



(12) 发明专利申请

(10) 申请公布号 CN 117745819 A

(43) 申请公布日 2024.03.22

(21) 申请号 202311759198.7

G06N 3/0464 (2023.01)

(22) 申请日 2023.12.18

G06N 3/08 (2023.01)

(71) 申请人 咪咕新空文化科技(厦门)有限公司

地址 361008 福建省厦门市思明区软件园
二期望海路8号

申请人 咪咕文化科技有限公司
中国移动通信集团有限公司

(72) 发明人 洪毅强 王琦 胡良军 陈佳悦
罗德海

(74) 专利代理机构 北京银龙知识产权代理有限公司 11243
专利代理人 崔姬玉

(51) Int.Cl.

G06T 7/73 (2017.01)

G06N 3/045 (2023.01)

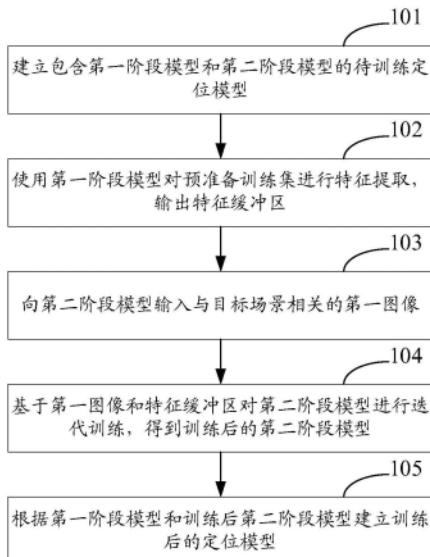
权利要求书2页 说明书11页 附图5页

(54) 发明名称

定位模型训练方法、图像定位方法、装置及可读存储介质

(57) 摘要

本申请公开了一种定位模型训练方法、图像定位方法、装置及可读存储介质，属于通信领域，定位模型训练方法包括：使用第一阶段模型对预准备训练集进行特征提取，输出特征缓冲区；基于第一图像和特征缓冲区对第二阶段模型进行迭代训练，得到训练后的第二阶段模型；根据第一阶段模型和训练后第二阶段模型建立训练后的定位模型；图像定位方法包括：获取目标图像；将目标图像输入定位模型，输出与目标图像对应的三维向量；根据三维向量确定目标图像的相机位姿；其中，定位模型为采用所述定位模型训练方法训练得到的定位模型。



1. 一种定位模型训练方法,其特征在于,所述方法包括:
建立包含第一阶段模型和第二阶段模型的待训练定位模型;
使用所述第一阶段模型对预准备训练集进行特征提取,输出特征缓冲区;
向所述第二阶段模型输入与目标场景相关的第一图像;
基于所述第一图像和所述特征缓冲区对所述第二阶段模型进行迭代训练,得到训练后的所述第二阶段模型;
根据所述第一阶段模型和训练后第二阶段模型建立训练后的定位模型;
其中,所述第一阶段模型的输出为所述第二阶段模型的输入,所述预准备训练集中包含基于多个场景以及每个场景对应的多个视角获取的多个训练图像。
2. 根据权利要求1所述的方法,其特征在于,所述第一阶段模型对所述预准备训练集进行特征提取,输出特征缓冲区,包括:
所述第一阶段模型采用离线模式对所述预准备训练集进行特征提取,输出特征缓冲区。
3. 根据权利要求1所述的方法,其特征在于,所述第一阶段模型对所述预准备训练集进行特征提取,输出特征缓冲区,包括:
对每个所述训练图像中的每个像素点,按照预设高度参数、预设宽度参数和预设通道参数确定一个对应的局部区域;
将每个所述局部区域输入所述第一阶段模型,输出每个所述局部区域对应的特征图;
将所有所述局部区域对应的特征图随机打乱存储,输出所述特征缓冲区。
4. 根据权利要求1至3任一项所述的方法,其特征在于,
所述第一阶段模型为包含卷积结构和跳跃连接结构的神经网络模型;
所述第二阶段模型为包含卷积结构的神经网络模型。
5. 根据权利要求3所述的方法,其特征在于,所述预准备训练集中还包含每个所述训练图像对应的相机位姿和相机参数,所述基于所述第一图像和所述特征缓冲区对所述第二阶段模型进行迭代训练,得到训练后的所述第二阶段模型,包括:
基于所述第一图像、所述特征缓冲区以及损失函数对所述第二阶段模型进行迭代训练;
在多次调整所述第二阶段模型的模型参数直至所述损失函数达到最小值的情况下,得到训练后的所述第二阶段模型;
其中,所述损失函数与所述第一图像的像素点坐标,所述第一图像对应的相机参数,所述第一图像对应的相机位姿,以及所述第二阶段模型预测的所述第一图像对应的三维向量相关。
6. 根据权利要求5所述的方法,其特征在于,所述基于所述第一图像、所述特征缓冲区以及损失函数对所述第二阶段模型进行迭代训练,包括:
通过以下公式对所述第二阶段模型进行迭代训练:
$$\arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T});$$
$$L = w(t) * ||p_i - K * \bar{T} * y_i||;$$
$$w(t) = \sqrt{1-t} * e_{\max} + e_{\min};$$

其中, $\arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T})$ 表示迭代所述第二阶段模型的模型参数,使得损失函数最小;

所述第二阶段模型的表达式为 $y_i = M(p_i|w)$, 所述 w 为所述第二阶段模型的模型参数, 所述 p_i 为所述第一图像的像素点坐标, 所述 R_i 为像素点 p_i 对应的局部区域, 所述 y_i 为所述第二阶段模型预测的所述第一图像对应的三维向量, 所述 L 为所述损失函数, 所述 K 为所述第一图像对应的相机参数, 所述 \bar{T} 为所述第一图像对应的相机位姿, 所述 $w(t)$ 为随训练时间变化的动态收缩因子, 所述 t 为训练时间, 所述 e_{\max} 为训练过程中出现过的最大重投影误差值, 所述 e_{\min} 为训练过程中出现过的最小重投影误差值, 所述重投影误差值为所述第二阶段模型预测的所述第一图像对应的三维向量与所述投影回所述第一图像后的误差值。

7. 一种图像定位方法,其特征在于,所述方法包括:

获取目标图像;

将所述目标图像输入定位模型,输出与所述目标图像对应的三维向量;

根据所述三维向量确定所述目标图像的相机位姿;

其中,所述定位模型为采用如权利要求1至6任一项所述的定位模型训练方法训练得到的定位模型。

8. 一种定位模型训练装置,其特征在于,所述装置包括:

第一建立模块,用于建立包含第一阶段模型和第二阶段模型的待训练定位模型;

特征提取模块,用于使用所述第一阶段模型对预准备训练集进行特征提取,输出特征缓冲区;

输入模块,用于向所述第二阶段模型输入与目标场景相关的第一图像;

训练模块,用于基于所述第一图像和所述特征缓冲区对所述第二阶段模型进行迭代训练,得到训练后的所述第二阶段模型;

第二建立模块,用于根据所述第一阶段模型和训练后第二阶段模型建立训练后的定位模型;

其中,所述第一阶段模型的输出为所述第二阶段模型的输入,所述预准备训练集中包含基于多个场景以及每个场景对应的多个视角获取的多个训练图像。

9. 一种图像定位装置,其特征在于,所述装置包括:

获取模块,用于获取目标图像;

处理模块,用于将所述目标图像输入定位模型,输出与所述目标图像对应的三维向量;

确定模块,用于根据所述三维向量确定所述目标图像的相机位姿;

其中,所述定位模型为采用如权利要求1至6任一项所述的定位模型训练方法训练得到的定位模型。

10. 一种可读存储介质,其特征在于,所述可读存储介质上存储程序或指令,所述程序或指令被处理器执行时实现如权利要求1至6任一项所述的定位模型训练方法的步骤,或者,执行时实现如权利要求7所述的图像定位方法的步骤。

定位模型训练方法、图像定位方法、装置及可读存储介质

技术领域

[0001] 本申请属于通信技术领域,具体涉及一种定位模型训练方法、图像定位方法、装置及可读存储介质。

背景技术

[0002] 现有的视觉重定位技术方案,首先基于运动结构(Structure-from-Motion,SFM)算法对目标场景进行三模建模,并存储关键帧、相机位姿、二维(2D)特征点及其匹配的三维(3D)地图点等信息。其次,执行定位时,对用户输入图像,基于图像检索的方法,找到最相似的关键帧,输入图像与相似关键帧进行2D特征匹配及3D信息关联,最后基于点透视(Perspective-n-Point,PnP)算法优化最小化重投影误差,估算出相机位姿,完成定位。

[0003] 现有技术存在以下缺点:

[0004] 1、针对每个场景,都需事先进行三维建模。

[0005] 2、存储的关键帧、特征点等信息,大小从几百M到几十G,需占用较大的容量空间,对于移动端或嵌入式设备,难以使用。

[0006] 3、对场景变化的适应性较差,当完成三维建模后,现场场景发生变化,定位的效果将受到较大影响。

发明内容

[0007] 本申请实施例提供一种定位模型训练方法、图像定位方法、装置及可读存储介质,能够解决上述现有视觉重定位技术中存在的问题。

[0008] 第一方面,提供了一种定位模型训练方法,所述方法包括:

[0009] 建立包含第一阶段模型和第二阶段模型的待训练定位模型;

[0010] 使用所述第一阶段模型对预准备训练集进行特征提取,输出特征缓冲区;

[0011] 向所述第二阶段模型输入与目标场景相关的第一图像;

[0012] 基于所述第一图像和所述特征缓冲区对所述第二阶段模型进行迭代训练,得到训练后的所述第二阶段模型;

[0013] 根据所述第一阶段模型和训练后第二阶段模型建立训练后的定位模型;

[0014] 其中,所述第一阶段模型的输出为所述第二阶段模型的输入,所述预准备训练集中包含基于多个场景以及每个场景对应的多个视角获取的多个训练图像。

[0015] 可选地,所述第一阶段模型对所述预准备训练集进行特征提取,输出特征缓冲区,包括:

[0016] 所述第一阶段模型采用离线模式对所述预准备训练集进行特征提取,输出特征缓冲区。

[0017] 可选地,所述第一阶段模型对所述预准备训练集进行特征提取,输出特征缓冲区,包括:

[0018] 对每个所述训练图像中的每个像素点,按照预设高度参数、预设宽度参数和预设

通道参数确定一个对应的局部区域；

[0019] 将每个所述局部区域输入所述第一阶段模型，输出每个所述局部区域对应的特征图；

[0020] 将所有所述局部区域对应的特征图随机打乱存储，输出所述特征缓冲区。

[0021] 可选地，所述第一阶段模型为包含卷积结构和跳跃连接结构的神经网络模型；

[0022] 所述第二阶段模型为包含卷积结构的神经网络模型。

[0023] 可选地，所述预准备训练集中还包含每个所述训练图像对应的相机位姿和相机参数，所述基于所述第一图像和所述特征缓冲区对所述第二阶段模型进行迭代训练，得到训练后的所述第二阶段模型，包括：

[0024] 基于所述第一图像、所述特征缓冲区以及损失函数对所述第二阶段模型进行迭代训练；

[0025] 在多次调整所述第二阶段模型的模型参数直至所述损失函数达到最小值的情况下，得到训练后的所述第二阶段模型；

[0026] 其中，所述损失函数与所述第一图像的像素点坐标，所述第一图像对应的相机参数，所述第一图像对应的相机位姿，以及所述第二阶段模型预测的所述第一图像对应的三维向量相关。

[0027] 可选地，所述基于所述第一图像、所述特征缓冲区以及损失函数对所述第二阶段模型进行迭代训练，包括：

[0028] 通过以下公式对所述第二阶段模型进行迭代训练：

$$[0029] \arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T});$$

$$[0030] L = w(t) * ||p_i - K * \bar{T} * y_i||;$$

$$[0031] w(t) = \sqrt{1-t} * e_{\max} + e_{\min};$$

[0032] 其中， $\arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T})$ 表示迭代所述第二阶段模型的模型参数，使得损失函数最小；

[0033] 所述第二阶段模型的表达式为 $y_i = M(p_i|w)$ ，所述 w 为所述第二阶段模型的模型参数，所述 p_i 为所述第一图像的像素点坐标，所述 R_i 为像素点 p_i 对应的局部区域，所述 y_i 为所述第二阶段模型预测的所述第一图像对应的三维向量，所述 L 为所述损失函数，所述 K 为所述第一图像对应的相机参数，所述 \bar{T} 为所述第一图像对应的相机位姿，所述 $w(t)$ 为随训练时间变化的动态收缩因子，所述 t 为训练时间，所述 e_{\max} 为训练过程中出现过的最大重投影误差值，所述 e_{\min} 为训练过程中出现过的最小重投影误差值，所述重投影误差值为所述第二阶段模型预测的所述第一图像对应的三维向量与所述投影回所述第一图像后的误差值。

[0034] 第二方面，提供了一种图像定位方法，所述方法包括：

[0035] 获取目标图像；

[0036] 将所述目标图像输入定位模型，输出与所述目标图像对应的三维向量；

[0037] 根据所述三维向量确定所述目标图像的相机位姿；

[0038] 其中，所述定位模型为采用如第一方面所述的定位模型训练方法训练得到的定位模型。

- [0039] 第三方面,提供了一种定位模型训练装置,所述装置包括:
- [0040] 第一建立模块,用于建立包含第一阶段模型和第二阶段模型的待训练定位模型;
- [0041] 特征提取模块,用于使用所述第一阶段模型对预准备训练集进行特征提取,输出特征缓冲区;
- [0042] 输入模块,用于向所述第二阶段模型输入与目标场景相关的第一图像;
- [0043] 训练模块,用于基于所述第一图像和所述特征缓冲区对所述第二阶段模型进行迭代训练,得到训练后的所述第二阶段模型;
- [0044] 第二建立模块,用于根据所述第一阶段模型和训练后第二阶段模型建立训练后的定位模型;
- [0045] 其中,所述第一阶段模型的输出为所述第二阶段模型的输入,所述预准备训练集中包含基于多个场景以及每个场景对应的多个视角获取的多个训练图像。
- [0046] 可选地,所述特征提取模块,具体用于:
- [0047] 所述第一阶段模型采用离线模式对所述预准备训练集进行特征提取,输出特征缓冲区。
- [0048] 可选地,所述特征提取模块,具体用于:
- [0049] 对每个所述训练图像中的每个像素点,按照预设高度参数、预设宽度参数和预设通道参数确定一个对应的局部区域;
- [0050] 将每个所述局部区域输入所述第一阶段模型,输出每个所述局部区域对应的特征图;
- [0051] 将所有所述局部区域对应的特征图随机打乱存储,输出所述特征缓冲区。
- [0052] 可选地,所述第一阶段模型为包含卷积结构和跳跃连接结构的神经网络模型;
- [0053] 所述第二阶段模型为包含卷积结构的神经网络模型。
- [0054] 可选地,所述预准备训练集中还包含每个所述训练图像对应的相机位姿和相机参数,所述训练模块,具体用于:
- [0055] 基于所述第一图像、所述特征缓冲区以及损失函数对所述第二阶段模型进行迭代训练;
- [0056] 在多次调整所述第二阶段模型的模型参数直至所述损失函数达到最小值的情况下,得到训练后的所述第二阶段模型;
- [0057] 其中,所述损失函数与所述第一图像的像素点坐标,所述第一图像对应的相机参数,所述第一图像对应的相机位姿,以及所述第二阶段模型预测的所述第一图像对应的三维向量相关。
- [0058] 可选地,所述训练模块,具体用于:
- [0059] 通过以下公式对所述第二阶段模型进行迭代训练:
- [0060]
$$\arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T});$$
- [0061]
$$L = w(t) * ||p_i - K * \bar{T} * y_i||;$$
- [0062]
$$w(t) = \sqrt{1-t} * e_{\max} + e_{\min};$$
- [0063] 其中, $\arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T})$ 表示迭代所述第二阶段模型的模型参数,使得

损失函数最小；

[0064] 所述第二阶段模型的表达式为 $y_i = M(p_i | w)$ ，所述w为所述第二阶段模型的模型参数，所述 p_i 为所述第一图像的像素点坐标，所述 R_i 为像素点 p_i 对应的局部区域，所述 y_i 为所述第二阶段模型预测的所述第一图像对应的三维向量，所述L为所述损失函数，所述K为所述第一图像对应的相机参数，所述 \bar{T} 为所述第一图像对应的相机位姿，所述 $w(t)$ 为随训练时间变化的动态收缩因子，所述t为训练时间，所述 e_{\max} 为训练过程中出现过的最大重投影误差值，所述 e_{\min} 为训练过程中出现过的最小重投影误差值，所述重投影误差值为所述第二阶段模型预测的所述第一图像对应的三维向量与所述投影回所述第一图像后的误差值。

[0065] 第四方面，提供了一种图像定位装置，所述装置包括：

[0066] 获取模块，用于获取目标图像；

[0067] 处理模块，用于将所述目标图像输入定位模型，输出与所述目标图像对应的三维向量；

[0068] 确定模块，用于根据所述三维向量确定所述目标图像的相机位姿；

[0069] 其中，所述定位模型为采用如第一方面所述的定位模型训练方法训练得到的定位模型。

[0070] 第五方面，提供了一种通信设备，该终端包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如第一方面所述的方法的步骤，或者，实现如第二方面所述的方法。

[0071] 第六方面，提供了一种可读存储介质，所述可读存储介质上存储程序或指令，所述程序或指令被处理器执行时实现如第一方面所述的方法的步骤，或者实现如第二方面所述的方法的步骤。

[0072] 第七方面，提供了一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现如第一方面所述的方法，或实现如第二方面所述的方法。

[0073] 第八方面，提供了一种计算机程序/程序产品，所述计算机程序/程序产品被存储在存储介质中，所述程序/程序产品被至少一个处理器执行以实现如第一方面所述的方法，或实现如第二方面所述的方法。

[0074] 在本申请实施例中，构建包含两个阶段模型的待训练定位模型，其中第一阶段模型对包含基于多个场景以及每个场景对应的多个视角获取的多个图像的预准备训练集进行特征提取，输出特征缓冲区，向第二阶段模型输入与目标场景相关的第一图像，基于第一图像和特征缓冲区对第二阶段模型进行迭代训练，根据第一阶段模型和训练后第二阶段模型建立训练后的定位模型第二阶段模型；

[0075] 这样，通过第一阶段模型实现与场景无关的特征提取，所获得的特征缓冲区具有通用性，可以作为固定参数使用，增强模型通用性，无需针对每个场景都重复进行三维建模、特征提取；后续在第二阶段模型训练过程中，为适用于特定场景，向第二阶段模型输入目标场景相关的第一图像，并基于第一图像和特征缓冲区对第二阶段模型进行迭代训练，一方面使用具有通用性的特征缓冲区能够提高训练第二阶段模型的梯度稳定性，另一方面，通用的特征缓冲区可以预先由第一阶段模型获得，且作为固定参数直接使用，这样在训练第二阶段模型时无需重复提取特征，只需基于与目标场景相关的第一图像和已获得的特

征缓冲区进行迭代训练,提高对场景变化的适应性,而且结合固定参数与特定场景相关的模型训练量相对较小,使保证有效预测结果的情况下,极大地加快了模型的训练速度。

附图说明

- [0076] 图1是本申请实施例提供的定位模型训练方法流程示意图;
- [0077] 图2a是本申请实施例提供的第一阶段模型结构示意图;
- [0078] 图2b是本申请实施例提供的第二阶段模型结构示意图;
- [0079] 图3是本申请实施例提供的图像定位方法流程示意图;
- [0080] 图4是本申请实施例提供的定位模型训练装置结构示意图;
- [0081] 图5是本申请实施例提供的图像定位装置结构示意图;
- [0082] 图6是本申请实施例提供的通信设备结构示意图。

具体实施方式

[0083] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚描述,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所获得的所有其他实施例,都属于本申请保护的范围。

[0084] 本申请的术语“第一”、“第二”等是用于区别类似的对象,而不用于描述特定的顺序或先后次序。应该理解这样使用的术语在适当情况下可以互换,以便本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施,且“第一”、“第二”所区别的对象通常为一类,并不限于对象的个数,例如第一对象可以是一个,也可以是多个。此外,本申请中的“和/或”表示所连接对象的至少其中之一。例如“A或B”涵盖三种方案,即,方案一:包括A且不包括B;方案二:包括B且不包括A;方案三:既包括A又包括B。字符“/”一般表示前后关联对象是一种“或”的关系。

[0085] 下面结合附图,通过一些实施例及其应用场景对本申请实施例提供的定位模型训练方法进行详细地说明。

- [0086] 参见图1,本申请实施例提供一种定位模型训练方法,方法包括:
 - [0087] 步骤101:建立包含第一阶段模型和第二阶段模型的待训练定位模型;
 - [0088] 步骤102:使用第一阶段模型对预准备训练集进行特征提取,输出特征缓冲区;
 - [0089] 步骤103:向第二阶段模型输入与目标场景相关的第一图像;
 - [0090] 步骤104:基于第一图像和特征缓冲区对第二阶段模型进行迭代训练,得到训练后的第二阶段模型;
 - [0091] 步骤105:根据第一阶段模型和训练后第二阶段模型建立训练后的定位模型;
- [0092] 上述第一阶段模型和第二阶段模型的模型结构将定位模型的训练分为两个阶段,第一阶段模型的输出为第二阶段模型的输入,即训练的过程的第一阶段由第一阶段模型完成,第二阶段由第二阶段模型完成。
- [0093] 上述预准备训练集中包含基于多个场景以及每个场景对应的多个视角获取的多个训练图像,即预先针对多个场景,按照每个场景多个视角进行图像获取,从而得到多场景多视角下的多个训练图像,例如由K个不同场景,每个场景中N个不同视角,最终获得K*N张图像。这样用于第一阶段特征提取的训练集中的图像具有通用性,避免提取特征仅能适用

于部分特定场景,即第一阶段模型是与场景无关的通用型模型。

[0094] 上述预准备训练集的获取,也即数据准备阶段,考虑到模型训练目标是根据采集的图像(记作I)预测相机位姿(记作 \bar{T}),在训练前,可以使用高精度设备或算法对场景采集进行一批次的图像带位姿数据的采集,以得到训练集。

[0095] 上述第二阶段模型使用第一阶段模型输出的特征缓冲区,并结合输入的与目标场景相关的第一图像,进行与场景相关的模型训练,即第二阶段模型是与具体目标场景相关的坐标回归模型。

[0096] 上述目标场景指的是具有应用定位模型进行图像定位的相关场景,该目标场景可以基于未来图像定位所需应用的场景确定,当需要针对特定场景进行图像定位时,只需使用该特定场景的图像对第二阶段模型进行重新训练,由于第二阶段模型已经预先获得了第一阶段模型输出的特征缓冲区,能够有效缩短基于特定场景的图像的训练过程,训练时间能够降低至数分钟,远低于现有技术中针对场景进行三维重建的数小时甚至1-2天。基于深度学习的特征提取方式,有较强的鲁棒性,能适应一定程度的场景变化。

[0097] 在本申请实施例中,构建包含两个阶段模型的待训练定位模型,其中第一阶段模型对包含基于多个场景以及每个场景对应的多个视角获取的多个图像的预准备训练集进行特征提取,输出特征缓冲区,向第二阶段模型输入与目标场景相关的第一图像,基于第一图像和特征缓冲区对第二阶段模型进行迭代训练,根据第一阶段模型和训练后第二阶段模型建立训练后的定位模型第二阶段模型;

[0098] 这样,通过第一阶段模型实现与场景无关的特征提取,所获得的特征缓冲区具有通用性,可以作为固定参数使用,增强模型通用性,无需针对每个场景都重复进行三维建模、特征提取;后续在第二阶段模型训练过程中,为适用于特定场景,向第二阶段模型输入目标场景相关的第一图像,并基于第一图像和特征缓冲区对第二阶段模型进行迭代训练,一方面使用具有通用性的特征缓冲区能够提高训练第二阶段模型的梯度稳定性,另一方面,通用的特征缓冲区可以预先由第一阶段模型获得,且作为固定参数直接使用,这样在训练第二阶段模型时无需重复提取特征,只需基于与目标场景相关的第一图像和已获得的特征缓冲区进行迭代训练,提高对场景变化的适应性,而且结合固定参数与特定场景相关的模型训练量相对较小,使保证有效预测结果的情况下,极大地加快了模型的训练速度。

[0099] 在一种可能的实施方式中,第一阶段模型对预准备训练集进行特征提取,输出特征缓冲区,包括:

[0100] 第一阶段模型采用离线模式对预准备训练集进行特征提取,输出特征缓冲区。

[0101] 在本申请实施例中,由于第一阶段模型具体用于提取具有通用性的特征缓冲区,这一步是与场景无关的特征提取处理,完成提取的特征缓冲区可以作为固定参数在后续第二阶段模型的多次重复训练中重复使用,因此第一阶段模型是更换新场景后,无需重新训练的模型,第一阶段模型可以基于预准备训练集消耗较长的训练时间来获得特征缓冲区,可选地,第一阶段模型训练得到特征缓冲区可以作为离线训练,无需在意训练时间的长短。

[0102] 在一种可能的实施方式中,第一阶段模型对预准备训练集进行特征提取,输出特征缓冲区,包括:

[0103] (1)对每个训练图像中的每个像素点,按照预设高度参数、预设宽度参数和预设通道参数确定一个对应的局部区域;

[0104] (2) 将每个局部区域输入第一阶段模型,输出每个局部区域对应的特征图;

[0105] (3) 将所有局部区域对应的特征图随机打乱存储,输出特征缓冲区。

[0106] 在本申请实施例中,不同于一般的特征提取模型,是直接对整张输入图像直接进行全局特征提取,得到稀疏的特征点位。本申请实施例中,第一阶段模型是针对像素级的点位进行特征预测,得到稠密的特征点位,该特征提取方式可以称之为点位特征提取,该第一阶段模型也可以称之为稠密特征提取模型,该方式需消耗的训练时间会远大于全局稀疏特征方式,但由于本步骤设计为场景无关的模型,即可作为离线训练,而无需在意训练时间的长短。

[0107] 具体涉及如下:

[0108] 对于输入图像 I ,在每一个像素点(坐标记为 p_i)附近,取一个局部区域,大小为 $H_p * W_p * C_3$,其中三个参数分别为区域大小的高,宽,通道(对应于RBG 3通道),可选的高、宽参数取值范围为(9x9 ~ 36x36,可以理解的是参数取值范围可以根据实际应用场景灵活调整,本申请实施例对参数取值范围不做具体限定)。取大量不同场景中不同视角(即相机位姿,记为 \bar{T}_i)的输入图像,按以上方式,进行局部区域分割,构建多批次训练数据集(即上述预准备训练集)。

[0109] 在一种可能的实施方式中,第一阶段模型为包含卷积结构和跳跃连接结构的神经网络模型;

[0110] 即基于卷积及跳跃连接方式,搭建第一阶段模型,输入为上述的批次训练集,输出为512维的高维度通用特征图(这里的512维度仅为举例,具有可以根据应用场景需求灵活调整)。一种可选的模型结构如图2a所示。

[0111] 对于上述预准备训练集的举例,对每个训练图像进行局部区域确定,得到K个不同场景,每个场景中N个不同视角图片,每张图片中M个不同局部区域组成的批次训练集,通过上述稠密特征提取模型后,得到K*N*M个512维的通用特征图。为了增强模型通用性及训练二阶段模型梯度稳定性,将所有特征图随机打乱存储,建立特征缓冲区。

[0112] 在一种可能的实施方式中,预准备训练集中还包含每个训练图像对应的相机位姿和相机参数,基于第一图像和特征缓冲区对第二阶段模型进行迭代训练,得到训练后的第二阶段模型,包括:

[0113] (1) 基于第一图像、特征缓冲区以及损失函数对第二阶段模型进行迭代训练;

[0114] (2) 在多次调整第二阶段模型的模型参数直至损失函数达到最小值的情况下,得到训练后的第二阶段模型;

[0115] 其中,损失函数与第一图像的像素点坐标,第一图像对应的相机参数,第一图像对应的相机位姿,以及第二阶段模型预测的第一图像对应的三维向量相关。

[0116] 在本申请实施例中,预准备训练集中还包含每个训练图像对应的相机位姿和相机参数,即在前述数据准备阶段,在获取训练图像时,连同图像 I_i 及其对应的相机位姿 \bar{T}_i ,相机参数K一并获取,共同组成预准备训练集。

[0117] 在本申请实施例中,更换新场景后,需重新训练模型,以取得最佳的预测效果,具体地使用与目标场景相关的第一图像对第二阶段模型进行迭代训练,可选地,第二阶段模型为包含卷积结构的神经网络模型,为了达到较强的实时性,适应场景更换需求,模型设计以1x1卷积为主,体积较小,可在10分钟以内完成训练,模型的输入为特征缓冲区中的512维

特征图,输出为一个3维向量,表示预测的场景中的3D点,一种可选的模型结构如图2b所示。

[0118] 进一步地,考虑到由于单纯用与目标场景相关的第一图像对第二阶段模型进行迭代训练,训练过程中,会产生较多的无效预测点(即预测的3D点投影回2D图像后,超出了图像的边界),为了引导模型更多的关注有效预测,在模型迭代训练中引入损失函数,即基于第一图像、特征缓冲区以及损失函数对第二阶段模型进行迭代训练。

[0119] 在一种可能的实施方式中,基于第一图像、特征缓冲区以及损失函数对第二阶段模型进行迭代训练,包括:

[0120] 通过以下公式对第二阶段模型进行迭代训练:

$$[0121] \arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T});$$

$$[0122] L = w(t) * ||p_i - K * \bar{T} * y_i||;$$

$$[0123] w(t) = \sqrt{1-t} * e_{\max} + e_{\min};$$

[0124] 其中, $\arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T})$ 表示迭代第二阶段模型的模型参数,使得损失函数最小;

[0125] 第二阶段模型的表达式为 $y_i = M(p_i|w)$, w 为第二阶段模型的模型参数, p_i 为第一图像的像素点坐标, R_i 为像素点 p_i 对应的局部区域, y_i 为第二阶段模型预测的第一图像对应的三维向量, L 为损失函数, K 为第一图像对应的相机参数, \bar{T} 为第一图像对应的相机位姿, $w(t)$ 为随训练时间变化的动态收缩因子, t 为训练时间, e_{\max} 为训练过程中出现过的最大重投影误差值, e_{\min} 为训练过程中出现过的最小重投影误差值, 重投影误差值为第二阶段模型预测的第一图像对应的三维向量与投影回第一图像后的误差值。

[0126] 在本申请实施例中,设计一种动态缩放的损失函数进行模型迭代训练,具体步骤如下:

[0127] 记当前待训练的图像为 I , 对应的相机位姿为 \bar{T} , 其像素点坐标集合为 $\{p_i\}$, 点 p_i 周边 $H_p * W_p$ 小区域为 $\{R_i\}$, 相机内参为 K , 步骤二所述的模型表达式记为 $y_i = M(p_i|w)$ (w 为模型待训练的参数, 表达式含义为输入图像2D像素点坐标 p_i , 根据模型参数 w , 预测对应的3D点 y_i)。

[0128] 模型的迭代训练方式如下:

$$[0129] \arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T})$$

[0130] 即不断迭代优化模型参数 w , 使得总损失函数 L 最小。

[0131] 损失函数 L 的定义如下:

$$[0132] L = w(t) * ||p_i - K * \bar{T} * y_i||$$

[0133] 其中 p_i 为图像像素点坐标, K 为相机参数, \bar{T} 为步骤一采集的相机位姿真值, y_i 为 $M(p_i|w)$, 即模型预测的3D点坐标, $w(t)$ 为随训练时间变化的动态收缩因子, 定义如下:

$$[0134] w(t) = \sqrt{1-t} * e_{\max} + e_{\min}$$

[0135] 其中 t 为训练时间, e_{\max} , e_{\min} 为当前训练过程中出现过的最大、最小重投影误差值。

[0136] 参见图3, 本申请实施例提供一种图像定位方法, 方法包括:

[0137] 步骤301: 获取目标图像;

- [0138] 步骤302:将目标图像输入定位模型,输出与目标图像对应的三维向量;
- [0139] 步骤303:根据三维向量确定目标图像的相机位姿;
- [0140] 上述定位模型为采用上文所述的定位模型训练方法,得到训练后的定位模型。
- [0141] 在本申请实施例中,在具体用户使用图像定位应用的过程中,当用户开启摄像头,请求进行视觉定位时,系统接受摄像头图像(即上述目标图像),并由上述定位模型进行预测,得到2D-3D点匹配关系(即目标图像与对应的三维向量的匹配关系)。有了匹配关系后,即可通过现有算法进行相机位姿的估算(例如PNP算法、随机抽样一致性(Random Sample Consensus,RANSAC)),完成定位。
- [0142] 参见图4,本申请实施例提供一种定位模型训练装置,所述装置包括:
- [0143] 第一建立模块401,用于建立包含第一阶段模型和第二阶段模型的待训练定位模型;
- [0144] 特征提取模块402,用于使用所述第一阶段模型对预准备训练集进行特征提取,输出特征缓冲区;
- [0145] 输入模块403,用于向所述第二阶段模型输入与目标场景相关的第一图像;
- [0146] 训练模块404,用于基于所述第一图像和所述特征缓冲区对所述第二阶段模型进行迭代训练,得到训练后的所述第二阶段模型;
- [0147] 第二建立模块405,用于根据所述第一阶段模型和训练后第二阶段模型建立训练后的定位模型;
- [0148] 其中,所述第一阶段模型的输出为所述第二阶段模型的输入,所述预准备训练集中包含基于多个场景以及每个场景对应的多个视角获取的多个训练图像。
- [0149] 可选地,所述特征提取模块,具体用于:
- [0150] 所述第一阶段模型采用离线模式对所述预准备训练集进行特征提取,输出特征缓冲区。
- [0151] 可选地,所述特征提取模块,具体用于:
- [0152] 对每个所述训练图像中的每个像素点,按照预设高度参数、预设宽度参数和预设通道参数确定一个对应的局部区域;
- [0153] 将每个所述局部区域输入所述第一阶段模型,输出每个所述局部区域对应的特征图;
- [0154] 将所有所述局部区域对应的特征图随机打乱存储,输出所述特征缓冲区。
- [0155] 可选地,所述第一阶段模型为包含卷积结构和跳跃连接结构的神经网络模型;
- [0156] 所述第二阶段模型为包含卷积结构的神经网络模型。
- [0157] 可选地,所述预准备训练集中还包含每个所述训练图像对应的相机位姿和相机参数,所述训练模块,具体用于:
- [0158] 基于所述第一图像、所述特征缓冲区以及损失函数对所述第二阶段模型进行迭代训练;
- [0159] 在多次调整所述第二阶段模型的模型参数直至所述损失函数达到最小值的情况下,得到训练后的所述第二阶段模型;
- [0160] 其中,所述损失函数与所述第一图像的像素点坐标,所述第一图像对应的相机参数,所述第一图像对应的相机位姿,以及所述第二阶段模型预测的所述第一图像对应的三

维向量相关。

[0161] 可选地,所述训练模块,具体用于:

[0162] 通过以下公式对所述第二阶段模型进行迭代训练:

[0163] $\arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T});$

[0164] $L = w(t) * ||p_i - K * \bar{T} * y_i||;$

[0165] $w(t) = \sqrt{1 - t} * e_{\max} + e_{\min};$

[0166] 其中, $\arg \min_w \sum_i L(p_i, M(R_i|w), \bar{T})$ 表示迭代所述第二阶段模型的模型参数,使得损失函数最小;

[0167] 所述第二阶段模型的表达式为 $y_i = M(p_i|w)$, 所述 w 为所述第二阶段模型的模型参数, 所述 p_i 为所述第一图像的像素点坐标, 所述 R_i 为像素点 p_i 对应的局部区域, 所述 y_i 为所述第二阶段模型预测的所述第一图像对应的三维向量, 所述 L 为所述损失函数, 所述 K 为所述第一图像对应的相机参数, 所述 \bar{T} 为所述第一图像对应的相机位姿, 所述 $w(t)$ 为随训练时间变化的动态收缩因子, 所述 t 为训练时间, 所述 e_{\max} 为训练过程中出现过的最大重投影误差值, 所述 e_{\min} 为训练过程中出现过的最小重投影误差值, 所述重投影误差值为所述第二阶段模型预测的所述第一图像对应的三维向量与所述投影回所述第一图像后的误差值。

[0168] 参见图5,本申请实施例提供了一种图像定位装置,所述装置包括:

[0169] 获取模块501,用于获取目标图像;

[0170] 处理模块502,用于将所述目标图像输入定位模型,输出与所述目标图像对应的三维向量;

[0171] 确定模块503,用于根据所述三维向量确定所述目标图像的相机位姿;

[0172] 其中,所述定位模型为采用如第一方面所述的定位模型训练方法训练得到的定位模型。

[0173] 参见图6,本发明实施例提供一种通信设备600,包括:处理器601、收发机602、存储器603和总线接口。

[0174] 其中,处理器601可以负责管理总线架构和通常的处理。存储器603可以存储处理器601在执行操作时所使用的数据。

[0175] 本发明实施例中,通信设备600还可以包括:存储在存储器603上并可在处理器601上运行的程序,该程序被处理器601执行时实现本发明实施例提供的方法的步骤。

[0176] 在图6中,总线架构可以包括任意数量的互联的总线和桥,具体由处理器601代表的一个或多个处理器和存储器603代表的存储器的各种电路链接在一起。总线架构还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起,这些都是本领域所公知的,因此,本发明实施例不再对其进行进一步描述。总线接口提供接口。收发机602可以是多个元件,即包括发送机和接收机,提供用于在传输介质上与各种其他装置通信的单元。

[0177] 本申请实施例还提供一种可读存储介质,所述可读存储介质上存储有程序或指令,该程序或指令被处理器执行时实现上述方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0178] 其中,所述处理器为上述实施例中所述的终端中的处理器。所述可读存储介质,包括计算机可读存储介质,如计算机只读存储器ROM、随机存取存储器RAM、磁碟或者光盘等。在一些示例中,可读存储介质可以是非瞬态的可读存储介质。

[0179] 本申请实施例另提供了一种芯片,所述芯片包括处理器和通信接口,所述通信接口和所述处理器耦合,所述处理器用于运行程序或指令,实现上述方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0180] 应理解,本申请实施例提到的芯片还可以称为系统级芯片,系统芯片,芯片系统或片上系统芯片等。

[0181] 本申请实施例另提供了一种计算机程序/程序产品,所述计算机程序/程序产品被存储在存储介质中,所述计算机程序/程序产品被至少一个处理器执行以实现上述方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0182] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。此外,需要指出的是,本申请实施方式中的方法和装置的范围不限按示出或讨论的顺序来执行功能,还可包括根据所涉及的功能按基本同时的方式或按相反的顺序来执行功能,例如,可以按不同于所描述的次序来执行所描述的方法,并且还可以添加、省去或组合各种步骤。另外,参照某些示例所描述的特征可在其他示例中被组合。

[0183] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助计算机软件产品加必需的通用硬件平台的方式来实现,当然也可以通过硬件。该计算机软件产品存储在存储介质(如ROM、RAM、磁碟、光盘等)中,包括若干指令,用以使得终端或者网络侧设备执行本申请各个实施例所述的方法。

[0184] 上面结合附图对本申请的实施例进行了描述,但是本申请并不局限于上述的具体实施方式,上述的具体实施方式仅仅是示意性的,而不是限制性的,本领域的普通技术人员在本申请的启示下,在不脱离本申请宗旨和权利要求所保护的范围情况下,还可做出很多形式的实施方式,这些实施方式均属于本申请的保护之内。

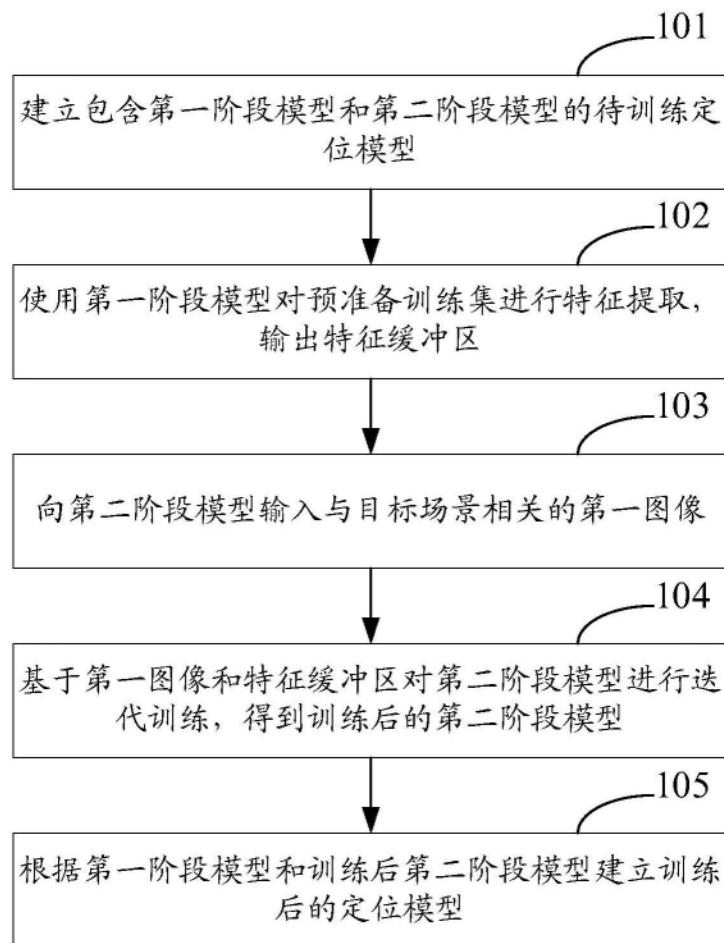


图1

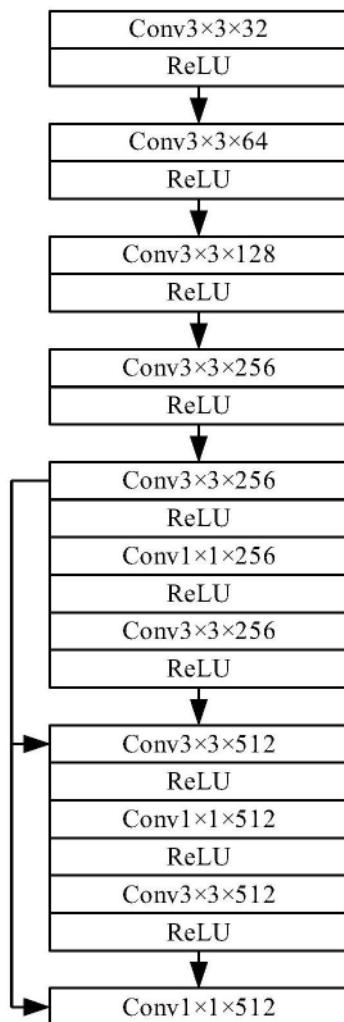


图2a

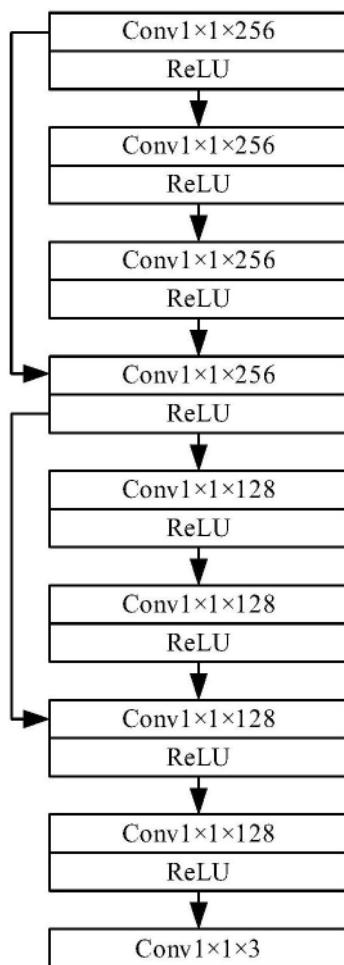


图2b

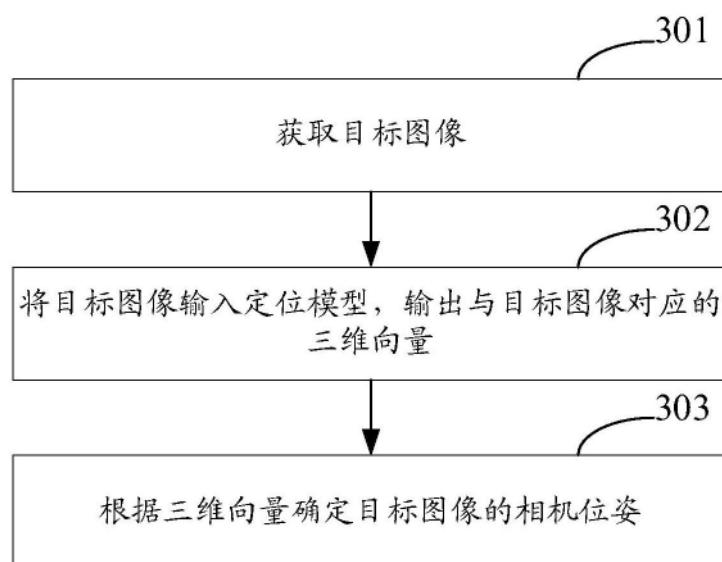


图3

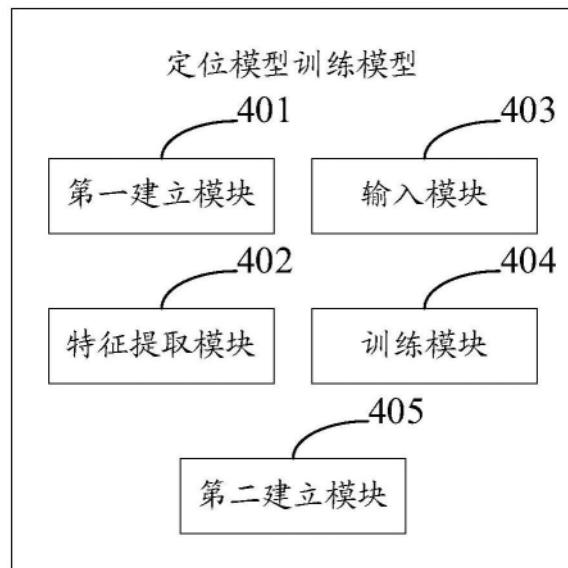


图4

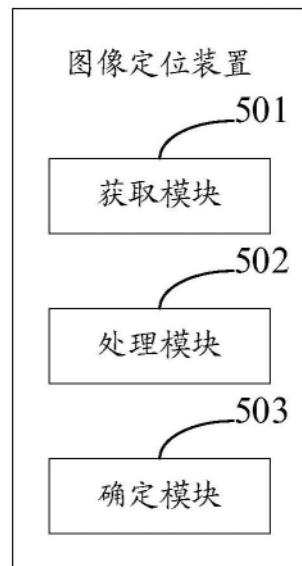


图5

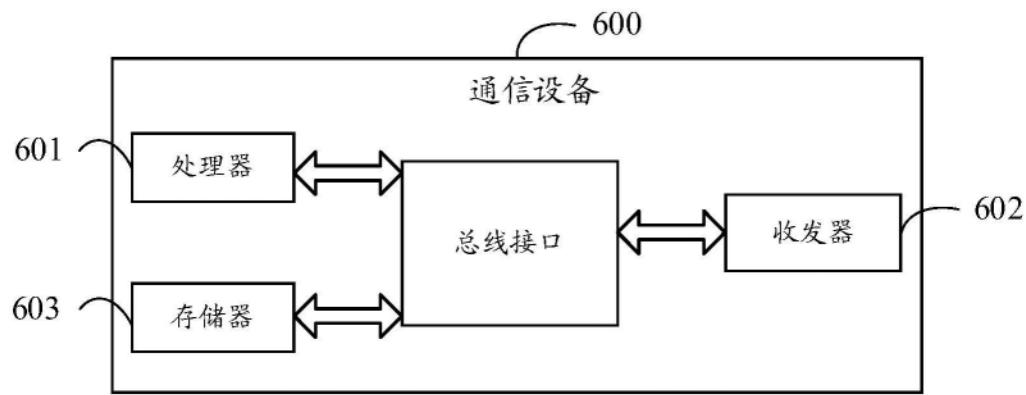


图6