(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110769985 B
(45)授权公告日 2023.10.17

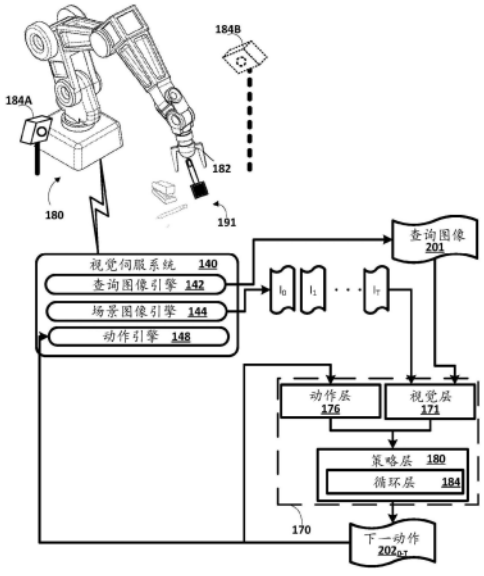
(21)申请号 201880040068.9
(22)申请日 2018.12.04
(65)同一申请的已公布的文献号
申请公布号 CN 110769985 A
(43)申请公布日 2020.02.07
(30)优先权数据
62/595,037 2017.12.05 US
(85)PCT国际申请进入国家阶段日
2019.12.12
(86)PCT国际申请的申请数据
PCT/US2018/063843 2018.12.04
(87)PCT国际申请的公布数据
WO2019/113067 EN 2019.06.13
(73)专利权人 谷歌有限责任公司
地址 美国加利福尼亚州
(72)发明人 A.托谢夫 F.萨德吉 S.莱维恩
(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 金玉洁
(51)Int.Cl.
B25J 9/16(2006.01)
G05B 13/02(2006.01)
G06N 3/08(2023.01)
(56)对比文件
US 2017334066 A1,2017.11.23
CN 103279039 A,2013.09.04
US 2017178346 A1,2017.06.22
US 9811074 B1,2017.11.07
Josh Tobin et al..“Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World”.《IEEE》.2017,第23-30页.
FINN CHELSEA RT AL..Deep visual foresight for planning robot motion.《IEEE》.2017,第2786-2793页.
审查员 李辉
权利要求书2页 说明书16页 附图6页

(54)发明名称

使用循环神经网络的机器人末端执行器的视点不变的视觉伺服

(57)摘要

训练和/或使用循环神经网络模型用于机器人的末端执行器的视觉伺服。在视觉伺服中,该模型可以用于在多个时间步中的每个生成动作预测,该动作预测表示应如何移动末端执行器以使末端执行器向目标对象移动的预测。该模型可以是视点不变的,因为它可以跨在各种视点具有视觉组件的各种机器人使用,和/或即使机器人的视觉组件的视点剧烈改变,也可以用于单个机器人。此外,可以基于大量的模拟数据来训练模型,该模拟数据基于关于模型执行模拟片段的模拟器。可以基于相对较少数量的真实训练数据来进一步训练模型的一个或多个部分。



CN 110769985 B

1.一种对机器人的末端执行器进行伺服的方法,包括:

确定查询图像,所述查询图像包括将由所述机器人的末端执行器进行交互的目标对象;

在第一时间步,基于使用神经网络模型处理所述查询图像、场景图像、以及先前动作表示生成动作预测,其中所述场景图像由与机器人相关联的视觉组件捕获,并捕获目标对象和机器人的末端执行器,其中所述先前动作表示是先前时间步的先前动作预测,并且其中神经网络模型包括一个或多个循环层,每个循环层包括多个记忆单元;

基于第一时间步的动作预测控制所述机器人的末端执行器;

在第二时间步,紧随在生成第一时间步的动作预测之后生成附加的动作预测,紧随在后的动作预测基于使用神经网络模型处理所述查询图像、附加场景图像和动作预测而生成,其中所述附加场景图像在基于第一时间步的动作预测控制末端执行器之后由视觉组件捕获,并捕获目标对象和末端执行器;以及

基于附加的动作预测,控制机器人的末端执行器。

2.根据权利要求1所述的方法,其中,基于使用神经网络模型处理所述查询图像、所述场景图像、以及所述先前动作表示而生成第一时间步的动作预测包括:

使用神经网络模型的视觉部分的多个视觉层来处理所述查询图像和所述场景图像,以生成视觉层输出;

使用神经网络模型的动作部分的一个或多个动作层处理所述先前的动作表示以生成动作输出;以及

将视觉层输出和动作输出组合并使用神经网络模型的多个策略层处理组合的视觉层输出和动作输出,所述策略层包括一个或多个循环层。

3.根据权利要求2所述的方法,其中,所述一个或多个循环层的多个记忆单元包括长短期记忆单元。

4.根据权利要求2或3所述的方法,其中,使用所述神经网络模型的视觉部分的多个视觉层来处理所述查询图像和所述场景图像以生成视觉层输出包括:

在所述视觉层的第一卷积神经网络部分上处理所述查询图像,以生成查询图像嵌入;

在所述视觉层的第二卷积神经网络部分上处理所述场景图像,以生成场景图像嵌入;以及

基于所述查询图像嵌入和所述场景图像嵌入生成所述视觉层输出。

5.根据权利要求4所述的方法,其中,基于所述查询图像嵌入和所述场景图像嵌入生成所述视觉层输出包括:在所述视觉层的一个或多个附加层上处理所述查询图像嵌入和所述场景图像嵌入。

6.根据权利要求1-3中的任一所述的方法,其中,所述第一时间步的动作预测表示在所述机器人的机器人坐标系中用于所述末端执行器的位移的速度矢量。

7.根据权利要求1-3中的任一所述的方法,其中,确定所述查询图像基于来自用户的用户接口输入。

8.根据权利要求7所述的方法,其中,所述用户接口输入是键入的或说出的用户接口输入,并且其中,基于来自用户的用户接口输入确定所述查询图像包括:

从多个库存图像中选择所述查询图像,基于与选择的查询图像关联的数据,匹配基于

用户接口输入确定的一个或多个项。

9.根据权利要求7所述的方法，其中，基于来自用户的用户接口输入确定所述查询图像包括：

使所述场景图像或先前场景图像经由计算设备呈现给用户；

其中，用户接口输入经由计算设备接收，并且指示所呈现的场景图像或先前的场景图像的子集；以及

基于所述场景图像或所述先前场景图像的裁剪生成所述查询图像，其中，基于用户接口输入确定裁剪。

10.根据权利要求1-3中的任一所述的方法，其中，基于由所述机器人的视觉组件捕获的图像来生成所述查询图像。

11.根据权利要求1-3中的任一所述的方法，其中，所述查询图像、所述场景图像和所述附加场景图像均为二维图像。

12.一种系统，包括：存储器，其存储指令；以及一个或多个处理器，可操作以运行所述指令来使得执行前述权利要求中的任一项所述的方法。

13.一种真实的机器人，包括：存储器，其存储指令；以及一个或多个处理器，可操作以运行所述指令来使得执行权利要求1-11中的任一项所述的方法。

14.一种计算机可读介质，其包括指令，当所述指令由计算机执行时，所述指令使得所述计算机执行权利要求1-11中任一项的方法。

使用循环神经网络的机器人末端执行器的视点不变的视觉伺服

背景技术

[0001] 许多机器人被配置为利用一个或多个末端执行器来执行一个或多个机器人任务，诸如抓握和/或其他操纵任务。例如，机器人可以利用抓握末端执行器，诸如“冲击式(impactive)”抓手或“侵入式(ingressive)”抓手(例如，使用大头针、针等物理穿透对象)来从第一位置拾取对象，移动对象到第二位置，并且在第二位置放下对象。可抓握对象的机器人末端执行器的一些其他示例包括“收缩式(astrictive)”末端执行器(例如，使用吸力或真空来拾取对象)和“接触式(contigutive)”末端执行器(例如，使用表面张力、冷冻或粘合剂来拾取对象)。

[0002] 已经提出了用于诸如抓握的机器人操纵任务的各种基于机器学习的方法。其中一些方法训练机器学习模型(例如，前馈深度神经网络)以生成用于机器人抓握的视觉伺服中利用的一个或多个预测，并使用基于来自试图对各种对象进行机器人抓握的真实世界的物理机器人的数据的训练示例来训练机器学习模型。例如，可以基于用于迭代的对应图像和用于迭代的候选运动矢量来训练机器学习模型，以预测在多个迭代中的每个迭代上成功抓握的可能性。对应图像可以是机器人的视觉传感器捕获的最新图像，并且候选运动矢量可以是机器人正在考虑实施的运动矢量。基于每次迭代成功抓握的可能性，可以确定是否试图抓握或替代地实现候选运动矢量，并执行预测成功抓握的可能性的另一迭代。

[0003] 然而，这些和/或其他方法可能具有一个或多个缺点。例如，至少部分地基于来自机器人的视觉组件的输入图像生成用于由机器人使用的预测的一些机器学习模型，在视觉组件具有相对于训练了机器学习模型的视点略有变化的视点的情况下可以是强健的，但是由于视点的严重变化可能不准确和/或失败。例如，各种方法训练机器学习模型以至少部分地基于训练示例的输入图像来生成对机器人的操纵预测，其中用于训练示例的输入图像全部是从相同或相似的视点捕获的。尽管这样的机器学习模型适用于从相同或相似的视点捕获图像的机器人，但它们用于在从不同的视点捕获图像的机器人中使用时可能不准确和/或失败。

[0004] 作为缺点的另一示例，各种方法严重地或排他地依赖于基于来自真实世界的物理机器人的数据而生成的训练示例，这在试图机器人抓握或其他操纵时需要大量使用物理机器人。这可能很费时间(例如，实际上试图大量抓握需要大量时间)，可能消耗大量资源(例如，操作机器人所需的电力)，可能给正在使用的机器人造成磨损，和/或可能需要很大的人工干预(例如，放置要抓握的对象、纠正错误条件)。

发明内容

[0005] 该说明书通常针对与机器人视觉伺服有关的方法和装置。更具体地，各种实现方式针对用于训练和/或使用机器人的末端执行器的视觉伺服的循环神经网络模型的技术。在视觉伺服中，可以使用循环神经网络模型在多个时间步中的每个生成动作预测，该动作预测表示应如何移动末端执行器(例如，移动方向)以引起末端执行器移向目标对象的预

测。例如,动作预测可以指示用于来回移动末端执行器的三维(3D)(例如,“X,Y,Z”)速度。在每个时间步,该预测可以基于使用循环神经网络模型处理捕获目标对象的查询图像(例如,聚焦在目标对象上的裁剪或缩放图像)、该时间步的当前场景图像(包括目标对象、机器人的末端执行器和可选的附加场景对象)、以及表示先前时间步的动作预测的先前动作表示。先前动作表示可以是神经网络模型的先前时间步的动作预测的内部表示。在初始时间步处,先前动作表示可以是“无效”动作预测(指示没有动作),或可以基于未参考使用循环神经网络模型生成的动作预测而先前实施的动作。

[0006] 循环神经网络模型是循环的,因为它包括至少一个循环层,该循环层包括多个存储器单元,诸如长短期记忆(“LSTM”)单元和/或门控循环单元(“GRU”)。这样的循环层使得循环神经网络模型能够维持先前动作和先前场景图像的“记忆”,并鉴于此类先前动作和场景图像来适应动作预测,使得在多个时间步朝着目标对象伺服末端执行器。这也可以使循环神经网络模型对于捕获使用循环神经网络模型处理的图像的视觉传感器(例如相机)的各种视点具有强健性,因为循环层的“记忆”能够实现先前动作的结果的有效观察和未来预测动作的适当适应。因此,循环神经网络模型可以跨具有各种视点的各种视觉组件的各种机器人使用和/或甚至当机器人的视觉组件的视点发生巨大变化时(在各伺服片段(episode)之间,或甚至在一伺服片段之间)也可以用于单个机器人。

[0007] 在许多实现方式中,可以利用基于在模拟环境中相互作用的模拟机器人生成的模拟数据来训练循环神经网络模型。模拟数据由一个或多个模拟器生成,每个模拟器在一个或多个计算设备上执行,因为这些模拟器用于模拟机器人、模拟环境、以及模拟环境中的机器人的模拟动作。如本文所述,可以利用强化学习来训练循环神经网络模型,并且可以基于在模拟环境中的模拟机器人的模拟控制来生成模拟数据,其中,模拟控制基于由循环神经网络模型在训练期间生成的动作预测。

[0008] 模拟器的使用可以能够使得使用模拟器跨训练片段从各种视点渲染图像,从而提高训练的循环神经网络模型对多个视点(包括训练期间看不到的观点)的强健性和/或准确性。模拟器的使用可以额外地或替代地能够实现跨训练片段的目标对象和环境对象的有效变化,从而提高了训练的循环神经网络模型对多个场景(包括训练期间看不到的场景)的强健性和/或准确性。此外,模拟器的使用可以附加地或替代地能够实现有效地确定训练期间用于更新循环神经网络的一个或多个奖励,诸如形态奖励(shaped reward)和/或稀疏奖励(sparse reward),其对应实例可以可选地应用于训练片段的每个步。例如,可以基于由时间步的动作预测所指示的方向与对于目标对象的“基准事实(ground truth)”方向的比较(例如,之间的欧式距离)来在每个时间步确定形态奖励。由于模拟器知道了模拟目标对象的姿态,因此可以在每个时间步高效确定到目标对象的基准事实方向。此外,模拟器的使用可以附加地或替代地能够实现通过演示的使用加快学习由循环神经网络模型表示的策略,每个演示都是基于朝向相应目标对象的相应最佳方向(可选地针对强健性进行扰动),其可以基于模拟机器人的相应已知姿态和模拟目标对象来确定。

[0009] 另外,与使用基于真实物理机器人的操作生成的真实世界数据相比,使用模拟数据可以导致各种高效性。例如,每个模拟片段可以比相应的真实世界抓握片段以更少的时间执行和/或可以在多个(例如,数百个、数千个)计算设备和/或处理器上并行执行,进一步提高了模拟片段的时间效率。这些和其他考虑因素可能导致消耗更少的资源(例如,模拟片

段比相应的真实抓取片段要消耗更少的电力),可能导致物理机器人的磨损减少(例如,由于真实片段的数量减少),和/或可能需要较少的人工干预(例如,对真实世界片段的监督(oversight)较少)。

[0010] 在各种实现方式中,基于真实世界数据对基于模拟数据训练的循环神经网络模型的一个或多个部分进行进一步训练,以使循环神经网络模型适应于在真实物理机器人上使用时改进的性能。例如,循环神经网络模型可以包括视觉部分,该视觉部分用于在每个时间步处理查询图像和对应的场景图像。可以通过基于训练示例的进一步训练来适应视觉部分,其中每个训练示例包括真实查询图像和相应的真实场景图像作为训练示例输入。训练示例可各自进一步包括与动作预测无关的训练示例输出。例如,训练示例的训练示例输出可以是手动标记的一个热矢量,其中"热"值指示存在相应真实查询图像的对象的情况下的真实场景图像中的位置。可以使用视觉部分,并且可选地使用在真实世界适应中使用的(并且不用于视觉伺服中的)一个或多个附加层(例如,仿射层)、基于生成的预测输出和训练示例输出的比较确定的误差、以及用于更新视觉部分的误差(例如,通过反向传播),处理训练示例输入。以这种方式,当通过真实的物理机器人在视觉伺服中采用时,可以基于真实的训练示例来进一步训练视觉部分,以使循环神经网络模型更强健和/或更准确。在各种实现方式中,可以基于真实训练示例来更新循环神经网络模型的视觉部分,而不更新循环神经网络模型的其他部分,诸如动作部分和/或策略部分。

[0011] 提供以上描述作为本公开的一些实现方式的概述。在下面更详细地描述了那些实现方式和其他实现方式的进一步说明。

[0012] 在一些实现方式中,提供了一种对机器人的末端执行器进行伺服的方法,包括:确定查询图像,所述查询图像捕获将由所述机器人的末端执行器与之交互的目标对象。所述方法还包括:基于使用神经网络模型处理所述查询图像、场景图像、以及先前动作表示,生成动作预测。场景图像由与机器人相关联的视觉组件捕获,并捕获目标对象和机器人的末端执行器。处理中使用的神经网络模型包括一个或多个循环层,每个循环层包括多个记忆单元。所述方法还包括:基于动作预测控制机器人的末端执行器。所述方法还可包括:紧随生成动作预测之后(但在基于动作预测控制末端执行器之后)生成附加的动作预测,以及基于附加的动作预测,控制机器人的末端执行器。紧随其后的动作预测可以基于使用神经网络模型处理查询图像、附加场景图像和动作预测而生成。附加场景图像可以在基于动作预测控制末端执行器之后由视觉组件捕获,并捕获目标对象和末端执行器。

[0013] 技术的这些和其他实现方式可以包括以下特征中的一个或多个。

[0014] 在一些实现方式中,基于使用神经网络模型处理查询图像、场景图像、以及先前动作表示而生成动作预测包括:使用神经网络模型的视觉部分的多个视觉层来处理查询图像和场景图像,以生成视觉层输出;使用神经网络模型的动作部分的一个或多个动作层处理先前的动作表示以生成动作输出;将视觉层输出和动作输出组合;并使用神经网络模型的多个策略层处理组合的视觉层输出和动作输出。在那些实现方式中,所述策略层包括一个或多个循环层。在那些实现方式中的一些中,所述一个或多个循环层的所述多个记忆单元包括长短期记忆单元。在那些实现方式中的一些中,使用所述神经网络模型的视觉部分的多个视觉层来处理所述查询图像和所述场景图像以生成视觉层输出包括:在所述视觉层的第一卷积神经网络部分上处理所述查询图像,以生成查询图像嵌入;在所述视觉层的第二

卷积神经网络部分上处理所述场景图像，以生成场景图像嵌入；以及基于查询图像嵌入和场景图像嵌入生成视觉层输出。基于所述查询图像嵌入和所述场景图像嵌入生成所述视觉层输出可以包括：在所述视觉层的一个或多个附加层上处理所述查询图像嵌入和所述场景图像嵌入。

[0015] 在一些实现方式中，所述动作预测表示在所述机器人的机器人坐标系(frame)中用于所述末端执行器的位移的速度矢量。

[0016] 在一些实现方式中，确定所述查询图像基于来自用户的用户接口输入。在那些实现方式的一些版本中，所述用户接口输入是键入的或说出的用户接口输入，并且基于用户接口输入来确定所述查询图像包括：从多个库存图像中选择所述查询图像，基于与选择的查询图像关联的数据，匹配基于用户接口输入确定的一个或多个项。在那些实现方式的一些其他版本中，基于用户接口输入确定查询图像包括：使场景图像或先前场景图像经由计算设备呈现给用户；以及基于场景图像或先前的场景图像的裁剪生成查询图像，其中，裁剪是基于用户接口输入确定的。并且其中，用户接口输入经由计算设备接收，并且指示所呈现的场景图像或先前的场景图像的子集。在一些实现方式中，基于由所述机器人的视觉组件捕获的图像来生成所述查询图像。

[0017] 在一些实现方式中，所述查询图像、所述场景图像和所述附加场景图像均为二维图像或均为2.5维图像。

[0018] 在一些实现方式中，提供了一种训练用于伺服机器人的末端执行器的神经网络模型的方法。训练的方法包括：对于使用机器人模拟器执行的多个模拟片段中的每一个，确定由机器人模拟器渲染的渲染查询图像。所述渲染查询图像捕获机器人模拟器的对应模拟环境的对应模拟目标对象。所述方法还包括，对于多个实例中的每个，对于模拟片段中的每个，直到满足一个或多个条件为止：生成动作预测，基于动作预测并至少部分基于来自机器人模拟器的基准事实数据生成实例的奖励；基于奖励更新神经网络模型的至少一部分；以及使得机器人模拟器基于所述实例的下一实例之前的动作预测控制模拟机器人的模拟的末端执行器。生成动作预测可以基于使用神经网络模型处理渲染查询图像、实例的渲染场景图像以及实例的先前动作表示。实例的渲染场景图像可以使用机器人模拟器渲染，并可以在实例的任何先前实例之后捕获模拟目标对象和模拟机器人的模拟末端执行器。先前动作表示可以基于实例的任何先前实例的紧挨的在前动作预测，并且神经网络模型可以包括一个或多个循环层，每个循环层包含多个记忆单元。

[0019] 技术的这些和其他实现方式可以包括以下特征中的一个或多个。

[0020] 在一些实现方式中，所述神经网络模型包括视觉部分，所述视觉部分用于处理片段的实例的渲染场景图像和渲染查询图像。在那些实现方式的一些中，所述方法还包括：利用真实图像进一步训练所述视觉部分。进一步训练真实图像的真实部分可以包括：在进一步训练期间生成损失，损失每个基于用于所述真实图像中的相应一个的相应的监督标签。所述监督标签可以用于与末端执行器伺服任务不同的任务。例如，与末端执行器伺服任务不同的任务可以是对象定位任务。

[0021] 在一些实现方式中，基于动作预测并至少部分基于来自机器人模拟器的基准事实数据生成实例的奖励包括：基于基准事实数据生成基准事实动作预测；以及基于动作预测和基准事实动作预测的比较生成奖励。在那些实现方式的一些中，所述基准事实数据包括

所述实例的模拟末端执行器的姿态和所述实例的模拟目标对象的姿态，并且基于基准事实数据生成基准事实动作预测包括：基于导致模拟末端执行器朝向模拟目标对象移动的所述基准事实动作生成所述基准事实动作预测。

[0022] 在一些实现方式中，对实例生成奖励还基于模拟片段是否导致模拟末端执行器成功到达模拟目标对象。

[0023] 在一些实现方式中，所述方法还包括，在所述多个模拟片段之前，对于使用所述机器人模拟器执行的多个先前模拟片段中的每个：对于生成奖励的模拟实现，选择特定的动作预测，基于所述特定动作预测基于按照基于来自模拟器的模拟数据确定的，朝向相应目标对象的相应最佳方向。在那些实现方式的一些中，所述特定动作预测中的一个或多个基于相应的最佳方向，具有注入的噪声。所注入的噪声可以是正态高斯噪声。

[0024] 在一些实现方式中，所述神经网络模型包括视觉部分，所述视觉部分用于处理渲染查询图像和所述片段的实例的渲染场景图像，并且所述方法还包括：识别真实的训练示例，基于使用视觉部分对真实训练示例输入的处理来生成预测输出；基于预测输出和训练示例输出确定误差；以及基于误差更新视觉部分。在那些实现方式的一些中，所述真实训练示例包括训练示例输入和训练示例输出。训练示例输入可以包括例如真实查询图像和真实场景图像，其中所述真实查询图像由真实视觉传感器捕获并捕获真实目标对象，所述真实场景图像由真实视觉传感器或附加真实视觉传感器捕获并捕获真实场景中的真实目标对象以及一个或多个附加对象。在那些实现方式的一些中，在满足所述一个或多个条件之后出现基于误差更新所述视觉部分。在那些实现方式的一些版本中，所述方法还包括，在更新视觉部分之后：提供所述神经网络模型以供一个或多个真实物理机器人进行视觉伺服使用。可选地，在提供所述神经网络模型以供所述一个或多个真实物理机器人进行视觉伺服使用之前，基于真实训练示例仅训练所述神经网络模型的视觉部分。

[0025] 其他实现方式可包括非暂时性计算机可读存储介质，其存储可由一个或多个处理器(例如，中央处理单元(CPU)、图形处理单元(GPU)、和/或张量处理单元(TPU))执行的指令，以执行诸如上述和/或本文其他地方所描述的方法中的一种或多种的方法。其他实现方式可包括一个或多个计算机的系统和/或一个或多个机器人，其包括一个或多个处理器，其可操作来运行存储的指令以执行诸如上述和/或本文其他地方所描述的方法中的一种或多种的方法。

[0026] 应当理解，本文中更详细描述的前述概念和附加概念的所有组合被认为是本文公开的主题的一部分。例如，出现在本公开的结尾处的所要求保护的主题的所有组合被认为是本文所公开的主题的一部分。

附图说明

[0027] 图1示出可以训练循环神经网络模型以用于机器人的末端执行器的视点不变视觉伺服的示例环境。

[0028] 图2示出在执行机器人的末端执行器的视觉伺服时的示例真实物理机器人以及机器人对循环神经网络模型的使用。

[0029] 图3示出图1和2的循环神经网络模型的一个示例。

[0030] 图4是示出训练用于在机器人的末端执行器的视点不变视觉伺服中使用的循环神

经网络模型的示例方法的流程图。

[0031] 图5是示出在执行机器人的末端执行器的视觉伺服时使用循环神经网络模型的示例方法的流程图。

[0032] 图6示意性地描绘机器人的示例架构。

[0033] 图7示意性地描绘计算机系统的示例架构。

具体实施方式

[0034] 本文所述的实现方式训练并利用循环神经网络模型，该循环神经网络模型可以在每个时间步用于：处理目标对象的查询图像，包括目标对象和机器人的末端执行器的当前场景图像、以及先前动作预测；并基于该处理生成预测动作，该预测动作指示对如何控制末端执行器以将末端移动到目标对象的预测。循环神经网络模型可以是视点不变的，因为它可以跨具有各种视点的视觉组件的各种机器人使用，并且/或者即使机器人的视觉组件的视点发生了巨大变化，也可以用于单个机器人。此外，可以基于大量的模拟数据来训练循环神经网络模型，该大量的模拟数据基于鉴于循环神经网络模型执行模拟片段(episode)的模拟器。可以基于相对较少量的真实训练数据来可选地进一步训练循环神经网络模型的一个或多个部分。例如，可以基于少量的真实训练数据来训练视觉部分(并且可选地，仅视觉部分)，以使循环神经网络模型适用于处理由真实机器人的视觉组件捕获的真实图像。

[0035] 即使在存在光学失真的情况下，人类也熟练地从范围广泛的各种视点和角度控制其四肢和工具。例如，大多数人可以轻松地执行任务，同时在镜子中看到自己。在机器人技术中，此类技能通常称为视觉伺服：主要使用视觉反馈将工具或端点移动到期望的位置。本文描述的实现方式涉及用于在机器人操纵场景中自动学习视点无关的视觉伺服技能的方法和装置。例如，实现方式涉及训练深层的循环神经网络模型，其可用于自动确定哪些动作将机器人手臂的端点移动到期望的对象。这样的实现方式即使在使用该模型处理的图像的视点严重变化的情况下，也能够实现使用循环神经网络模型来确定要实现的动作。本文所述的视觉伺服系统的实现方式利用了过去运动的记忆(经由循环神经网络模型的循环层)，以从被用来捕获图像的视觉组件的当前视点了解动作如何影响机器人运动，纠正实现的动作的错误以及逐渐移动靠近目标。这与许多视觉伺服技术形成了鲜明的对比，后者采用已知的动力学或涉及校准阶段。

[0036] 因此，本文所述的实现方式训练了深度神经网络，其以用于记忆的循环连接增强，以用于视点不变视觉伺服。在经典机器人技术中，视觉伺服是指控制机器人以实现图像空间中的定位目标，其通常由手动设计的关键特征的位置指定。本文公开的实现式采用了更为开放的视觉伺服的方法：通过为神经网络模型提供所期望的对象的目标“查询图像”来指定目的，并且在没有任何手动指定的特征，并且存在严重的视点变化的情况下，利用神经网络模型来选择将导致机器人到达该对象的动作。这能够实线这样的视觉伺服技术，其可以伺服目标对象(例如，经由查询图像的用户指定的用户选择)，只要与机器人相关联的视觉组件(例如，相机)实际上可以看到机器人(例如，末端执行器和可选的控制末端执行器的链路)和目标对象。根据本文公开的实现方式训练的神经网络模型被训练，以自动并且隐式学习识别动作如何影响图像空间运动，并可以概括到训练期间未见的新颖对象，该模型通过合成图像(例如，模拟环境和/或模拟机器人的渲染图像)、以及可选地使用弱标记的真实世

界视频(图像序列)的适应过程进行训练。

[0037] 因此,本文描述的各种实现方式提供了可将机器人手臂伺服到先前未见过的对象的学习视觉伺服机制。为此,其中一些实现施方式将新颖的循环卷积神经网络体系结构用于学习的视觉伺服,和/或利用使用强标记的合成图像与少量弱标记的真实世界数据相结合的新颖的训练过程。此外,在这些实现方式中的一些中,可以在模拟中生成绝大多数训练数据,并且仅使用适量的真实机器人的视频(和/或其他真实世界的图像)来通过辅助注意损失使模型适应真实世界。这种传输方法有效地将视觉表示精细调整为真实视频,同时保持网络的策略/电机控制层固定不变。

[0038] 在视觉伺服中由循环神经网络模型迭代生成的动作预测使机器人的末端执行器能够到达放置在一个或多个表面(例如,桌子)上的多个对象中的目标对象。可以通过从任意角度紧凑裁剪的该对象的图像来指定目标对象。例如,可以基于来自用户的用户接口输入(例如,绘制或以其他方式指示用于生成紧凑裁剪的图像的约束框)来指定对象,或可以基于来自更高级任务规划器的输出(例如,指示“对象X”应被遍历到下一个,并提供“对象X”的“库存”图像或“对象X”的渲染图像(例如,基于“对象X”的模型渲染)来指定对象。当利用动作预测到达目标对象时,可以使用机器人末端执行器来操纵目标对象。例如,机器人末端执行器可用于抓握、推动、拉动和/或以其他方式操纵目标对象。

[0039] 本文所述的伺服技术基于视觉反馈来适应末端执行器的控制。例如,由于循环神经网络模型用于在未探索的设置中生成动作预测,因此它响应于那些动作预测的实现和自我校准来观察其自己的运动。因此,循环神经网络模型的学习策略可以概括到新的设置或处理当前设置的变化,这在大多数先前的应用中都是经由繁琐的校准过程完成的。此外,视觉机器人系统可以意识到其自己的物理特性而无需精确模型,这使得这种方法比校准更通用。

[0040] 可以基于变化的场景布局(变化的表面、变化的表面纹理、各种各样的对象)来训练循环神经网络模型,并且可以对其进行训练和配置以了解目标对象的语义,因为它不是要到达任何目标对象,而用于指定的目标对象。这样,循环神经网络模型在3D中执行隐式对象定位。同样,通过整个训练过程中目标对象和场景对象的变化,训练了循环神经网络模型以概括其在不同形状之间的策略。

[0041] 本文描述的实现方式利用策略π_θ,该策略π_θ被实现为具有参数θ的深度神经网络。该策略输出表示在机器人坐标系中手臂的末端执行器的位移的动作$\alpha = (\partial_x, \partial_y, \partial_z)$。它是在有限时段贬损(finite-horizon discounted)的Markov决策过程(MDP)(S,A,P,R,γ)上的强化学习来训练的。状态空间S的可观察部分是场景和手臂的图像,在时间t处表示为o_t。动作空间$A=[-d,d]^3$是允许位移赞同(commend)的连续3维空间。训练时使用的形态奖励函数捕获手臂和目标对象之间的距离,并经由计算到目标对象的基准事实方向和预测方向之间的欧几里得距离来定义。除了(或代替)形态奖励函数,可以使用基于达成成功和失败的稀疏奖励函数。这样的稀疏奖励函数可以使用多步展示(rollout)和蒙特卡洛回报估计来估计并分配给训练片段(episode)期间的每个步。作为一个示例,如果末端执行器d到目标对象的距离小于预定义的阈值τ,则稀疏奖励为$r=1$,否则为0。对策略进行训练,以使从策略采样的轨迹$T=o_1,a_1,\cdots,o_T$的期望贬损奖励最大化:

[0042] $$\theta^* = \arg\max_{\theta} \mathbb{E}_{\tau\sim\pi_\theta}\left(\sum_{t=1}^{T}\gamma^t R(a_t, O_t)\right)$$

[0043] 可以使用循环神经网络模型的相应卷积神经网络(CNN)将场景图像和查询图像每个嵌入到相应的嵌入矢量中。例如,可以使用CNN层(诸如VGG16网络的Conv5-3层)的输出来可选地将场景图像调整大小(例如,将其缩放为256x256)并嵌入。另外,例如,可以以类似的方式使用VGG16可选地将查询图像调整大小(例如,将其缩放为32×32)并嵌入。

[0044] 视觉伺服模型是在观察和动作序列$s_{1:t}=(o_1,a_2,\cdots,o_t,a_t)$上的循环神经网络模型。循环神经网络模型包括一个或多个循环层,诸如维度为512的单层LSTM。可以使用带有ReLU的单个层的全连接网络部分将动作嵌入到动作矢量(例如64维矢量)中,并且可以在每个时间步将动作嵌入与观察嵌入拼接。每个时间步的观察嵌入可以基于该时间步的查询图像嵌入和场景图像嵌入的连结(concatenate)。请注意,在片段期间,查询图像的嵌入可以在每个时间步保持相同(查询图像保持不变),而场景图像嵌入可以在每个时间步变化(因为每个场景图像嵌入基于新的“当前”场景图像)。

[0045] 循环神经网络的循环层(例如,LSTM层)中的隐藏状态捕获片段中观察的全部历史。因此,它可以基于来自多个先前动作预测的实现的观察到的反馈来执行隐式校准。

[0046] 在每个时间步的损失可以基于执行动作之后末端执行器位置与目标对象位置之间的欧几里得距离。由x_t表示在世界坐标系(world frame)中步t处末端执行器的位置,其可以表示为$x_t=x_{t-1}+a_{t-1}$,,其中a_{t-1}由控制器生成。如果目标对象位置在同一世界坐标系中为l则损失为:

[0047] $$\sum_{t=1}^{T-1} \| x_{t-1} + a_{t-1} - l \|^2$$

[0048] 为了将动作预测幅度保持在界限内,可以学习归一化的动作方向矢量并且可以利用恒定速度。即,使用循环神经网络模型生成的动作预测可以是归一化的动作矢量,并且可以指示速度方向,其中速度是恒定速度(即方向将变化,但是速度的幅度将恒定)。

[0049] 本文公开的实现方式所解决的视图不变查询目标到达任务将超维引入到状态空间,并且由于样本复杂性,经由深度强化学习为这种复杂任务学习最优策略可能具有挑战性。因此,本文公开的一些实现方式通过使用演示来加速策略学习。例如,可以在演示的多个时间步中的每个确定朝向模拟中的目的的最佳方向,并且可以可选地干扰演示。例如,可以通过向演示注入正态高斯噪声来干扰一个或多个演示,以学习强健的策略。

[0050] 假设片段的长度为T,则可以基于多步蒙特卡洛(MC)策略评估为每个时间步收集无策略(off-policy)数据并确定奖励。蒙特卡洛返回估计提供了Bellman更新的简化版本,并具有没有Bellman更新不稳定性的好处。使用这些MC返回估计,在给定当前状态$v_t=\sum_i^H \gamma^i r_{t+1}$的情况下,可以利用循环神经网络模型为任何候选动作产生奖励值。

[0051] 在运行时,可以对利用循环神经网络模型产生的预测动作a_t进行小的优化。这样的优化可以提供更好的结果并在运行时改善性能。可以使用各种优化,诸如交叉熵方法(CEM)优化,它是一种无导数优化方法。CEM通过拟合以预测动作矢量a_t为中心的高斯分布对一批候选动作进行采样,并根据值网络对它们进行评估。具有最高值的候选动作将被选

择作为下一个要执行的动作。

[0052] 在实际的机器人上使用神经网络模型之前，至少可以对模型的视觉部分进行适应，可选地使模型的策略/控制部分固定。模型的视觉部分应了解与任务相关的场景的相关语义，同时允许伺服。为了确保两个性质真实，可以基于训练示例对模型的视觉部分进行精细调整，该训练示例基于相关(但截然不同)任务，诸如粗略的对象定位。例如，可以在视觉部分的最后特征图上使用软注意力(soft attention)。软注意力可以是最后特征图中所有位置上的softmax，这对应于一小组潜在的目标对象位置。训练示例可以每个包括带有目标对象的真实查询图像和包括目标对象(以及其他对象和/或机器人末端执行器)的真实场景图像的训练示例输入、以及作为目标对象(在真实场景图像中)的真实位置的手动标记的训练示例输出。损失可以基于真实位置和注意力矢量之间的交叉熵来确定，该真实位置在训练示例输出中表示为一个热矢量，该注意力矢量是对所有位置的分数的softmax运算。因此，根据本文描述的实现方式的神经网络模型的网络体系结构提供了经由辅助注意力损失来使感知与控制脱离的灵活性。这种灵活性能够实现在精细调整中适应视觉层。

[0053] 为了在模拟中训练循环神经网络模型，可以利用模拟机器人(例如，多自由度机器人臂)和模拟环境来使用模拟器(例如，BULLET物理引擎)。在模拟设置中，可以将随机模拟对象随机放置在模拟臂前面的一个或多个表面(例如桌子)上。为了激励模型学习对目标对象的形状和外观以及场景外观不变的强健的策略，可以利用多种对象集，并使用纹理随机化、光照随机化和/或其他技术以指数方式增强环境的视觉多样性。在如此多样化的模拟环境中进行训练导致学习可快速适应新测试场景的通用策略。

[0054] 现在转到图，图1示出可以训练循环神经网络模型以用于机器人的末端执行器的视点不变的视觉伺服的示例环境。

[0055] 图1包括模拟器训练系统120，其由一个或多个计算机系统实现。在生成用于训练循环神经网络模型170的模拟数据时，模拟器训练系统120与一个或多个模拟器110接口。模拟器110也由一个或多个计算机系统来实现，其可以与用于实现模拟器训练系统120的计算机系统相同和/或不同。模拟器110可用于模拟包括相应环境对象的各种环境，模拟在该环境中操作的机器人，模拟响应于各种模拟的机器人动作的虚拟实现的该机器的响应，并模拟响应于模拟的机器人动作在机器人与环境对象之间的交互。可以利用各种模拟器，诸如模拟碰撞检测、柔体和刚体动力学的物理引擎等。这种模拟器的一个非限制示例是BULLET物理引擎。

[0056] 模拟器训练系统120包括场景配置引擎121、渲染查询图像引擎122、渲染场景图像引擎123、动作引擎124和奖励信号引擎125。模拟器训练系统120导致使用模拟器110执行大量(例如，数千、数十万，数百万)模拟片段，并且在执行此类片段时与循环神经网络模型170交互。每个模拟片段可以在存在相应模拟环境对象的相应模拟环境中执行。场景配置引擎121改变片段中的模拟环境和/或模拟环境对象，并为片段选择变化的目标对象。例如，第一组的一个或多个模拟片段可以以5个模拟盘子、3个模拟叉子、4个模拟杯子、以及所有搁置在模拟桌子上面的模拟餐巾纸出现。一个或多个对象的起始姿势可以可选地在第一组的一个或多个片段之间变化，目标对象可以可选地在第一组的一个或多个片段之间变化，桌子的纹理可以可选地在第一组的一个或多个片段之间变化，和/或模拟的照明可以可选地在第一组的一个或多个片段之间变化。第二组一个或多个模拟片段可以以不同的模拟表面上

的8个模拟叉子和2个模拟杯子出现。第二组的片段之间的变化同样可以发生。

[0057] 对于每个模拟片段，选择模拟环境中的模拟目标对象，并且由渲染查询图像引擎122来渲染目标对象的渲染查询图像。例如，图1示出可以对片段渲染的对象的查询图像101。查询图像101可以高度聚焦于目标对象上，并且可以可选地包括与片段的场景相符的背景(如果有的话)。可以从与对该片段的生成的渲染场景图像中利用的视点相同或不同的视点渲染查询图像101。

[0058] 每个模拟片段由T个单独的时间步或实例组成。渲染场景图像引擎123对每个时间步渲染场景图像，其中每个渲染场景图像来自相应的视点，并在相应的时间步捕获模拟环境。例如，每个渲染场景图像可以在相应的时间步捕获模拟的末端执行器和/或其他模拟的机器人组件、模拟的目标对象以及可选的其他模拟的环境对象。如本文所述，用于渲染场景图像的视点可以在各片段之间广泛变化，以提供合成训练数据中的多样性以及循环神经网络模型170对各种视点的强健性。

[0059] 图1示出可以在整个片段中渲染场景图像I_0,I_1,…I_T，其中每个场景图像在相应的时间步渲染并在相应的时间步使用循环神经网络模型170进行处理。如本文所述，作为利用循环神经网络模型170生成的动作预测的实现的结果，场景图像I_0,I_1,…I_T将随时间步而变化。给定片段的场景图像I_0,I_1,…I_T可以可选地从相同的视点渲染(尽管不同的视点可以用于不同的片段)。

[0060] 动作引擎124在片段的每个时间步实现使用循环神经网络模型170对该时间步生成的相应预测动作。例如，动作引擎124使模拟器110根据在每个时间步生成的预测动作遍历模拟机器人的模拟末端执行器，从而在模拟环境中对模拟末端执行器进行伺服。

[0061] 奖励信号引擎125向奖励引擎132提供一个或多个奖励信号，以供奖励引擎132用于确定训练期间用于更新循环神经网络模型170的奖励。例如，奖励引擎132可以在每个时间步确定奖励，并基于奖励在每个时间步更新循环神经网络模型170。例如，每个更新可以是损失，其基于奖励并且跨循环神经网络模型170的一个或多个(例如，全部)部分反向传播，以更新那些部分的参数。由奖励信号引擎125提供的奖励信号可以包括例如基准事实方向103A和/或成功/失败指示103B——可以在每个时间步更新二者之一或两者。例如，奖励引擎132可以使用基准事实方向103A，以基于该时间步的动作预测所指示的方向与为该时间步提供的基准事实方向103A的比较(例如，之间的欧几里得距离)来确定该时间步的形态奖励。每个基准事实方向指示对该时间步到目标对象的对应方向，并且可以在每个时间步基于该时间步处的模拟末端执行器的姿态以及基于该时间步的模拟目标物的姿态而确定。

[0062] 在片段期间的每个时间步，使用循环神经网络模型170处理查询图像、对应的场景图像和对应的先前动作，以生成预测动作。预测动作被提供给动作引擎124，以用于在模拟器110中实现预测动作。此外，奖励由奖励引擎132确定，并被用于更新循环神经网络模型170。例如，如图1所示，在给定片段的每个时间步，可以在模型170的视觉层171上处理查询图像101和场景图像I_0,I_1,…I_T中的相应一个，以生成视觉输出(即，嵌入)。此外，可以在模型的动作层172上处理先前动作$102_{0\text{-}T}$(即，最近生成的动作预测)的对应一个，以生成动作输出(即，嵌入)。视觉输出和动作输出可以在包括循环层184的模型170的策略层180上被连结和处理，以生成预测动作$104_{0\text{-}T}$中的相应一个。预测动作$104_{0\text{-}T}$中的相应一个被提供给动作引擎124以在模拟器110中实现，并且被提供给奖励引擎132。奖励引擎132利用预测动作

$104_{0\text{-}T}$中的相应一个，以及对应的基准事实方向103A和/或成功/失败指示1033来确定奖励，并为循环神经网络模型170提供相应的更新$105_{0\text{-}T}$。可以对该片段的每个时间步重复此操作，并对大量片段中的每一个进一步重复以训练循环神经网络模型170。

[0063] 同样如图1所示，是视觉适应引擎134，其可以用于在将循环神经网络模型170部署在一个或多个物理机器人中之前，基于真实的训练示例135来进一步训练至少(并且可选地仅)视觉层171。在图1中示出了训练示例135之一的一个示例，并且该示例包括训练示例输入135A1和训练示例输出135A2，训练示例输入135A1包括真实查询图像和真实场景图像，训练示例输出135A2包括训练示例输入135A1的真实场景图像中的(训练示例输入135A1的真实查询图像的)目标对象的位置指示。例如，训练示例输出135A2可以是一个热矢量，其中“热”值指示在存在训练示例输入135A1的相应真实查询图像的对象的情况下训练示例输入135A1的真实场景图像中的位置。视觉适应引擎134可以使用视觉层171以及可选地在真实世界适应中使用(并且不用于视觉伺服)的一个或多个附加层(例如，仿射层)来处理训练示例输入135A1，基于所生成的预测输出和训练示例输出135A2的比较来确定误差，以及确定用于更新视觉层171的误差(例如，通过反向传播)。以这种方式，可以基于真实的训练示例来进一步训练视觉部分，以使得当被真实的物理机器人用于视觉伺服时循环神经网络模型更加强健和/或更准确。

[0064] 图2示出在执行对机器人180的末端执行器182的视觉伺服中的示例真实物理机器人180以及使用循环神经网络模型170的示例。视觉伺服系统140可以用于执行视觉伺服并且例如可以由机器人180的一个或多个处理器实现。机器人180可以与由图1中的机器人模拟器110模拟的模拟机器人相同和/或相似。机器人180是具有多个自由度(例如，在每个致动器处一个)的“机器人臂”，以使得能够沿着多个潜在路径中的任何一个来来回移动抓握末端执行器182，以将抓握末端执行器182定位在期望的位置。机器人180进一步控制抓握末端执行器182的两个相对的“爪”，以至少在打开位置和关闭位置(和/或可选的多个“部分关闭”位置)之间致动爪。

[0065] 图2中还示出示例视觉组件184A，并且来自视觉组件184A的图像被应用于视觉伺服中的循环神经网络模型170。在图2中，在第一固定视点处提供视觉组件184A。视觉组件184B也在图2中以虚线示出，并且相对于视觉组件184A处于非常不同的视点。提供视觉组件184B以说明由于模型170的视点不变特性，即使使用视觉组件184B代替视觉组件184A，视觉伺服仍可以有效执行。视觉组件184A生成与在传感器的视线内的对象的形状、颜色、深度和/或其他特征有关的图像。视觉组件184A可以是例如单镜头相机(例如，生成2D RGB图像)、立体相机(例如，生成2.5D RGB图像)和/或激光扫描仪(例如，生成2.5D“点云”图像)。要理解，模拟数据的渲染图像(图1)将被渲染为具有与视觉组件184A生成的图像相同的类型。例如，两者都可以是2.5D RGBD图像。还要理解，视觉组件替代地可以直接耦合到机器人180。例如，视觉组件可以耦合到机器人180的链路，诸如在末端执行器182上游的链路。此外，要理解，视觉组件184A、视觉组件184B和/或直接耦合到机器人180的视觉组件可以可选地被独立地调节(例如，它可以被独立地平移、倾斜和/或缩放)。

[0066] 视觉组件184A具有机器人180的工作空间的至少一部分(诸如包括示例对象191的工作空间的一部分)的视场。尽管在图2中未示出用于对象191的搁置表面，但是那些对象可以搁置在桌子、托盘和/或其他表面上。对象191包括刮铲、订书机和铅笔，并且可以可选地

与训练循环神经网络模型170中使用的模拟对象不同。尽管图2中示出了特定机器人180，但是可以利用附加的和/或替代的机器人(物理的和/或模拟的)，包括与机器人180相似的附加的机器人手臂、具有其他机器人手臂形式的机器人、具有人形的机器人、具有动物形式的机器人、经由一个或多个轮子移动的机器人、无人机(“UAV”)等。此外，尽管在图2中示出了特定的抓握末端执行器182，但是可以使用附加的和/或替代的末端执行器(物理的和/或模拟的)，诸如替代的冲击式抓握末端执行器(例如，带有抓“板”的末端执行器，带有更多或更少的“手指”/“爪”的末端执行器)、“侵入式”抓握末端执行器、“收缩式”抓握末端执行器或“接触式”抓握末端执行器、或非抓握末端执行器。

[0067] 在图2中，查询图像引擎142确定目标对象的查询图像201。可以例如基于来自用户的用户接口输入来确定查询图像201(例如，查询图像可以是紧凑裁剪的图像，其基于在显示由视觉组件184A捕获的图像的计算设备处经由用户的用户接口输入由用户绘制的或以其他方式指示的边界框裁剪)，或基于更高级别任务规划器的输出。查询图像201可以可选地基于相机184A捕获的图像，或者可以基于来自单独相机的图像，甚至是“库存”图像。作为一个示例，基于指定目标对象的一个或多个语义特性的用户的说出或键入的用户接口输入，通过确定与那些一个或多个语义特性匹配的库存图像，查询图像201可以是库存图像。例如，响应于“抓住订书机”的说出的用户接口输入，可以基于被术语“订书机”索引或以其他方式与术语“订书机”相关联的库存图像来选择库存图像作为查询图像201。作为另一示例，基于更高级任务规划器指定目标对象的一个或多个语义特性，通过确定与那些一个或多个语义特性匹配的库存图像，查询图像201可以是库存图像。在每个时间步，场景图像引擎144为该时间步提供场景图像I_0，I_1，…I_T中的相应当前图像，其中每个当前场景图像由相机184A在相应的时间步捕获。

[0068] 在每个时间步，视觉伺服系统140在循环神经网络模型上，与下一动作$202_{0\text{-}T}$中的对应在前动作(如果有的话)一起，处理查询图像201和场景图像I_0，I_1，…I_T中的对应一个，以生成下一动作$202_{0\text{-}T}$的对应一个。下一动作$202_{0\text{-}T}$的对应一个提供给动作引擎148，其生成并向机器人180的一个或多个致动器提供控制命令，以使末端执行器182根据该动作移动。迭代执行此操作以在每个时间步(基于查询图像、先前的下一个动作和当前场景图像)生成新的下一动作，从而在多个时间步上将末端执行器182朝向查询图像201指示的目标对象伺服。一旦达到目标对象，就可以可选地抓握或以其他方式操纵目标对象。

[0069] 尽管未在图2中示出，但是在许多实现方式中，奖励引擎可以在伺服片段期间被利用，并且可以被用来继续生成奖励以用于在伺服片段期间对模型的进一步训练(例如，基于本文所述的基于蒙特卡洛的技术)。

[0070] 图3示出图1和2的循环神经网络模型170的一个示例。在图3中，查询图像101和场景图像I_t在视觉层171上被处理。具体地，查询图像101在视觉层的第一CNN 172上被处理，并且场景图像I_t在视觉层171的第二CNN173上被处理。在卷积层174上处理基于CNN 172和173的处理生成的输出，并在池化层175上处理来自卷积层174上的处理的输出。

[0071] 来自池化层175上的处理的输出被应用于策略层180。来自池化层175上的处理的输出与来自动作层176的输出一起被应用于策略层180。基于在动作层176的完全连接层177和平铺层178上的先前动作102_t的处理(例如，来自紧挨的在先时间步的预测动作)生成来自动作层的输出。

[0072] 在策略层180的卷积层181、最大池化层182、完全连接的层183和循环层184上对来自池化层175上的处理的输出和来自动作层176的输出进行处理。在策略层180上生成的输出在完全连接层185上进行处理以生成动作预测104_t。动作预测104_t也使用策略层180的完全连接层186进行处理，并且来自该处理的输出连同来自循环层184的输出一起在另一个完全连接层187上进行处理。

[0073] 如132A所示，基于动作预测104_t和基准事实方向(例如，参见图1)来确定形态奖励。例如，形态奖励可以基于动作预测104_t指示的方向和基准事实方向之间的距离。应用形态奖励(例如，作为反向传播的损失)以更新循环神经网络170。

[0074] 如132B所示，还基于基于完全连接的层187上的处理生成的输出来确定稀疏奖励。稀疏奖励可以基于如本文所述的MC返回估计来生成，并且也可以被应用(例如，作为反向传播损失)于更新循环神经网络170。132A和132B指示的奖励可以由奖励引擎132(图1)应用。

[0075] 在图3中也示出了卷积层174A，其未在视觉伺服中使用，但是可用于基于本文所述的真实训练示例(例如，在基于模拟数据的训练之后)来适应卷积层174以及CNN 172和173。这在图3中示出为基于真实图像的损失134A，其可以由视觉适应引擎134(图1)基于在CNN 172和173以及卷积层174、174A上对真实训练示例输入的处理以生成预测的输出、以及预测输出与真实训练示例输出的比较来确定。

[0076] 图4是示出训练用于在机器人的末端执行器的视点不变视觉伺服中使用的循环神经网络模型的示例方法400的流程图。为了方便起见，参考执行操作的系统来描述方法400的操作。该系统可以包括计算系统的一个或多个组件。尽管以特定顺序示出了流程图的操作，但这并不意味着限制。一个或多个操作可重新排序、省略或添加。

[0077] 在452，模拟片段开始。

[0078] 在步骤454，系统在模拟器中为模拟片段配置模拟场景。

[0079] 在步骤456，系统渲染用于模拟片段的查询图像。查询图像属于模拟场景中模拟目标对象。

[0080] 在步骤458，系统基于模拟场景的当前状态以及由模拟器模拟的模拟机器人，对模拟片段的当前时间步渲染当前场景图像。当前场景图像至少捕获模拟机器人的末端执行器和模拟目标对象。

[0081] 在步骤460，系统在循环神经网络模型上处理查询图像、当前场景图像和先前动作(如果有)。

[0082] 在步骤462，系统基于步骤460的处理生成动作预测。

[0083] 在步骤464，系统确定奖励并基于该奖励更新循环神经网络模型。

[0084] 在步骤466，系统在模拟器中实现模拟动作预测。

[0085] 在步骤468，系统确定是否已经到达片段的终点。这可以基于执行的实例的阈值数量、阈值时间量的超过和/或确定模拟末端执行器已经到达模拟目标对象(例如，基于来自模拟器的反馈)。

[0086] 如果在步骤468的迭代，系统确定尚未达到片段的终点，则系统返回至458并渲染另一个当前场景图像(它将反映在步骤466的在先迭代处的动作预测的实现)，然后执行方框460、462、464、466和468的另一迭代。

[0087] 如果在步骤468的迭代，系统确定已经达到片段的终点，则系统进行到步骤470。

[0088] 在步骤470，系统确定是否执行另一片段。这可以基于执行的片段的阈值数量，阈值时间量的超过和/或以其他方式确定循环神经网络模型已得到充分训练。

[0089] 如果在步骤470的迭代，系统确定执行另一片段，则系统进行到452，并开始另一模拟片段。

[0090] 如果在步骤470的迭代，系统确定不执行另一片段，则系统进行到步骤472。

[0091] 在步骤472，系统基于真实图像训练示例来适应循环神经网络的视觉层。然后，系统进行到方框474，并提供适应的循环神经网络以供在视觉伺服中由一个或多个机器人使用。

[0092] 图5是示出在执行机器人的末端执行器的视觉伺服时使用循环神经网络模型的示例方法500的流程图。为了方便起见，参考执行操作的系统来描述方法500的操作。该系统可以包括机器人的一个或多个处理器。尽管以特定顺序示出了流程图的操作，但这并不意味着限制。一个或多个操作可重新排序、省略或添加。

[0093] 在552，机器人末端执行器的视觉伺服开始。

[0094] 在步骤554，系统确定查询图像。查询图像属于目标对象。查询图像用于指示或识别目标对象。确定查询图像可以包括检索目标对象的图像(例如，通过根据图像的主体(corpus)中选择图像，通过裁剪由与机器人相关联的视觉组件(例如，相机)捕获的图像以产生目标对象的图像，或通过其他任何合适的技术)。

[0095] 在步骤556，系统使用与机器人相关联的相机捕获当前场景图像。当前场景图像至少捕获机器人的末端执行器和目标对象。

[0096] 在步骤558，系统在循环神经网络模型上处理查询图像、当前场景图像和先前动作(如果有)。

[0097] 在步骤560，系统基于步骤558的处理生成动作预测。

[0098] 在可选步骤562，系统确定奖励并基于该奖励更新循环神经网络模型。

[0099] 在步骤564，系统通过基于动作预测来控制机器人的末端执行器来实现动作预测。例如，该系统可以向机器人的一个或多个致动器提供控制命令，该控制命令控制末端执行器的位置，以使末端执行器根据动作预测移动。

[0100] 在步骤566，系统确定是否已经到达片段的终点。这可以基于执行的实例的阈值数量、阈值时间量的超过和/或确定末端执行器已经到达目标对象(例如，基于步骤560的最新迭代处的动作预测，该动作预测指示需要很少或无需进一步的末端执行器的移动来到达目标对象)。

[0101] 如果在步骤566的迭代，系统确定尚未达到片段的终点，则系统返回至556，并捕获另一个当前场景图像26(这将反映在步骤564的在先迭代处的动作预测的实现)，然后执行方框558、560、可选地562、564和566的另一次迭代。

[0102] 如果在步骤568的迭代，系统确定已经达到片段的终点，则系统前进至方框568，并等待新目标对象的新查询图像。当接收到对新目标对象的新查询图像时，系统继续回到552，然后此时再次基于新查询图像执行视觉伺服。

[0103] 图6示意性地描绘了机器人625的示例架构。机器人625包括机器人控制系统660、一个或多个操作组件625a-625n以及一个或多个传感器642a-642m。传感器642a-642m可以包括例如视觉组件、光传感器、压力传感器、压力波传感器(例如，麦克风)、接近传感器、加

速计、陀螺仪、温度计、气压计等。虽然传感器642a-m被描绘为与机器人625集成在一起，但这并不意味着是限制性的。在一些实现方式中，传感器642a-m可以例如作为独立单元位于机器人625的外部。

[0104] 操作组件625a-625n可包括例如一个或多个末端执行器和/或一个或多个伺服电机或其他致动器，以实现机器人的一个或多个组件的运动。例如，机器人625可以具有多个自由度，并且每个致动器可以响应于控制命令而在一个或多个自由度内控制机器人625的致动。如本文中所使用的，除了可能与致动器相关联以及将接收的控制命令转换成用于驱动致动器的一个或多个信号的任何驱动器之外，术语致动器包括产生运动的机械或电设备(例如，电机)。因此，向致动器提供控制命令可以包括向驱动器提供控制命令，该驱动器将该控制命令转换为用于驱动电或机械设备以产生期望运动的适当信号。

[0105] 机器人控制系统660可以在机器人625的一个或多个处理器(诸如CPU、GPU和/或其他控制器)中实现。在一些使得方式中，机器人625可以包括“大脑箱(brain box)”，该“大脑箱”可以包括控制系统660的所有或各方面。例如，大脑箱可以向操作组件625a-n提供数据的实时脉冲串，其中每个实时脉冲串包括一组的一个或多个控制命令，这些命令尤其指示对于一个或多个操作组件625a-n的每个的运动参数(如果有)。在一些实现方式中，机器人控制系统660可以执行本文所述的一个或多个方法的一个或多个方面。

[0106] 如本文中所描述，在一些实现方式中，在伺服末端执行器时由控制系统660生成的控制命令的全部或各方面可以基于利用本文中所描述的循环神经网络模型生成的预测动作。尽管控制系统660在图6中示出为机器人625的整体构成部分，但是在一些实现方式中，控制系统660的所有或各方面可以在与机器人625分离但与其通信的组件中实现。例如，可以在与机器人625进行有线和/或无线通信的一个或多个计算设备(诸如计算设备710)上实现控制系统660的所有或各方面。

[0107] 图7是可以可选地用于执行本文描述的技术的一个或多个方面的示例计算设备710的框图。计算设备710通常包括至少一个处理器714，该处理器714经由总线子系统712与多个外围设备进行通信。这些外围设备可以包括存储子系统724(包括例如存储器子系统725和文件存储子系统726)、用户接口输出设备720、用户接口输入设备722和网络接口子系统716。输入和输出设备允许用户与计算设备710交互。网络接口子系统716提供到外部网络的接口，并耦合到其他计算设备中的相应接口设备。

[0108] 用户接口输入设备722可以包括键盘，诸如鼠标、轨迹球、触摸板或图形输入板的指向设备，扫描仪，结合到显示器的触摸屏，诸如语音识别系统、麦克风的音频输入设备，和/或其他类型的输入设备。通常，术语“输入设备”的使用旨在包括将信息输入到计算设备710或通信网络中的所有可能的设备以及方式的类型。

[0109] 用户接口输出设备720可以包括显示子系统、打印机、传真机或诸如音频输出设备的非视觉表现器。显示子系统可以包括阴极射线管(CRT)、平板设备(诸如液晶显示器(LCD))、投影设备或用于创建可见图像的一些其他机制。显示子系统也可以提供非视觉表现，诸如经由音频输出设备。通常，术语“输出设备”的使用旨在包括将信息从计算设备710输出到用户或另一机器或计算设备的所有可能的设备以及方式的类型。

[0110] 存储子系统724存储提供本文所述的一些或所有模块的功能的程序和数据构造。例如，存储子系统724可以包括执行本文描述的一种或多种方法的所选择的方面的逻辑。

[0111] 这些软件模块通常由处理器714单独执行或与其他处理器组合执行。存储子系统724中使用的存储器725可以包括多个存储器，包括用于在程序执行期间存储指令和数据的主随机存取存储器(RAM)730以及存储固定指令的只读存储器(ROM)732。文件存储子系统726可以为程序和数据文件提供持久存储，并且可以包括硬盘驱动器、软盘驱动器以及相关联的可移除介质、CD-ROM驱动器、光盘驱动器或可移除介质盒。实现某些使得方式的功能的模块可以由存储子系统724中的文件存储子系统726存储或存储在处理器714可访问的其他机器中。

[0112] 总线子系统712提供了一种机制，用于使计算设备710的各个组件和子系统按预期的方式相互通信。尽管总线子系统712被示意性地示出为单个总线，但是总线子系统的替代实现方式可以使用多个总线。

[0113] 计算设备710可以具有各种类型，包括工作站、服务器、计算集群、刀片服务器、服务器场或任何其他数据处理系统或计算设备。由于计算机和网络的日新月异的特性，图7中所描绘的计算设备710的说明仅旨在作为特定示例用于说明一些实现方式。具有比图7中描绘的计算设备更多或更少的组件的计算设备710的许多其他配置是可能的。

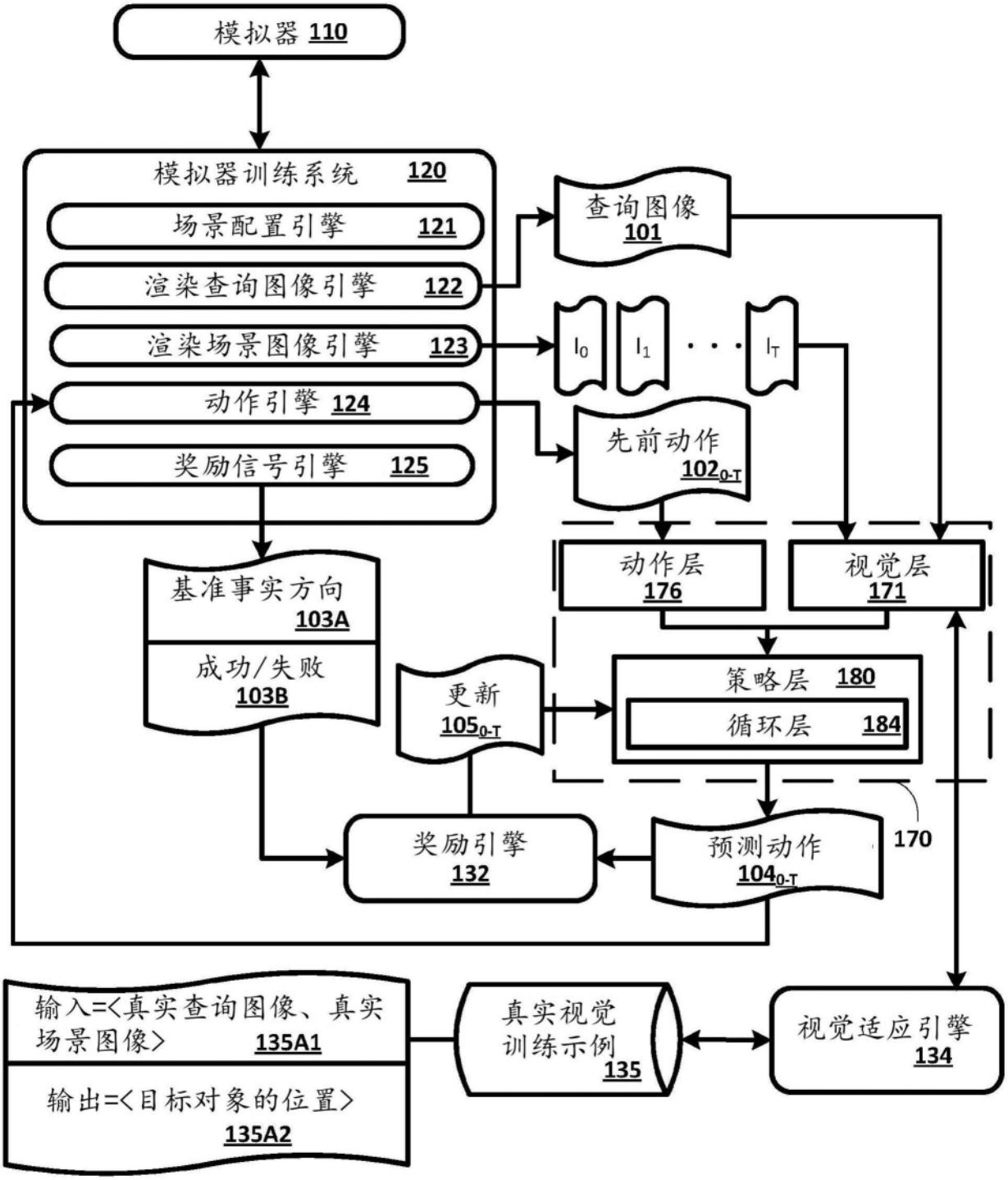

模拟器 110
模拟器训练系统 120
场景配置引擎 121
渲染查询图像引擎 122
渲染场景图像引擎 123
动作引擎 124
奖励信号引擎 125
查询图像 101
I_0 I_1 . . . I_T
先前动作 102_{0-T}
基准事实方向 103A
成功/失败 103B
更新 105_{0-T}
动作层 176
视觉层 171
策略层 180
循环层 184
奖励引擎 132
预测动作 104_{0-T}
170
输入=<真实查询图像、真实场景图像> 135A1
输出=<目标对象的位置> 135A2
真实视觉训练示例 135
视觉适应引擎 134

图1

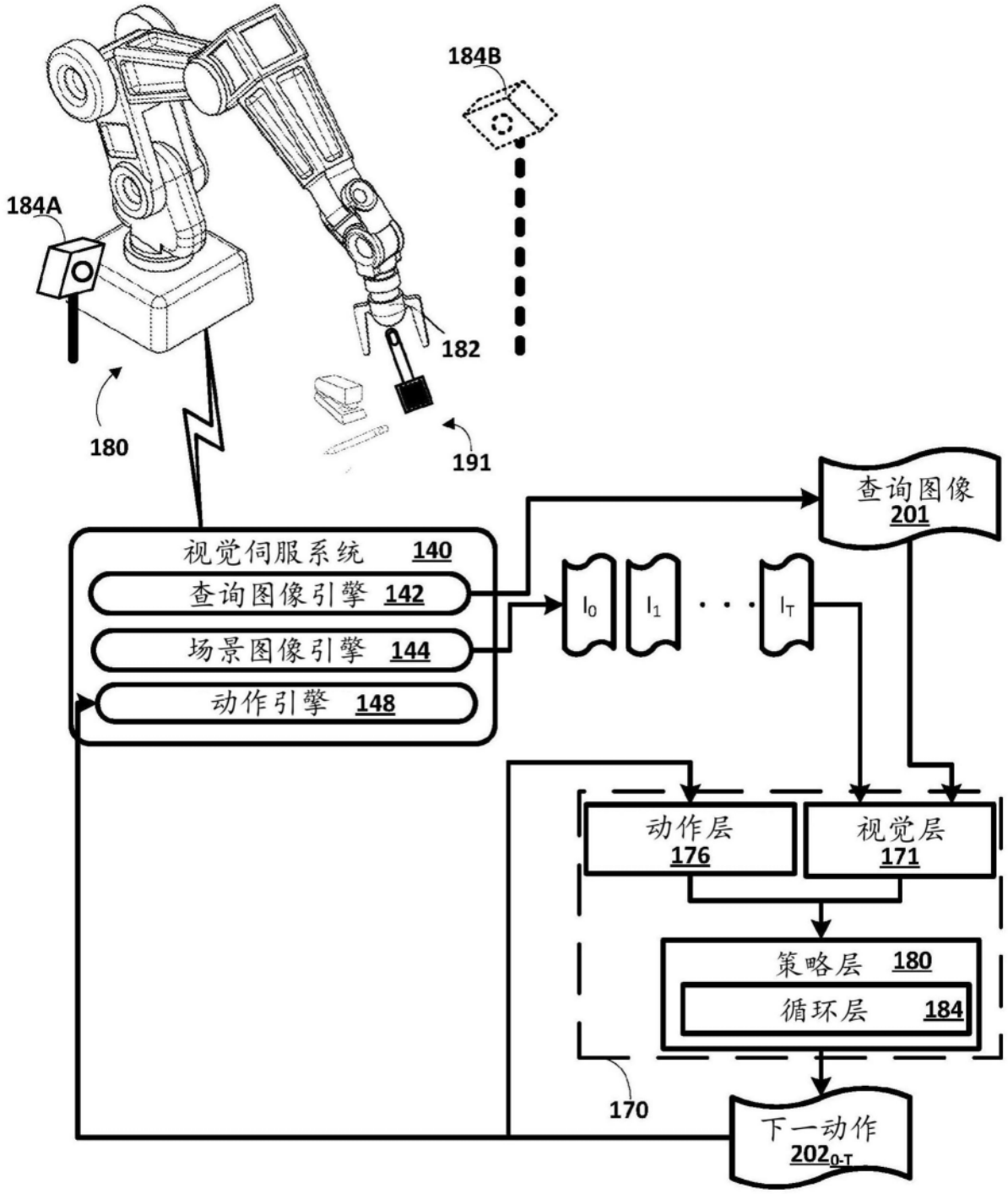

184B
184A
182
180
191
查询图像
201
视觉伺服系统 140
查询图像引擎 142
场景图像引擎 144
动作引擎 148
I_0
I_1
I_T
动作层
176
视觉层
171
策略层 180
循环层 184
170
下一动作
202_{0-T}

图2

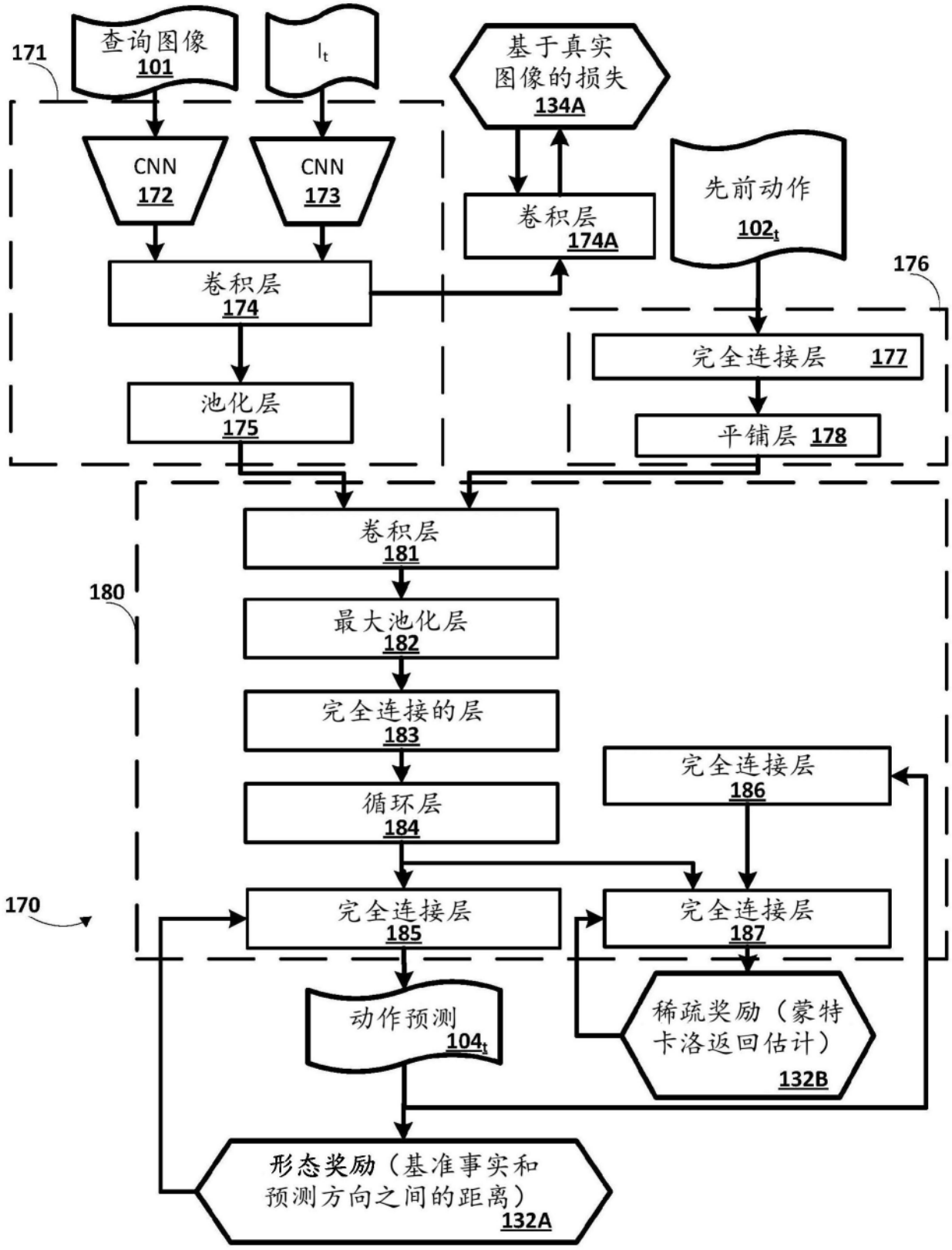

171
查询图像
101
I_t
基于真实图像的损失
134A
CNN
172
CNN
173
卷积层
174A
先前动作
102_t
176
卷积层
174
完全连接层 177
池化层
175
平铺层 178
卷积层
181
180
最大池化层
182
完全连接的层
183
完全连接层
186
循环层
184
170
完全连接层
185
完全连接层
187
动作预测
104_t
稀疏奖励（蒙特卡洛返回估计）
132B
形态奖励（基准事实和预测方向之间的距离）
132A

图3

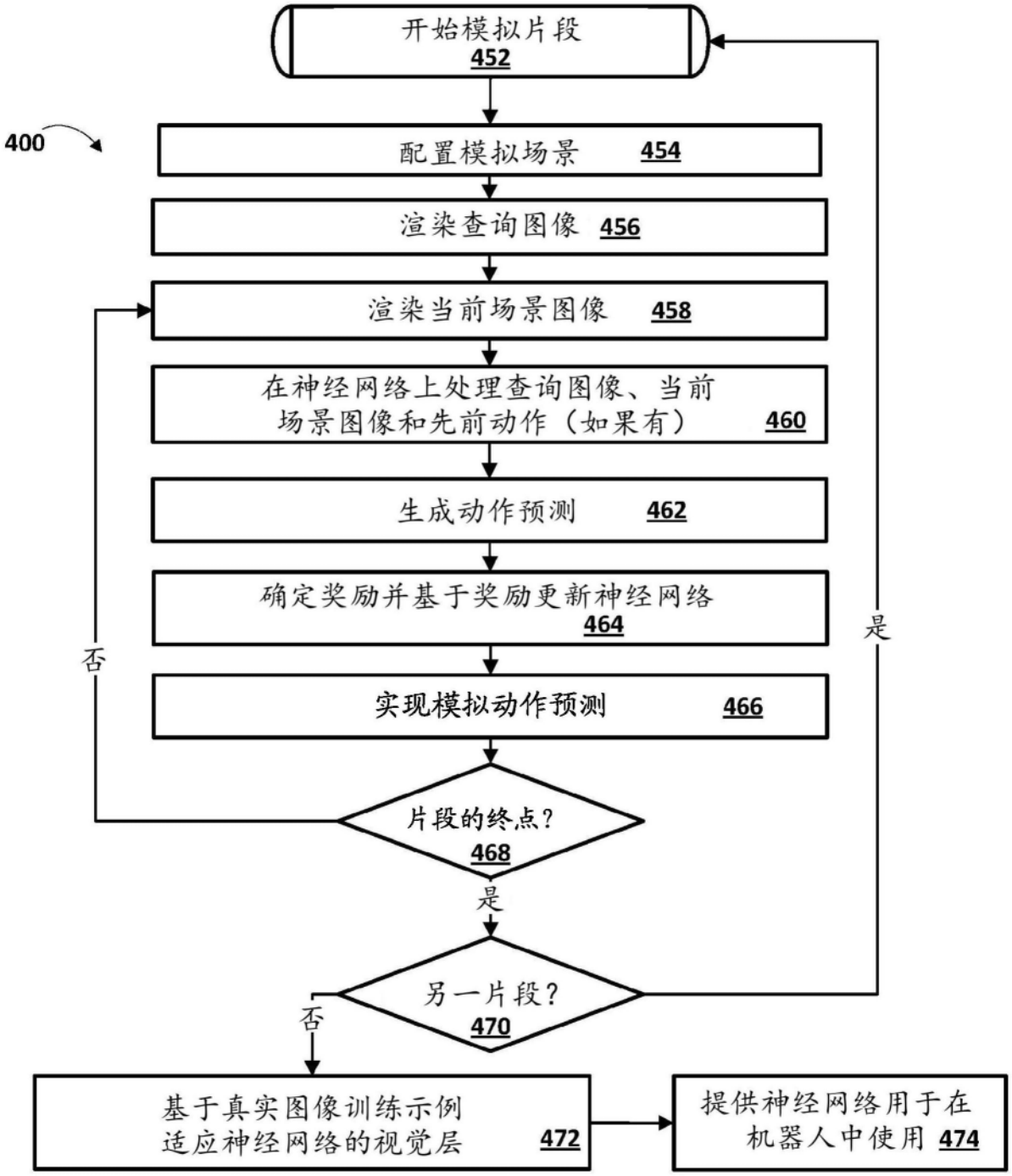

开始模拟片段
452
400
配置模拟场景 454
渲染查询图像 456
渲染当前场景图像 458
在神经网络上处理查询图像、当前场景图像和先前动作（如果有） 460
生成动作预测 462
确定奖励并基于奖励更新神经网络 464
实现模拟动作预测 466
片段的终点？ 468
否
是
是
另一片段？ 470
否
基于真实图像训练示例适应神经网络的视觉层 472
提供神经网络用于在机器人中使用 474

图4

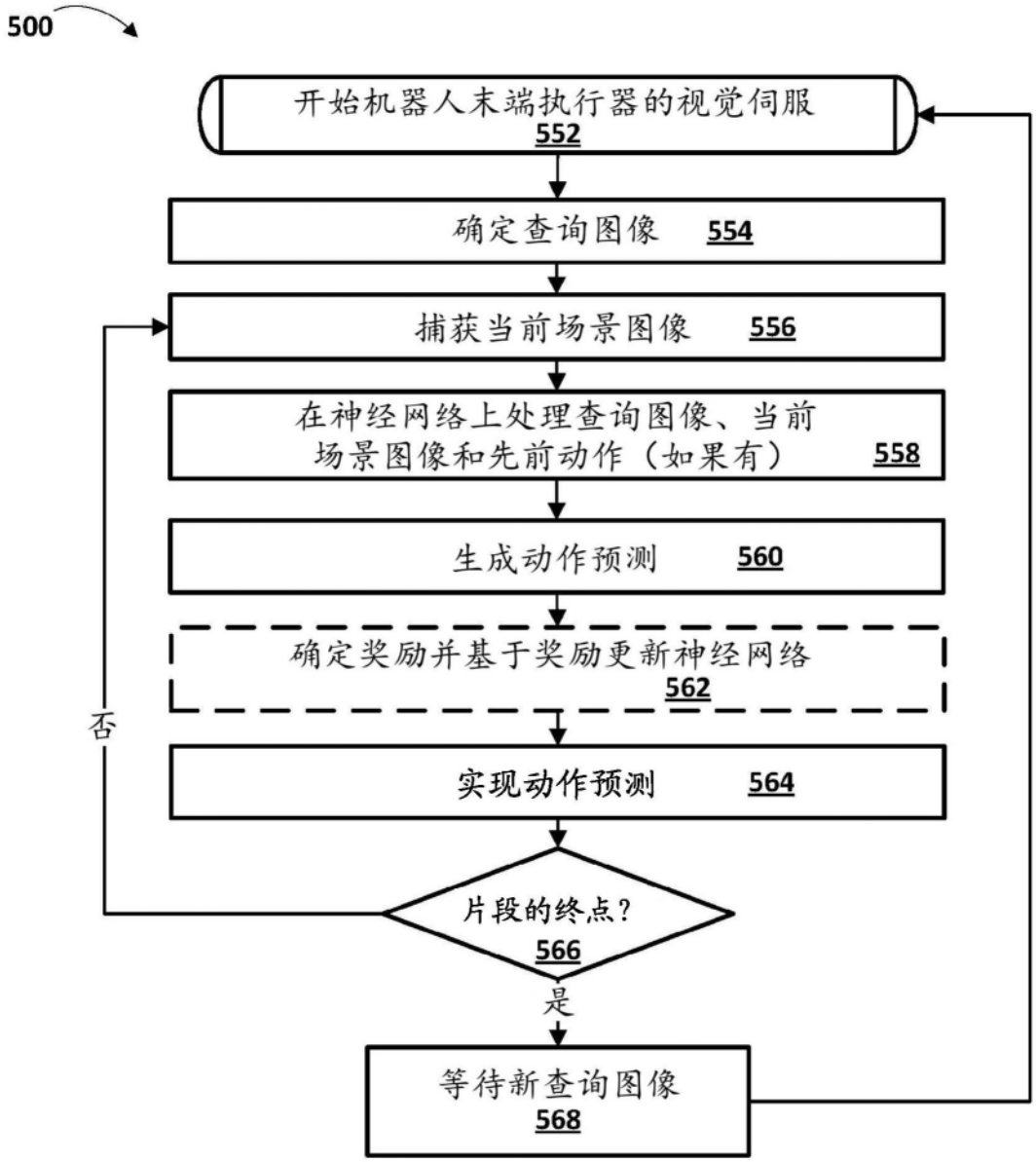

500
开始机器人末端执行器的视觉伺服
552
确定查询图像 554
捕获当前场景图像 556
在神经网络上处理查询图像、当前
场景图像和先前动作（如果有） 558
生成动作预测 560
确定奖励并基于奖励更新神经网络
562
实现动作预测 564
片段的终点?
566
否
是
等待新查询图像
568

图5

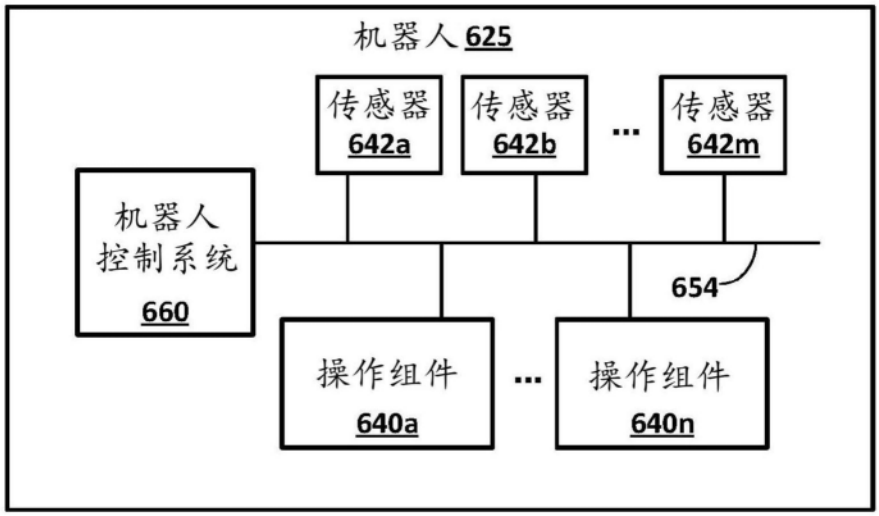

机器人 625
传感器 642a
传感器 642b
...
传感器 642m
机器人控制系统 660
654
操作组件 640a
...
操作组件 640n

图6

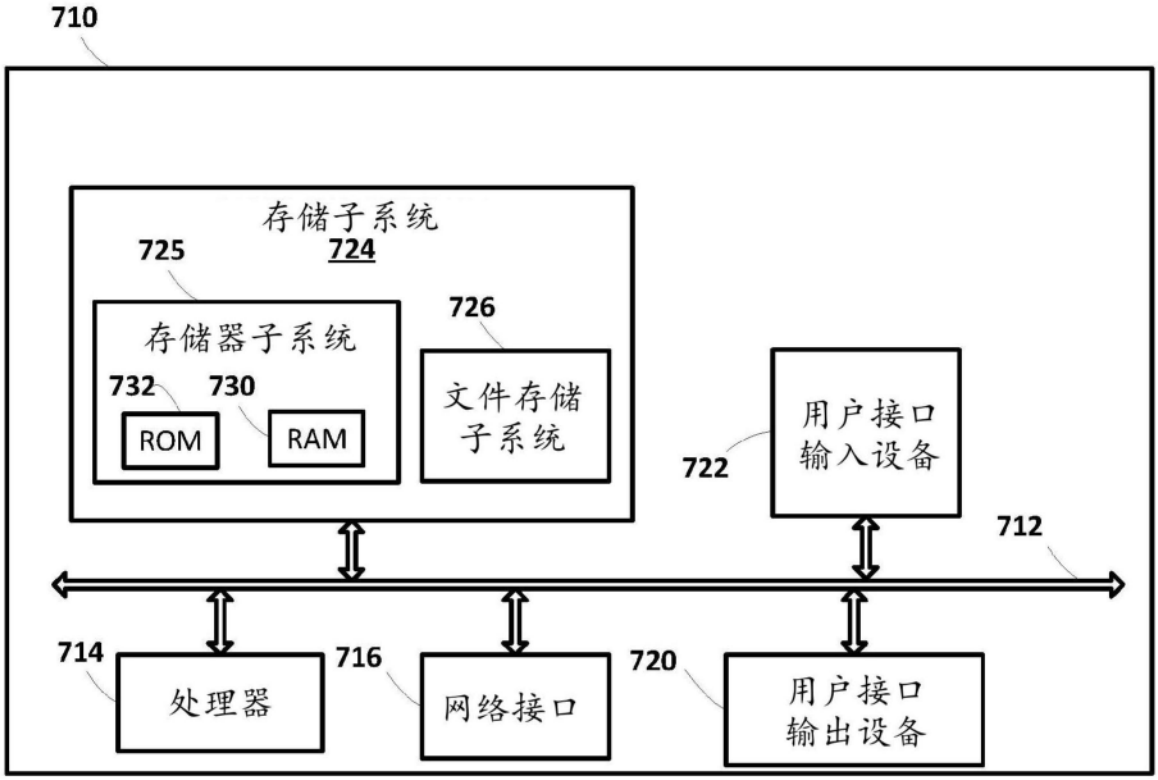

710
存储子系统
725
724
存储器子系统
732
730
ROM
RAM
726
文件存储子系统
722
用户接口输入设备
712
714
处理器
716
网络接口
720
用户接口输出设备

图7