又一方面,还提供了一种用于优化预测雷达回波图像的设备,其中,该设备包括:

[0115] 一个或多个处理器;以及

[0116] 存储有计算机可读指令的存储器,所述计算机可读指令在被执行时使所述处理器执行如前述方法的操作。

[0117] 例如,计算机可读指令在被执行时使所述一个或多个处理器:首先获取图像训练集,其中,所述图像训练集中的每个样本包含图像和与该图像对应的文字说明,然后基于所述图像训练集中的样本,确定所述样本中图像的图像编码向量及与该图像对应的文字说明的文字编码向量;再将所述图像编码向量及所述文字编码向量进行关联,获得所述样本的关联特征向量;再接着将所述关联特征向量输入神经网络进行训练,当所述神经网络的损失函数满足预设条件,完成所述神经网络的训练,以获得训练好的网络模型。还可以继续获取待生成文字说明的目标图像;接着基于所述目标图像,确定所述目标图像的图像编码向量;然后将所述图像编码向量输入所述训练好的网络模型,以获得所述目标图像对应的文字说明。再可以将所述目标图像的图片风格与所述图像训练集中的图片风格进行统一。

[0118] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。装置权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一,第二等词语用来表示名称,而并不表示任何特定的顺序。

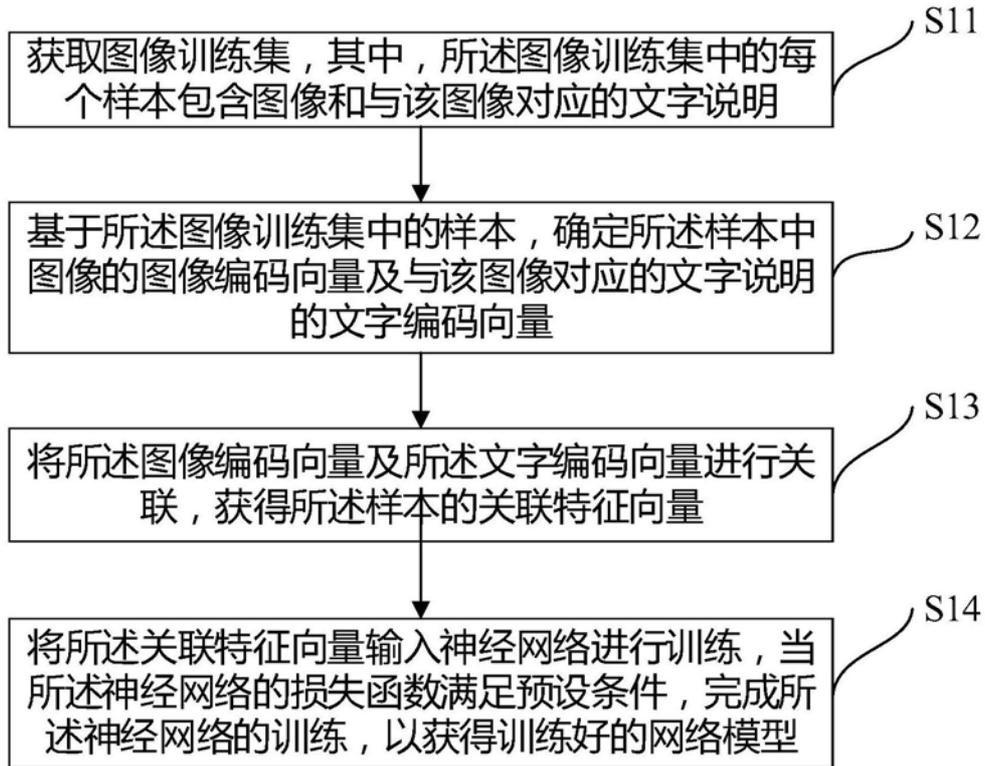


图1

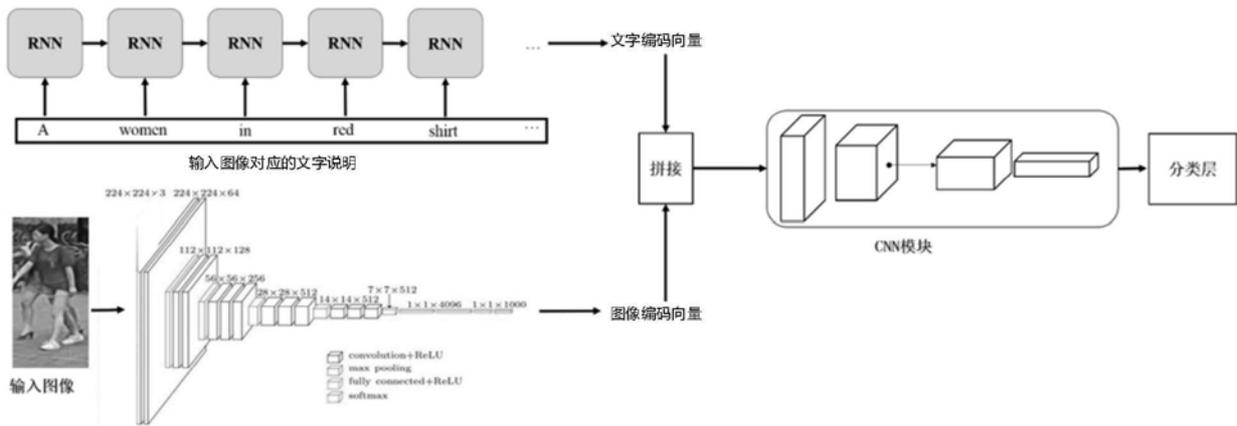


图2

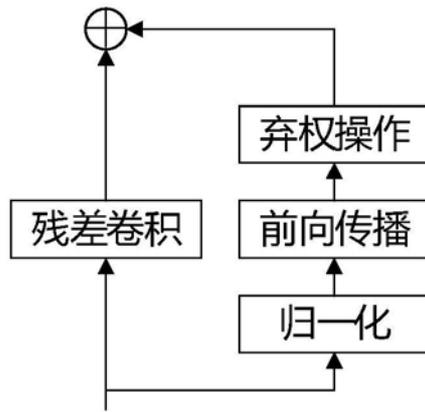


图3

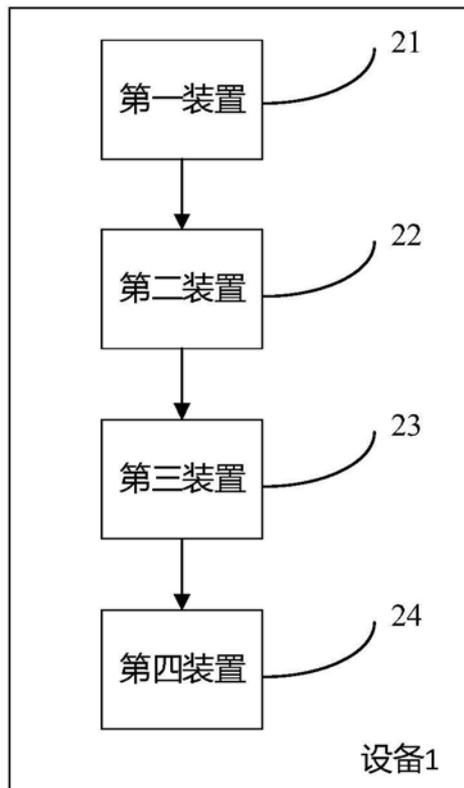


图4