



(21) 申请号 202380076697.8

(22) 申请日 2023.09.29

(30) 优先权数据

63/377,954 2022.09.30 US

63/378,012 2022.09.30 US

(85) PCT国际申请进入国家阶段日

2025.04.29

(86) PCT国际申请的申请数据

PCT/US2023/034168 2023.09.29

(87) PCT国际申请的公布数据

W02024/073076 EN 2024.04.04

(71) 申请人 特斯拉公司

地址 美国得克萨斯州

(72) 发明人 T·扎曼 Y·孙 J·鲍尔斯

I·戈扎里

(74) 专利代理机构 北京市金杜律师事务所

11256

专利代理师 张维

(51) Int.Cl.

G06N 3/0455 (2023.01)

G06N 20/00 (2019.01)

G06N 3/08 (2023.01)

H04N 19/105 (2014.01)

H04N 19/159 (2014.01)

H04N 19/172 (2014.01)

H04N 19/177 (2014.01)

B60W 60/00 (2020.01)

G06T 1/60 (2006.01)

H04N 19/85 (2014.01)

H04N 23/00 (2023.01)

权利要求书4页 说明书17页 附图7页

(54) 发明名称

用于机器学习模型的加速的基于视频的训练的系统和方法

(57) 摘要

系统和方法可以包括计算系统,该计算系统接收视频序列的比特流和指示视频序列的一个或多个图像帧的一个或多个指示,通过解析比特流来确定视频序列的图像帧的时间戳、在比特流内的位置和类型,并且使用一个或多个指示以及视频序列的图像帧的时间戳、在比特流内的位置和类型来确定比特流的一个或多个片段以用于进行解码以提取一个或多个图像帧。针对一个或多个图像帧中的图像帧,对应的片段表示视频序列中的图像帧的对应参考链。计算系统可以解码比特流的一个或多个片段并且使用一个或多个图像帧来训练ML模型。

300



1. 一种方法,包括:

由处理器接收视频序列的比特流和指示所述视频序列的一个或多个图像帧的一个或多个指示,以用于训练机器学习模型;

由所述处理器通过解析所述比特流来确定所述视频序列的图像帧的时间戳、在所述比特流内的位置和类型;

由所述处理器使用所述一个或多个指示以及所述视频序列的所述图像帧的所述时间戳、在所述比特流内的位置和类型来确定所述比特流的一个或多个片段以用于进行解码以提取所述一个或多个图像帧,针对所述一个或多个图像帧中的图像帧,对应的片段表示所述视频序列的图像帧的对应参考链;

由所述处理器解码所述比特流的所述一个或多个片段;以及

由所述处理器使用所述一个或多个图像帧来训练所述机器学习模型。

2. 根据权利要求1所述的方法,还包括:

由所述处理器在所述处理器的存储器内分配存储器区域;

由所述处理器将所述比特流存储在分配的所述存储器区域中;以及

由所述处理器在解码所述一个或多个片段之后将所述一个或多个图像帧存储在分配的所述存储器区域中。

3. 根据权利要求1所述的方法,其中针对所述一个或多个图像帧中的每个图像帧,所述图像帧的对应参考链在所述视频序列的帧内帧(I-帧)处开始,并且在所述一个或多个图像帧中的所述图像帧处结束。

4. 根据权利要求1所述的方法,其中确定所述视频序列的所述图像帧的所述时间戳、在所述比特流中的位置和类型包括生成一个或多个数据结构,所述一个或多个数据结构指示:针对所述视频序列的每个图像帧,表示在所述比特流中的所述图像帧的压缩数据的对应位置的对应时间戳和对应偏移;以及所述视频序列中的特定类型的图像帧。

5. 根据权利要求1所述的方法,其中所述处理器是图形处理单元。

6. 根据权利要求1所述的方法,其中所述比特流的所述一个或多个片段由集成在所述处理器中的硬件解码器解码。

7. 根据权利要求1所述的方法,其中所述一个或多个指示包括一个或多个时间值,并且针对所述一个或多个时间值中的每个时间值,确定所述比特流的所述一个或多个片段包括:

确定所述视频序列的所述图像帧的所述时间戳之中最接近所述时间值的第一时间戳,所述第一时间戳对应于所述视频序列中的第一图像帧;

确定所述视频序列的I-帧的第二时间戳,所述第二时间戳被确定为小于或等于所述第一时间戳的最接近所述第一时间戳的I-帧时间戳;

使用所述第二时间戳来确定在所述图像帧的所述位置之中的所述比特流中的所述I-帧的起始位置;

使用所述第一时间戳来确定在所述图像帧的所述位置之中的所述比特流中的所述第一图像帧的结束位置;以及

确定在所述I-帧的所述起始位置与所述第一图像帧的所述结束位置之间延伸的所述比特流的片段。

8. 根据权利要求1所述的方法, 其中所述比特流是由第一相机捕获的第一压缩视频序列的第一比特流, 并且所述方法还包括:

由所述处理器接收由第二相机捕获的第二视频序列的第二比特流;

由所述处理器通过解析所述第二比特流来确定所述第二视频序列的图像帧的时间戳、在所述第二比特流内的位置和类型;

由所述处理器使用所述一个或多个指示以及所述第二视频序列的所述图像帧的所述时间戳、所述第二比特流内的位置和类型来确定所述第二比特流的一个或多个片段以用于进行解码以提取所述第二视频序列的一个或多个第二图像帧, 针对所述一个或多个第二图像帧中的图像帧, 所述第二比特流的对应的片段表示所述第二视频序列的图像帧的对应参考链;

由所述处理器解码所述第二比特流的所述一个或多个片段; 以及

由所述处理器使用所述一个或多个第二图像帧来训练所述ML模型。

9. 根据权利要求8所述的方法, 其中所述第一相机和所述第二相机彼此不同步。

10. 根据权利要求9所述的方法, 其中所述第一比特流的所述一个或多个片段和所述第二比特流的所述一个或多个片段中的至少两个片段由集成在所述处理器中的至少两个硬件解码器并行解码。

11. 一种计算设备, 包括:

存储器; 以及

处理电路系统, 所述处理电路系统被配置为:

接收视频序列的比特流和指示所述视频序列的一个或多个图像帧的一个或多个指示, 以用于训练机器学习模型;

通过解析所述比特流来确定所述视频序列的图像帧的时间戳、在所述比特流内的位置和类型;

使用所述一个或多个指示和所述视频序列的所述图像帧的所述时间戳、在所述比特流内的位置和类型来确定所述比特流的一个或多个片段以用于进行解码以提取所述一个或多个图像帧, 针对所述一个或多个图像帧中的图像帧, 对应的片段表示所述视频序列的图像帧的对应参考链;

解码所述比特流的所述一个或多个片段; 以及

使用所述一个或多个图像帧来训练所述机器学习模型。

12. 根据权利要求11所述的计算设备, 其中所述处理电路系统还被配置为:

在所述存储器内分配存储器区域;

将所述比特流存储在分配的所述存储器区域中; 以及

在解码所述一个或多个片段之后, 将所述一个或多个图像帧存储在分配的所述存储器区域中。

13. 根据权利要求11所述的计算设备, 其中针对所述一个或多个图像帧中的每个图像帧, 所述图像帧的对应参考链在所述视频序列的帧内帧(I-帧)处开始, 并且在所述一个或多个图像帧中的所述图像帧处结束。

14. 根据权利要求11所述的计算设备, 其中在确定所述视频序列的所述图像帧的所述时间戳、在所述比特流中的位置和类型时, 所述处理电路系统被配置为生成一个或多个数

据结构,所述一个或多个数据结构指示:

针对所述视频序列的每个图像帧,指示在所述比特流中的所述图像帧的压缩数据的对应位置的对应时间戳和对应偏移;以及

所述视频序列中的特定类型的图像帧。

15. 根据权利要求11所述的计算设备,其中所述计算设备是图形处理单元(GPU)。

16. 根据权利要求15所述的计算设备,其中所述比特流的所述一个或多个片段由集成在GPU中的硬件解码器解码。

17. 根据权利要求11所述的计算设备,其中所述一个或多个指示包括一个或多个时间值,并且在确定所述比特流的所述一个或多个片段时,所述处理电路系统被配置为,针对所述一个或多个时间值中的每个时间值:

确定所述视频序列的所述图像帧的所述时间戳之中最接近所述时间值的第一时间戳,所述第一时间戳对应于所述视频序列中的第一图像帧;

确定所述视频序列的I-帧的第二时间戳,所述第二时间戳被确定为小于或等于所述第一时间戳的最接近所述第一时间戳的I-帧时间戳;

使用所述第二时间戳来确定在所述图像帧的所述位置之中的所述比特流中的所述I-帧的起始位置;

使用所述第一时间戳来确定在所述图像帧的所述位置之中的所述比特流中的所述第一图像帧的结束位置;以及

确定在所述I-帧的所述起始位置与所述第一图像帧的所述结束位置之间延伸的所述比特流的片段。

18. 根据权利要求11所述的计算设备,其中所述比特流是由第一相机捕获的第一压缩视频序列的第一比特流,并且所述处理电路系统还被配置为:

接收由第二相机捕获的第二视频序列的第二比特流;

确定解析所述第二比特流、所述第二视频序列的图像帧的时间戳、在所述第二比特流内的位置和类型;

使用所述一个或多个指示以及所述第二视频序列的所述图像帧的所述时间戳、所述第二比特流内的位置和类型来确定所述第二比特流的一个或多个片段以用于进行解码以提取所述第二视频序列的一个或多个第二图像帧,针对所述一个或多个第二图像帧中的图像帧,所述第二比特流的对应的片段表示所述第二视频序列的图像帧的对应参考链;

解码所述第二比特流的所述一个或多个片段;以及

使用所述一个或多个第二图像帧来训练所述机器学习模型。

19. 根据权利要求18所述的计算设备,其中所述第一相机和所述第二相机彼此不同步。

20. 根据权利要求19所述的计算设备,其中所述第一比特流的所述一个或多个片段和所述第二比特流的所述一个或多个片段中的至少两个片段由集成在所述计算设备中的至少两个硬件解码器并行解码。

21. 一种非暂态计算机可读介质,所述非暂态计算机可读介质在其上存储计算机代码指令,所述计算机代码指令在由处理器来执行时使得所述处理器:

接收视频序列的比特流和指示所述视频序列的一个或多个图像帧的一个或多个指示,以用于训练机器学习模型;

通过解析所述比特流来确定所述视频序列的图像帧的时间戳、在所述比特流内的位置和类型；

使用所述一个或多个指示和所述视频序列的所述图像帧的所述时间戳、在所述比特流内的位置和类型来确定所述比特流的一个或多个片段以用于进行解码以提取所述一个或多个图像帧，针对所述一个或多个图像帧中的图像帧，对应的片段表示所述视频序列的图像帧的对应参考链；

解码所述比特流的所述一个或多个片段；以及

使用所述一个或多个图像帧来训练所述机器学习模型。

用于机器学习模型的加速的基于视频的训练的系统和方法

[0001] 交叉引用相关专利申请

[0002] 本申请要求于2022年9月30日提交的美国临时申请号63/377,954和于2022年9月30日提交的美国临时申请号63/378,012的优先权,出于所有目的,这些申请中的每个申请的全部内容均通过引用并入本文。

技术领域

[0003] 本公开总体上涉及机器学习(ML)模型的视频训练。具体地,本公开涉及用于利用视频数据的ML模型的加速训练的系统和方法。

背景技术

[0004] 由于计算机技术的快速发展,用于自主车辆和机器人(统称为自体)的自主导航技术已经变得无处不在。这些进步允许自体的更安全、更可靠的自主导航。自体通常需要在复杂和动态的环境和地形中导航,这些环境和地形可以包括车辆、交通、行人、骑自行车的人以及各种其他静态或动态障碍物。理解自体的周围环境针对明智和有能力的决策以避免冲突是必要的。

发明内容

[0005] 本文中描述的系统、设备和方法提供了机器学习(ML)模型的加速训练。具体地,针对利用视频数据训练的ML模型,本文中描述的系统、设备和方法实现了对视频数据的快速解码以及对计算和存储器资源的有效使用。针对被用于预测或感知自体的周围环境的ML模型或人工智能(AI)模型,诸如占用网络,此类模型的训练极其耗时。本文中描述的系统、设备和方法显著加速了对此类模型的训练和/或验证。

[0006] 在一个实施例中,一种方法可以包括:由处理器接收视频序列的比特流和指示视频序列的一个或多个图像帧的一个或多个指示,以用于训练机器学习(ML)模型;由处理器通过解析比特流来确定视频序列的图像帧的时间戳、在比特流内的位置和类型;由处理器使用一个或多个指示以及视频序列的图像帧的时间戳、在比特流内的位置和类型来确定比特流的一个或多个片段以用于进行解码以提取一个或多个图像帧,使得针对该一个或多个图像帧中的图像帧,对应的片段表示视频序列的图像帧的对应参考链;由处理器解码比特流的一个或多个片段;以及由处理器使用一个或多个图像帧来训练ML模型。

[0007] 该方法还可以包括:由处理器在处理器的存储器内分配存储器区域;由处理器将比特流存储在分配的存储器区域中;以及由处理器在解码一个或多个片段之后将一个或多个图像帧存储在分配的存储器区域中。

[0008] 针对一个或多个图像帧中的每个图像帧,图像帧的对应参考链可以在视频序列的帧内帧(I-帧)处开始,并且在一个或多个图像帧中的图像帧处结束。

[0009] 确定视频序列的图像帧的时间戳、在比特流中的位置和类型可以包括生成一个或多个数据结构,该一个或多个数据结构存储:(i) 针对视频序列的每个图像帧,指示在比特

流中的图像帧的压缩数据的对应位置的对应时间戳和对应偏移,以及(ii)针对视频序列中的特定类型的图像帧的每个图像帧,该特定类型的对应指示。

[0010] 处理器可以是图形处理单元(GPU)。在一些实现中,比特流的一个或多个片段可以由集成在处理器中的硬件解码器解码。

[0011] 一个或多个指示可以包括一个或多个时间值,并且确定比特流的一个或多个片段包括,针对一个或多个时间值中的每个时间值,可以包括:确定视频序列的图像帧的时间戳之中最接近该时间值的第一时间戳,使得第一时间戳对应于视频序列中的第一图像帧;确定视频序列的I-帧的第二时间戳,该第二时间戳被确定为小于或等于第一时间戳的最接近第一时间戳的I-帧时间戳;使用第二时间戳来确定在图像帧的位置之中的比特流中的I-帧的起始位置;使用第一时间戳来确定在图像帧的位置之中的比特流中的第一图像帧的结束位置;以及确定在I-帧的起始位置与第一图像帧的结束位置之间延伸的比特流的片段。

[0012] 比特流可以是由第一相机捕获的第一压缩视频序列的第一比特流,并且该方法还可以包括:由处理器接收由第二相机捕获的第二视频序列的第二比特流;由处理器通过解析第二比特流来确定第二视频序列的图像帧的时间戳、在第二比特流内的位置和类型;由处理器使用一个或多个指示以及第二视频序列的图像帧的时间戳、在第二比特流内的位置和类型来确定第二比特流的一个或多个片段以用于进行解码以提取第二视频序列的一个或多个第二图像帧,使得针对该一个或多个第二图像帧中的图像帧,第二比特流的对应的片段表示第二视频序列的图像帧的对应参考链;由处理器解码第二比特流的一个或多个片段;以及由处理器使用一个或多个第二图像帧来训练ML模型。第一相机和第二相机可以不彼此同步。

[0013] 第一比特流的一个或多个片段和第二比特流的一个或多个片段中的至少两个片段可以由集成在处理器中的至少两个硬件解码器并行解码。

[0014] 在另一实施例中,计算设备可以包括存储器和处理电路系统。处理电路系统可以被配置为接收视频序列的比特流和指示视频序列的一个或多个图像帧的一个或多个指示,以用于训练机器学习(ML)模型;通过解析比特流来确定视频序列的图像帧的时间戳、在比特流内的位置和类型;使用一个或多个指示和视频序列的图像帧的时间戳、在比特流内的位置和类型来确定比特流的一个或多个片段以用于进行解码以提取一个或多个图像帧,使得针对该一个或多个图像帧中的图像帧,对应的片段表示视频序列的图像帧的对应参考链;解码比特流的一个或多个片段;以及使用一个或多个图像帧来训练ML模型。

[0015] 处理电路系统还可以被配置为在存储器内分配存储器区域;将比特流存储在分配的存储器区域中;并且在解码一个或多个片段之后,将一个或多个图像帧存储在分配的存储器区域中。

[0016] 针对一个或多个图像帧中的每个图像帧,图像帧的对应参考链可以在视频序列的帧内帧(I-帧)处开始,并且在一个或多个图像帧中的图像帧处结束。

[0017] 在确定视频序列的图像帧的时间戳、在比特流中的位置和类型时,处理电路系统可以被配置为生成一个或多个数据结构。一个或多个数据结构可以存储:(i)针对视频序列的每个图像帧,指示在比特流中的图像帧的压缩数据的对应位置的对应时间戳和对应偏移,以及(ii)针对视频序列中的特定类型的图像帧中的每个图像帧,该特定类型的对应指示。

[0018] 计算设备可以是图形处理单元 (GPU)。比特流的一个或多个片段可以由集成在GPU中的硬件解码器解码。

[0019] 一个或多个指示可以包括一个或多个时间值,并且在确定比特流的一个或多个片段时,处理电路系统可以被配置为,针对一个或多个时间值中的每个时间值:确定视频序列的图像帧的时间戳之中最接近该时间值的第一时间戳,该第一时间戳对应于视频序列中的第一图像帧;确定视频序列的I-帧的第二时间戳,该第二时间戳被确定为小于或等于第一时间戳的最接近第一时间戳的I-帧时间戳;使用第二时间戳来确定在图像帧的位置之中的比特流中的I-帧的起始位置;使用第一时间戳来确定在图像帧的位置之中的比特流中的第一图像帧的结束位置;以及确定在I-帧的起始位置与第一图像帧的结束位置之间延伸的比特流的片段。

[0020] 比特流可以由第一相机捕获的第一压缩视频序列的第一比特流,并且处理电路系统还可以被配置为接收由第二相机捕获的第二视频序列的第二比特流;确定解析第二比特流、第二视频序列的图像帧的时间戳、在第二比特流内的位置和类型;使用一个或多个指示以及第二视频序列的图像帧的时间戳、在第二比特流内的位置和类型来确定第二比特流的一个或多个片段以用于进行解码以提取第二视频序列的一个或多个第二图像帧,针对该一个或多个第二图像帧中的图像帧,第二比特流的对应的片段表示第二视频序列的图像帧的对应参考链;解码第二比特流的一个或多个片段;以及使用一个或多个第二图像帧来训练ML模型。第一相机和第二相机可以彼此不同步。

[0021] 第一比特流的一个或多个片段和第二比特流的一个或多个片段中的至少两个片段可以由集成在计算设备中的至少两个硬件解码器并行解码。

[0022] 在又一实施例中,非暂态计算机可读介质可以在其上存储计算机代码指令。计算机代码指令当由处理器执行时可以使处理器:接收视频序列的比特流和指示视频序列的一个或多个图像帧的一个或多个指示,以用于训练机器学习 (ML) 模型;通过解析比特流来确定视频序列的图像帧的时间戳、在比特流内的位置和类型;使用一个或多个指示以及视频序列的图像帧的时间戳、在比特流中的位置和类型来确定比特流的一个或多个片段以用于进行解码以提取一个或多个图像帧,针对该一个或多个图像帧中的图像帧,对应的片段表示视频序列的图像帧的对应参考链;解码比特流的一个或多个片段;以及使用一个或多个图像帧来训练ML模型。

附图说明

[0023] 本公开的非限制性实施例通过与附图相关的示例来描述,附图是示意性的并且不旨在按比例绘制。除非被指示为表示背景技术,否则附图表示本公开的方面。

[0024] 图1A图示了根据实施例的AI使能的视觉数据分析系统的组件。

[0025] 图1B图示了根据实施例的与自体相关联的各种传感器。

[0026] 图1C图示了根据实施例的车辆组件。

[0027] 图2图示了根据实施例的视频训练系统的框图。

[0028] 图3图示了根据实施例的用于利用视频数据的机器学习 (ML) 模型的加速训练的方法的流程图。

[0029] 图4图示了根据实施例的描绘视频序列中的一组选定图像帧和要被解码的对应图

像帧的图。

[0030] 图5图示了根据实施例的对应于图4的视频序列的比特流、对应于选定图像帧的压缩数据和对应于要被解码的帧的压缩数据的图。

具体实施方式

[0031] 现在将参考附图中描绘的说明性实施例,并且这里将使用特定语言来描述它们。然而,应当理解,并不旨在因此限制权利要求或本公开的范围。相关领域的并且拥有本公开的技术人员可以想到的本文所示的发明特征的改变和进一步修改、以及本文所示主题内容的原理的附加应用将被认为在本文所公开的主题内容的范围内。在不脱离本公开的精神或范围的情况下,可以使用其他实施例和/或可以进行其他改变。具体实施方式中描述的说明性实施例并不意味着限制所呈现的主题内容。

[0032] 用视频数据训练ML模型通常非常耗时。此外,处理(例如,解码)视频序列消耗显著的处理能力和存储器。针对要利用大量视频数据进行训练的ML模型或AI模型,对模型的训练(或验证)可能是极为耗时的,并且在处理、存储器和带宽资源方面要求极高。例如,训练ML模型或AI模型以预测或感知自体的周围环境涉及使用数百万甚至数十亿个视频帧用于训练数据。视频帧通常被存储为压缩视频数据。解码和处理这种大量的视频数据以训练(多个)ML模型可能花费数千小时。本文中描述的系统、设备和方法提供了对计算和存储器资源的加速训练和更有效使用。具体地,本文中描述的系统、设备和方法实现了对被配置为预测或感知自体的三维周围环境的占用网络的加速训练和/或验证。

[0033] 图1A是可以在其中实现本文中讨论的方法和系统的系统的组件的非限制性示例。例如,分析服务器可以训练AI模型,并且使用经训练的AI模型来生成用于一个或多个自体的占用数据集和/或地图。图1A图示了AI使能的视觉数据分析系统100的组件。系统100可以包括分析服务器110a、系统数据库110b、管理员计算设备120、自体140a至140b(统称为(多个)自体140)、自体计算设备141a至141c(统称为自体计算设备141)和服务器160。系统100不限于本文中描述的组件,并且可以包括为简洁起见而未示出的附加或其他组件,这些组件将被认为在本文中描述的实施例的范围内。

[0034] 上述组件可以通过网络130连接。网络130的示例可以包括但不限于私有或公共LAN、WLAN、MAN、WAN和互联网。网络130可以包括根据一个或多个标准和/或经由一个或多个传输介质的有线和/或无线通信。

[0035] 网络130上的通信可以根据各种通信协议(诸如传输控制协议和互联网协议(TCP/IP)、用户数据报协议(UDP)和IEEE通信协议)来执行。在一个示例中,网络130可以包括根据蓝牙规范集或另一标准或专有无线通信协议的无线通信。在另一示例中,网络130还可以包括蜂窝网络上的通信,包括例如GSM(全球移动通信系统)、CDMA(码分多址)或EDGE(用于全球演进的增强型数据)网络。

[0036] 系统100图示了可以被用于训练和执行一个或多个AI模型(诸如(多个)AI模型110c)的系统架构和组件的示例。具体地,如图1A所描绘和本文中描述的,分析服务器110a可以使用本文中讨论的方法以使用从自体140取回的数据(例如,通过使用数据流172和174)来训练(多个)AI模型110c。当(多个)AI模型110c已经被训练时,自体140中的每个自体可以访问并且执行(多个)经训练的AI模型110c。例如,具有自体计算设备141a的车辆140a

可以将其相机馈送传输到(多个)经训练的AI模型110c,并且可以确定其周围环境(例如,数据流174)的占用状态。此外,由(多个)AI模型110c相对于自体140(在推理时)摄取和/或预测的数据也可以被用于改进(多个)AI模型110c。因此,系统100描绘了连续的循环,该循环可以周期性地提高(多个)AI模型110c的准确性。此外,系统100描绘了循环,其中除了推理阶段外,自体140所接收的数据还可以在训练阶段被使用。

[0037] 分析服务器110a可以被配置为收集、处理和分析导航数据(例如,在导航时捕获的图像)和从自体140收集的各种传感器数据。然后,收集的数据可以被处理并且准备成训练数据集。然后,训练数据集可以被用于训练一个或多个AI模型,诸如AI模型110c。分析服务器110a还可以被配置为从自体140收集视觉数据。使用AI模型110c(使用本文中讨论的方法和系统训练),分析服务器110a可以生成用于自体140的数据集和/或占用地图。分析服务器110a可以在自体140上显示占用地图和/或将占用地图/数据集传输到自体计算设备141、管理员计算设备120和/或服务器160。

[0038] 在图1A中,AI模型110c被图示为系统数据库110b的组件,但AI模型110c可以被存储在不同或单独的组件中,诸如云存储装置或分析服务器110a可访问的任何其他数据存储库。

[0039] 分析服务器110a还可以被配置为显示电子平台,该电子平台图示了用于训练AI模型110c的各种训练属性。电子平台可以被显示在管理员计算设备120上,使得分析师可以监测AI模型110c的训练。由分析服务器110a生成和托管的电子平台的示例可以是被配置为显示从自体140收集的训练数据集和/或AI模型110c的训练状态/度量的基于web的应用或网站。

[0040] 分析服务器110a可以是包括能够执行本文中描述的各种任务和过程的处理器和非暂态机器可读存储装置的任何计算设备。这种计算设备的非限制性示例可以包括工作站计算机、膝上型计算机、服务器计算机等。虽然系统100包括单个分析服务器110a,但系统100可以包括在分布式计算环境(诸如云环境)中操作的任何数目的计算设备。

[0041] 自体140可以表示将与其先前或当前导航会话相关联的数据传输到分析服务器110a的各种电子数据源。自体140可以是被配置用于导航的任何装置,诸如车辆140a和/或卡车140c。自体140不限于是车辆,并且也可以包括机器人设备。例如,自体140可以包括机器人140b,该机器人140b可以表示能够导航各种地形的通用、双足、自主的人形机器人。机器人140b可以配备有实现平衡、导航、感知或与物理世界交互的软件。机器人140b还可以包括被配置为将视觉数据传输到分析服务器110a的各种相机。

[0042] 尽管本文中被称为“自体”,但自体140可以是也可以不是被配置用于自动导航的自主设备。例如,在一些实施例中,自体140可以由人类操作员或远程处理器控制。自体140可以包括各种传感器,诸如图1B中所描绘的传感器。传感器可以被配置为在自体140导航各种地形(例如,道路)时收集数据。分析服务器110a可以收集由自体140提供的数据。例如,分析服务器110a可以从各种传感器获得导航会话和/或道路/地形数据(例如,在道路上导航的自体140的图像),使得收集的数据最终被AI模型110c使用用于训练目的。

[0043] 如本文中使用的,导航会话与自体140行进路线的行程相对应,而不管该行程是自主的还是由人类控制的。在一些实施例中,导航会话可以用于数据收集和模型训练目的。然而,在一些其他实施例中,自体140可以是指由消费者购买的车辆,并且行程的目的可以被

归类为日常使用。当自体140从非移动位置移动超过阈值距离(例如,0.1英里、100英尺)或超过阈值速率(例如,超过0mph、超过1mph、超过5mph)时,导航会话可以开始。当自体140返回到非移动位置和/或关闭时(例如,当驾驶员离开车辆时),导航会话可以结束。

[0044] 自体140可以表示由分析服务器110a监测以训练(多个)AI模型110c的自体的集合。例如,用于车辆140a的驾驶员可以授权分析服务器110a监测与其相应车辆相关联的数据。作为结果,分析服务器110a可以利用本文中讨论的各种方法来收集传感器/相机数据,并且生成训练数据集来相应地训练(多个)AI模型110c。分析服务器110a然后可以应用(多个)经训练的AI模型110c来分析自体140相关联的数据,并且预测用于自体140的占用地图。此外,与自体140相关联的附加/正在进行的数据也可以被处理并且添加到训练数据集中,以便分析服务器110a相应地重新校准(多个)AI模型110c。因此,系统100描绘了其中从自体140接收的导航数据可以被用于训练(多个)AI模型110c的循环。自体140可以包括出于导航目的执行(多个)经训练的AI模型110c的处理器。在导航时,自体140可以收集关于其导航会话的附加数据,并且这些附加数据可以被用于校准(多个)AI模型110c。也就是说,自体140表示可以被用于训练、执行/使用和重新校准(多个)AI模型110c的自体。在非限制性示例中,自体140表示由客户购买的车辆,这些车辆可以使用(多个)AI模型110c自主导航,同时改进(多个)AI模型110c。

[0045] 自体140可以配备有允许自体从其周围环境收集数据并且(可能)自主导航的各种技术。例如,自体140可以配备有推理芯片来运行自动驾驶软件。

[0046] 用于每个自体140的各种传感器可以监测收集的与不同导航会话相关联的数据,并且将该数据传输到分析服务器110a。图1B至图1C图示了根据实施例的集成在自体140内的传感器的框图。关于图1B至图1C而讨论的每个传感器的数目和位置可以取决于图1A中讨论的自体的类型。例如,机器人140b可以包括与车辆140a或卡车140c不同的传感器。例如,机器人140b可以不包括安全气囊激活传感器170q。此外,车辆140a和卡车140c的传感器可以与图1C所示的不同地被定位。

[0047] 如本文所讨论的,集成在每个自体140内的各种传感器可以被配置为测量与每个导航会话相关联的各种数据。分析服务器110a可以定期收集由这些传感器监测和收集的数据,其中数据根据本文中描述的方法被处理,并且被用于训练AI模型110c和/或执行AI模型110c以生成占用地图。

[0048] 自体140可以包括用户界面170a。用户界面170a可以是指自体计算设备(例如,图1A中的自体计算设备141)的用户界面。用户界面170a可以被实现为与车辆内部集成或被耦合至车辆内部的显示屏、平视显示器、触摸屏等。用户界面170a可以包括输入设备,诸如触摸屏、旋钮、按钮、键盘、鼠标、手势传感器、方向盘等。在各种实施例中,用户界面170a可以适于向自体140的其他设备或传感器(例如,图1B所示的传感器)(诸如控制器170c)提供用户输入(例如,作为一种信号和/或传感器信息)。

[0049] 用户界面170a还可以用一个或多个逻辑设备来实现,这些逻辑设备可以适于执行指令,诸如软件指令,以实现本文中描述的各种过程和/或方法中的任何过程和/或方法。例如,用户界面170a可以适于形成通信链路、传输和/或接收通信(例如,传感器信号、控制信号、传感器信息、用户输入和/或其他信息)或执行各种其他过程和/或方法。在另一示例中,驾驶员可以使用用户界面170a来控制自体140的温度或激活其特征(例如,自动驾驶或转向

系统170o)。因此,用户界面170a可以结合本文中描述的其他传感器来监测和收集驾驶会话数据。用户界面170a还可以被配置为显示由分析服务器110a和/或AI模型110c生成/预测的各种数据。

[0050] 取向传感器170b可以被实现为指南针、浮子、加速度计、和/或能够测量自体140的取向(例如,相对于一个或多个参考取向(诸如重力和/或磁北)的横滚、俯仰和/或偏航的幅度和方向)的其他数字或模拟设备中的一种或多种。取向传感器170b可以适于向自体140提供航向测量。在其他实施例中,取向传感器170b可以适于使用取向测量的时间序列向自体140提供横滚、俯仰和/或偏航率。取向传感器170b可以被定位和/或适于相对于自体140的特定坐标系进行取向测量。

[0051] 控制器170c可以被实现为任何适当的逻辑设备(例如,处理设备、微控制器、处理器、应用专用集成电路(ASIC)、现场可编程门阵列(FPGA)、存储器存储设备、存储器读取器、或其他设备或设备组合),该逻辑设备可以适于执行、存储和/或接收适当的指令,诸如实现用于控制自体140的各种操作的控制回路的软件指令。这种软件指令还可以实现用于处理传感器信号、确定传感器信息、提供用户反馈(例如,通过用户界面170a)、向设备查询操作参数、选择用于设备的操作参数或执行本文中描述的各种操作中的任何操作的方法。

[0052] 通信模块170e可以被实现为任何有线和/或无线接口,该接口被配置为将传感器数据、配置数据、参数和/或其他数据和/或信号传送到图1A所示的任何特征(例如,分析服务器110a)。如本文所述,在一些实施例中,通信模块170e可以以分布式方式实现,使得通信模块170e的部分在图1B所示的一个或多个元素和传感器内实现。在一些实施例中,通信模块170e可以延迟传送传感器数据。例如,当自体140不具有网络连接时,通信模块170e可以将传感器数据存储在临时数据存储装置中,并且在自体140被标识为具有适当的网络连接时传输传感器数据。

[0053] 速度传感器170d可以被实现为电子皮托管、计量齿轮或轮子、水速传感器、风速传感器、风速率传感器(例如,方向和幅度)、和/或能够测量或确定自体140的线速度(例如,在周围介质中和/或与自体140纵轴对准)并且提供这种测量作为传感器信号(其可以被传送给各种设备)的其他设备。

[0054] 陀螺仪/加速度计170f可以被实现为一个或多个电子六分仪、半导体设备、集成芯片、加速度计传感器、或能够测量自体140的角速度/加速度和/或线性加速度(例如,方向和幅度)并且提供这种测量作为传感器信号(其可以被传送给各种设备,诸如分析服务器110a)的其他系统或设备。陀螺仪/加速度计170f可以被定位和/或适于相对于自体140的特定坐标系进行这种测量。在各种实施例中,陀螺仪/加速度计170f可以与图1B所描绘的其他元素在公共外壳和/或模块中被实现,以确保公共参考系或参考系之间的已知变换。

[0055] 全球导航卫星系统(GNSS)170h可以被实现为全球定位卫星接收器和/或能够基于例如从空间和/或地面源接收的无线信号来确定自体140的绝对和/或相对位置并且能够提供诸如传感器信号的测量(其可以被传送到各种设备)的另一设备。在一些实施例中,GNSS170h可以适于确定自体140的速率、速度和/或偏航率(例如,使用位置测量的时间序列),诸如自体140的绝对速率和/或角速率的偏航分量。

[0056] 温度传感器170i可以被实现为热敏电阻、电传感器、电温度计、和/或能够测量与自体140相关联的温度并且提供这种测量作为传感器信号的其他设备。温度传感器170i可

以被配置为测量与自体140相关联的环境温度,诸如驾驶舱或仪表板温度,例如,该环境温度可以被用于估计自体140的一个或多个元素的温度。

[0057] 湿度传感器170j可以被实现为相对湿度传感器、电传感器、电相对湿度传感器、和/或能够测量与自体140相关联的相对湿度并且提供这种测量作为传感器信号的另一设备。

[0058] 转向传感器170g可以适于根据由逻辑设备(诸如控制器170c)提供的一个或多个控制信号和/或用户输入来物理地调节自体140的航向。转向传感器170g可以包括自体140的一个或多个致动器和控制表面(例如,方向舵或其他类型的转向或调整机构),并且可以适于将控制表面物理地调节到各种正和/或负转向角度/位置。转向传感器170g还可以适于感测这种转向机构的当前转向角度/位置,并且提供这种测量。

[0059] 推进系统170k可以被实现为螺旋桨、涡轮机或其他基于推力的推进系统、机械轮式和/或履带式推进系统、基于风/帆的推进系统、和/或可以被用于向自体140提供动力的其他类型的推进系统。推进系统170k还可以监测自体140的原动力和/或推力相对于自体140的参考坐标系的方向。在一些实施例中,推进系统170k可以被耦合至传感器170g和/或与转向传感器170g集成。

[0060] 乘客约束传感器170l可以监测安全带检测和锁定/解锁组件以及其他乘客约束子系统。乘客约束传感器170l可以包括各种环境和/或状态传感器、致动器、和/或促进与自体140的操作相关联的安全机制的操作的其他设备。例如,乘客约束传感器170l可以被配置为从图1B所描绘的其他传感器接收运动和/或状态数据。乘客约束传感器170l可以确定安全测量(例如,安全带)是否正在被使用。

[0061] 如图1C所描绘的,相机170m可以是指集成在自体140内的一个或多个相机,并且可以包括集成(或改装)到自体140中的多个相机。相机170m可以是自体140的面向内部或外部的相机。例如,如图1C所描绘的,自体140可以包括一个或多个面向内部的相机,这些相机可以监测和收集自体140的乘客的镜头。自体140可以包括八个面向外部的相机。例如,自体140可以包括前置相机170m-1、前视侧相机170m-2、前视侧相机170m-3、每个前挡泥板上的后视侧相机170m-4、每侧的相机170m-5(例如,集成在B柱内)和后置相机170m-6。

[0062] 参照图1B,雷达170n和超声传感器170p可以被配置为监测自体140到其他对象(诸如其他车辆或不可移动对象(例如,树木或车库门))的距离。自体140还可以包括被配置为使用经由各种传感器(例如,雷达170n、速度传感器170d和/或超声传感器170p)收集的数据来自导航自体140的自动驾驶或转向系统170o。

[0063] 因此,自动驾驶或转向系统170o可以分析由本文中描述的一个或多个传感器收集的各种数据,以标识驾驶数据。例如,自动驾驶或转向系统170o可以基于自体140的速度和其到道路上的另一车辆的距离来计算前方碰撞的风险。自动驾驶或转向系统170o还可以确定驾驶员是否正在触摸方向盘。自动驾驶或转向系统170o可以将分析的数据传输到本文中讨论的各种特征,诸如分析服务器。

[0064] 气囊激活传感器170q可以预测或检测碰撞,并且引起一个或多个气囊的激活或展开。气囊激活传感器170q可以传输关于气囊展开的数据,包括与引起展开的事件相关联的数据。

[0065] 再次参照图1A,管理员计算设备120可以表示由系统管理员操作的计算设备。管理

员计算设备120可以被配置为显示由分析服务器110a取回或生成的数据(例如,各种分析度量和风险得分),其中系统管理员可以监测由分析服务器110a利用的各种模型,审阅反馈,和/或促进由分析服务器110a维持的(多个)AI模型110c的训练。

[0066] (多个)自体140可以是被配置为导航各种路线的任何设备,诸如车辆140a或机器人140b。如关于图1B至图1C所讨论的,自体140可以包括各种遥测传感器。自体140还可以包括自体计算设备141。具体地,每个自体可以具有其自己的自体计算设备141。例如,卡车140c可以具有自体计算设备141c。为简洁起见,自体计算设备被统称为(多个)自体计算设备141。自体计算设备141可以控制自体140的信息娱乐系统上的内容呈现,处理与信息娱乐系统相关联的命令,聚合传感器数据,管理数据到电子数据源的通信,接收更新和/或传输消息。在一个配置中,自体计算设备141与电子控制单元通信。在另一配置中,自体计算设备141是电子控制单元。自体计算设备141可以包括能够执行本文中描述的各种任务和过程的处理器和非暂态机器可读存储介质。例如,本文中描述的(多个)AI模型110c可以由自体计算设备141存储和执行(或直接访问)。自体计算设备141的非限制性示例可以包括车辆多媒体和/或显示系统。

[0067] 在如何用视频数据加速(多个)AI模型110c和/或其他ML模型的训练的一个示例中,分析服务器110a可以包括多个图形处理单元(GPU),多个图形处理单元被配置为并行训练(多个)AI模型110c。例如,分析服务器110a可以包括超级计算机。每个GPU可以接收视频数据(例如,一个或多个比特流)和要从视频数据中被提取并且被用于训练(多个)AI模型110c的视频帧(或图像帧)的指示。GPU可以仅对解码选定图像帧所需的(多个)比特流的部分进行解码,并且使用解码形式的选定图像帧来训练(多个)AI模型110c。

[0068] 在一些实现中,每个GPU可以被配置或被设计为独立地执行训练步骤,而无需使用外部资源。具体地,GPU可以接收视频数据,解码相关部分或片段以提取选定图像帧,从选定图像帧中提取特征,并且使用提取的特征来训练(多个)AI模型110c,而无需使用任何外部存储器或处理资源。换句话说,从接收视频数据的点到(多个)AI模型110c的训练的所有处理和数据处理都可以在GPU内被内部地并且独立地执行。

[0069] 每个GPU可以包括被集成在其中的一个或多个硬件视频解码器,以加速视频解码。GPU可以包括多个视频解码器以并行化视频解码。并行化可以以各种方式来实现,例如按视频片段、按比特流或按训练会话。

[0070] 在一些实现中,GPU可以具有足够的内部存储器(例如高速缓存存储器)来存储执行训练步骤所需的数据。GPU可以在存储器内分配存储器区域来存储与训练步骤相关联的数据,并且在整个训练步骤中使用分配的存储器区域以用于数据存储。

[0071] 图2图示了根据实施例的用于训练ML模型的计算机环境200的框图。计算机环境200可以包括用于训练ML模型的训练系统202和用于存储训练数据和/或验证数据的数据存储系统204。训练系统202可以包括多个训练节点(或处理节点)206。每个训练节点206可以包括相应的数据加载器(或数据加载设备)208和相应的图形处理单元(GPU)210。每个GPU210可以包括存储器(例如,高速缓存存储器)212、处理电路系统214和一个或多个视频解码器216。

[0072] 数据存储系统204可以包括分布式存储系统,或者可以是分布式存储系统。例如,数据存储系统204可以具有可以跨多个物理服务器(诸如超级计算机)拆分数据的基础设

施。数据存储系统204可以包括存储单元的一个或多个存储集群,具有用于并行和加速访问来自(多个)存储集群的多个节点或存储单元的数据的机制和基础设施。例如,数据存储系统204可以包括足够的数据链路和带宽,以并行地或同时地将数据递送到训练节点206。

[0073] 数据存储系统204可以包括足够的存储器容量来例如以压缩形式存储数百万甚至数十亿个视频帧。例如,数据存储系统204可以具有存储多个拍字节(例如10、20或30拍字节)的数据的存储器容量。数据存储系统204可以允许数千个视频序列在任何时间实例移入和/或移出数据存储系统204。存储系统204的相对巨大的尺寸和带宽允许一个或多个ML模型的并行训练,如下面所讨论的。

[0074] 训练系统202可以被实现为一个或多个物理服务器,诸如服务器110a。例如,训练系统202可以被实现为一个或多个超级计算机。每个超级计算机可以包括数千个处理或训练节点206。训练节点206可以被配置或被设计为支持一个或多个ML模型(诸如(多个)AI模型110c)的并行训练。每个训练节点206可以被通信耦合到存储系统204以访问存储在其中的数据训练数据和/或验证数据。

[0075] 每个训练节点206可以包括彼此通信耦合的相应数据加载器208和相应GPU 210。数据加载器208可以是(或可以包括)处理器或中央处理单元(CPU)以用于处理对应的GPU 210(例如,同一训练节点206中的GPU 210)与数据存储系统204之间的数据请求或数据传送。例如,GPU 210可以请求由关于图1C所描述的相机170m中的一个或多个相机捕获的一个或多个视频序列。例如,前置或前视相机170m-1、170m-2和170m-3、后视侧相机170m-4、侧相机170m-5和后视相机170m-6可以同时捕获视频序列,并且将视频序列发送到数据存储系统204并且将视频序列存储在数据存储系统204中以用于存储。在一些实现中,数据存储系统204可以将由自体140的多个相机(诸如相机170m)同时捕获的视频序列存储为一组视频序列或视频序列的组合,这些视频序列可以一起被递送到训练节点206。例如,数据存储系统204可以维护附加数据,该附加数据指示哪些视频序列是由相机170m同时捕获的,或者表示来自不同相机角度的同一场景。数据存储系统204可以维护(例如,针对每个存储的视频序列)指示与视频序列相关联的自体标识符、相机标识符和时间实例的数据。

[0076] 训练节点206可以使用由自体140的相机170m捕获并且存储在数据存储系统204中的视频数据同时(例如并行)训练一个或多个ML模型。在一些实现中,视频数据可以由多个自体140的相机170m捕获。在训练节点206中,对应的数据加载器208可以从数据存储系统204中请求由自体140的一个或多个相机170m在时间间隔期间同时捕获的一个或多个视频序列的视频数据,并且将接收到的视频数据发送到对应的GPU 210以用于在训练ML模型时使用来执行训练步骤(或验证步骤)。视频数据可以是压缩形式。例如,视频序列可以由集成在或实现在自体140中的编码器来编码。每个数据加载器208可以具有足够的处理能力和带宽,以保持GPU 210繁忙的方式来将视频数据递送到对应的GPU 210。换句话说,GPU 210可以被配置或被设计为,例如,在处理能力和带宽方面,从数据存储系统204中请求一组压缩视频序列的视频数据,并且在小于或等于GPU 210处理一组视频序列所消耗的平均时间的持续时间内将视频递送到GPU 210。

[0077] 每个GPU 210可以包括对应的内部存储器212,诸如高速缓存存储器,以存储用于执行本文所述过程的可执行指令、一个或多个视频序列的接收视频数据、解码的视频帧、从解码的视频帧中提取的特征、参数或经训练的ML模型的数据、和/或被用于训练ML模型的其

他数据。存储器212可以足够大,以存储执行单个训练步骤所需的所有数据。如本文中使用的,训练步骤可以包括接收和解码一个或多个视频序列(例如,由自体140的一个或多个相机170m同时捕获的一组视频序列)的视频数据,从解码的视频数据中提取特征,并且使用提取的特征来更新正被训练或验证的ML模型的参数。

[0078] 每个GPU 210可以包括处理电路系统214以执行本文中描述的过程或方法。处理电路系统214可以包括一个或多个微处理器、多核处理器、数字信号处理器(DSP)、一个或多个逻辑电路或其组合。处理电路系统214可以执行计算机代码指令(例如,存储在存储器212中),以执行本文中描述的过程或方法。

[0079] GPU 210可以包括用于解码从数据存储系统204接收的视频数据的一个或多个视频解码器216。一个或多个视频解码器216可以包括集成在GPU 110中的(多个)硬件视频解码器以加速视频解码。一个或多个视频解码器216可以是处理电路系统214的一部分,或者可以包括通信耦合到处理电路系统214的单独的(多个)电子电路。

[0080] 每个GPU 210可以被配置或被设计为在不使用任何外部资源的情况下处理或执行训练步骤。GPU 210可以包括足够的存储器容量和处理能力来执行训练步骤。下面将结合图3至图6进一步详细描述由训练节点208或对应的GPU 210执行的过程。

[0081] 现在参照图3,根据实施例,用于利用视频数据的机器学习(ML)模型的加速训练的方法300的流程图。简而言之,方法300可以包括接收视频序列的比特流和指示视频序列的一个或多个图像帧的一个或多个指示,以用于训练机器学习(ML)模型(步骤302),并且通过解析比特流来确定视频序列的图像帧的时间戳、在比特流内的位置和类型(步骤304)。方法300可以包括使用一个或多个指示以及视频序列的图像帧的时间戳、在比特流内的位置和类型来确定比特流的一个或多个片段以用于进行解码以提取一个或多个图像帧(步骤306)。针对该一个或多个图像帧中的图像帧,对应的片段可以表示视频序列的图像帧的对应参考链。方法300可以包括解码比特流的一个或多个片段(步骤308),并且使用一个或多个图像帧来训练ML模型(步骤310)。

[0082] 方法300可以由GPU 210完全实现、执行或实行。GPU 210可以在不使用任何外部存储器或处理资源的情况下执行步骤302至步骤310。具有由单个GPU 210完全执行的方法300引起对(多个)ML模型的加速训练。具体地,通过在单个GPU 210内完全执行方法300,可以通过避免GPU 210与任何外部资源之间的数据交换来减少处理时间。例如,在GPU 210外部解码视频数据或存储解码的视频数据可以引入与GPU 210与任何外部资源之间的压缩和/或解码的视频数据的交换相关联的延迟。

[0083] 方法300可以包括GPU 210接收视频序列的比特流和指示视频序列的一个或多个图像帧的一个或多个指示,以用于训练机器学习(ML)模型(步骤302)。在一些实现中,GPU 210或处理电路系统214可以在存储器212内分配存储器区域以用于执行训练步骤。例如,在接收比特流和一个或多个指示之前,处理电路系统214可以分配存储器区域以存储用于下一训练步骤的数据,诸如从存储系统204接收的压缩的视频数据、解码的视频或图像帧、从视频帧中提取的特征和/或其他数据。在一些实现中,处理电路系统214可以在每个训练步骤开始时分配存储器区域,或者可以在训练会话开始时分配存储器区域,其中分配的存储器区域可以被用于连续的训练步骤。在一些实现中,处理电路系统214可以覆写分配的存储器区域(或存储器212)中存储不再需要的数据的片段,以有效利用存储器212。

[0084] GPU 210可以经由加载器208从存储系统204接收一个或多个视频序列的一个或多个比特流。例如, GPU 210可以接收(例如, 由自体140的相机170m同时捕获的)多个压缩的视频序列的多个比特流。压缩的视频序列可以在自体140处被编码, 并且可以不被同步。例如, 每个相机170m可以具有单独的时间线(根据该时间线捕获图像帧)和用于对捕获的图像帧进行编码的单独的编码器。用于不同相机170m的图像捕获时间实例可以不是时间对齐的。此外, 与不同相机170m相关联的编码器可以不被同步。如此, 当由单独的编码器以不同的比特流对由不同的相机170m在同一时刻(或基本在同一时刻, 考虑到用于由不同的相机170m捕获图像帧的时间线的任何差异)所捕获的图像帧进行编码时, 所捕获的图像帧可以具有不同的时间戳。

[0085] GPU 210可以接收一个或多个指示, 该一个或多个指示指示从一个或多个视频序列中选择或者将被选择以用于训练ML模型的图像帧。一个或多个指示符可以由训练系统202的用户指定, 并且由GPU 210作为输入(例如, 经由数据加载器208)进行接收。一个或多个指示可以包括一个或多个时间值。每个时间值可以指示每个接收的比特流中的单独图像帧。例如, 如果由自体140的八个不同相机170m捕获的八个视频序列的八个比特流由GPU 210接收, 则每个时间值将指示八个图像帧, 例如每个视频序列中的图像帧。时间值可以指示选择的或要被选择的图像帧的时间戳, 但是不一定完全等于这些时间戳。GPU 210可以将接收到的(多个)比特流和一个或多个指示存储在分配的存储器区域中。

[0086] 方法300可以包括GPU 210通过解析比特流来确定视频序列的图像帧的时间戳、在比特流内的位置和类型(步骤304)。每个视频比特流可以包括跨比特流分布的多个标头。视频比特流可以包括比特流中的用于每个压缩图像帧的单独标头。每个图像帧标头可以紧接在图像帧的压缩数据之前, 并且可以包括指示关于图像帧和对应压缩数据的信息的数据。标头可以包括图像帧的时间戳、图像帧的类型、压缩图像帧数据的大小、压缩图像帧数据在比特流中的位置和/或其他数据。

[0087] 图像帧的类型可以包括帧内帧(I-帧)类型或预测帧(P帧)。使用来自另一个先前编码的图像帧的数据对P帧进行编码。为了解码P帧, 解码器将在解码P帧之前解码P帧所依赖的任何其他图像帧。I-帧也被称为参考帧, 并且可以独立于任何其他图像帧地被解码。针对图像帧, 对应的时间戳表示图像帧的呈现的时间, 例如相对于视频序列中的第一图像帧。压缩图像帧数据在比特流中的位置可以是偏移值, 例如以字节为单位, 其指示压缩图像帧数据在比特流中在何处开始。

[0088] 处理电路系统214可以解析每个接收到的比特流以标识或确定比特流中的每个压缩图像帧的标头。处理电路系统214可以读取每个比特流中的标头以确定每个比特流中的每个图像帧的时间戳、压缩图像帧数据的位置和类型。与解码单个比特流或其部分相比, 解析(多个)比特流在处理能力和处理时间方面要轻松得多。通过解析(多个)比特流而确定的图像帧的时间戳、在比特流内的位置和类型实现了要被解码以提取(多个)选定图像帧的视频数据量的显著减少, 并且因此显著加速训练过程。

[0089] 在确定(多个)视频序列的图像帧的时间戳、在比特流中的位置和类型时, GPU 210或处理电路系统214可以生成一个或多个数据结构来存储(多个)视频序列的图像帧的时间戳、在比特流中的位置和类型。一个或多个数据结构可以包括表、数据文件和/或一些其他类型的数据结构。例如, 处理电路系统214可以针对每个比特流生成分离的表, 类似于下面

的表1。表1的最左边第一列可以包括视频序列中的所有图像帧的时间戳,例如按升序顺序,第二列可以包括每个图像帧的压缩帧数据的位置(或偏移),并且最右列可以包括每个图像帧的类型。表1的每行对应于视频序列中的单独图像帧。

[0090]	时间戳	位置/偏移	类型
	T_1	O_1	I
	T_2	O_2	P
	T_3	O_3	P
	T_4	O_4	P
	T_5	O_5	I
[0091]	.	.	.
	.	.	.
	T_n	O_n	P

[0092] 表1

[0093] 在一些实现中,一个或多个数据结构可以包括用于I-帧的第一数据结构和用于视频序列中的所有帧的第二数据结构。第一和第二数据结构中的每个数据结构都可以是表或数据文件。第一数据结构的示例可以是下面的表2。表2的每行表示视频序列中的单独I-帧。第一列(例如,最左边)可以包括I-帧的时间戳(例如,按升序顺序),并且第二列可以包括对应的位置或偏移(例如,以字节为单位)。表2中的数据允许快速确定比特流中的用于任何I-帧的压缩数据的位置。

[0094]	时间戳	位置/偏移
	$T_{1,I}$	$O_{1,I}$
	$T_{2,I}$	$O_{2,I}$
	$T_{3,I}$	$O_{3,I}$
	.	.
	.	.
[0095]	.	.
	$T_{n,I}$	$O_{n,I}$

[0095] 表2

[0096] 第一数据结构的示例可以是下面的表3。表3的每行表示视频序列中的单独的图像

帧。第一列(例如,最左边)可以包括图像帧的时间戳(例如,按升序顺序),并且第二列可以包括对应的位置或偏移(例如,以字节为单位)。表3中的数据允许确定比特流中的用于任何图像帧的压缩数据的位置。

[0097]	时间戳	位置/偏移
	T ₁	O ₁
	T ₂	O ₂
	T ₃	O ₃
[0098]	.	.
	.	.
	.	.
	T _n	O _n

[0099] 表3

[0100] 上面表中的任何表的数据都可以被存储在数据文件中。与表1相比,表2和表3可以不包括图像帧类型的指示。相反,可以从特定于I-帧的表2中标识I-帧。在一些实现中,可以生成其他数据结构(例如,代替表1、表2和/或表3中的任何表或与其相组合)。一旦被生成,GPU 210或处理电路系统214可以将一个或多个数据结构存储在存储器212中。一个或多个数据结构可以指示:(i) 针对视频序列的每个图像帧,指示在比特流中的图像帧的压缩数据的对应位置的对应时间戳和对应偏移,以及(ii) 视频序列中的特定类型的图像帧。

[0101] 方法300可以包括GPU 210使用一个或多个指示以及视频序列的图像帧的时间戳、在比特流内的位置和类型来确定比特流的一个或多个片段以用于进行解码以提取一个或多个图像帧(步骤306)。一个或多个指示可以包括一个或多个时间值以用于使用来标识或确定用于使用来训练ML模型的选定图像帧。每个时间值可以指示接收到的一个或多个接收比特流中的一个或多个对应图像帧或(多个)对应时间戳。如果对应于多个相机170m的多个比特流被接收,则每个指示符或时间值可以指示接收比特流中的每个比特流中的单独图像帧或对应时间戳。

[0102] 针对一个或多个时间值中的每个时间值,GPU 210或处理电路系统214可以针对每个接收比特流确定最接近时间值的对应时间戳。例如,处理电路系统214可以使用针对给定比特流的表3(或表1)来确定该比特流的最接近该时间值的时间戳。处理电路系统214可以通过使用对应的数据结构(例如,表3或表1)来针对每个比特流确定最接近时间值的单独时间戳。每个确定的时间戳指示经由时间值而指示或选择的对应比特流中的相应图像帧。例如,如果对应于八个相机170m的八个比特流被接收,则处理电路系统214可以针对每个时间值(或指示符)确定指示八个选定图像帧的八个对应时间戳,其中每个比特流中有一个选定图像帧。如果GPU 210接收到三个指示符和对应于八个相机170m的八个比特流,则处理电路系统214可以确定总共24个时间戳,这些时间戳对应于24个选定图像帧,其中针对每个比特流确定对应于三个选定图像帧的三个时间戳。

[0103] 处理电路系统214可以针对每个时间值(或每个指示符)确定针对每个接收比特流的单独的第二时间戳。针对每个比特流,第二时间戳指示比特流的对应I-帧。针对每个时间值,处理电路系统214可以将针对给定比特流的第二时间戳确定为最接近该时间值(或该时间值指示的比特流的选定帧的时间戳)的该比特流的I-帧时间戳,该I-帧时间戳小于或等于该时间值(或小于或等于由该时间值指示的该比特流的选定帧的时间戳)。例如,处理电路系统214可以使用比特流的表2(或表1)来确定最接近给定时间值(或指示符)的并且小于或等于该时间值(或指示符)的该比特流的I-帧时间戳。

[0104] 针对给定的时间值(或指示符)和给定的比特流,如果选定图像帧的时间戳和对应的I-帧时间戳相等,则意味着选定图像帧(或由该时间值指示的图像帧)是I-帧。然而,如果两个时间戳不同,则处理电路系统214可以确定选定图像帧(或由该时间值指示的图像帧)不是I-帧。

[0105] 针对每个时间值(或指示符)和每个接收比特流,处理电路系统214可以使用I-帧时间戳和一个或多个数据结构来确定比特流中的I-帧的位置或偏移(例如,起始位置)。例如,处理电路系统214可以使用表2(或表1)来确定与所确定的I-帧时间戳相对应的位置或偏移。例如,如果所确定的I-帧时间戳为 $T_{3,I}$,则对应的偏移为 $0_{3,I}$ 。

[0106] 处理电路系统214可以针对每个时间值(或指示符)和每个接收比特流,使用对应的选定图像帧的时间戳来确定对应的选定图像帧的结束位置。处理电路系统214可以将比特流中的选定图像帧的结束位置确定为比特流中的下一(或之后)图像帧的起始位置。处理电路系统214可以将比特流中的选定图像帧的结束位置确定为选定图像帧的起始位置加上比特流的压缩数据的大小。处理电路系统214可以通过解析比特流或帧标头并且将大小记录在一个或多个数据结构中来确定比特流中的每个图像帧的大小。

[0107] 处理电路系统214可以针对每个时间值(或指示符)和每个接收比特流,确定在对应I-帧的起始位置与对应选定图像帧的结束位置之间延伸的比特流的对应片段。所确定的片段表示要被解码以解码选定图像帧的压缩视频数据的最小量。如果对应于不同时间值(或不同指示符)但是来自同一比特流的两个或多个片段重叠,则处理电路系统214可以考虑最长的片段用于解码并且忽略较短的(多个)片段。

[0108] 需要注意的是,虽然上面将指示符描述为时间值,但是其他实现也是可能的。例如,由GPU 210接收的指示符可以是图像帧的索引。

[0109] 现在参照图4,根据实施例,示出了描绘视频序列400中的一组选定图像帧和要被解码的对应图像帧的图。使用三个不同的指示符或时间值,处理电路系统214可以将图像帧402、406和412确定(例如,如上面所讨论的)为选定图像帧或被指示要由指示符或时间值选择的图像帧。图像帧402是I-帧,而图像帧406和412是P帧。关于选定图像帧402,处理电路系统214确定仅图像帧402要被解码,因为它是I-帧并且不依赖于任何其他图像帧。针对选定图像帧406(其是P帧),处理电路系统214确定最接近的前一I-帧是图像帧404,并且帧404和406两者都要被解码,以便获得解码形式的选定图像帧406。针对选定图像帧412(其是P帧),处理电路系统214确定最接近的前一I-帧是图像帧408,并且帧408、410和412要被解码,以便获得解码形式的选定图像帧412。

[0110] 图5图示了根据实施例的与图4的视频序列400相对应的比特流500、与选定图像帧相对应的压缩数据以及与要被解码的帧相对应的压缩数据的图。压缩数据片段502表示图

像帧402的压缩数据。压缩数据片段504表示图像帧404和406的压缩数据。压缩数据片段506表示图像帧408、410和412的压缩数据。处理电路系统214可以只将压缩数据片段502、504和506馈送到(多个)视频解码器216以用于解码,而不是解码整个比特流500。

[0111] 针对每个选定图像帧(或每个被指示用于选择的图像帧),要被解码的对应压缩数据片段可以被视为表示图像帧的对应参考链。参考链以选定帧之前最接近的I-帧开始,并且以选定图像帧结束。参考链表示具有依赖性(帧参考)的相互依赖的图像帧链,该依赖性从选定图像帧处开始并且一直向后到第一遇到的I-帧。例如,在由图像帧408、410和412形成的参考链中,图像帧412参考图像帧410,并且后者参考图像帧410,其是I-帧。

[0112] 再次参照图3,方法300可以包括GPU 210解码比特流的一个或多个片段(步骤308)。处理电路系统214可以将确定的压缩数据片段提供或馈送到(多个)视频解码器216以用于解码。通过仅对解码选定图像帧所需的压缩数据进行解码,GPU 210显著减少了解码选定图像帧所消耗的处理时间和处理能力(例如,与解码整个比特流相比)。处理电路系统214可以将解码的视频数据存储在存储器212中。针对GPU 210中的存储器资源的有效使用,视频处理电路系统214可以覆写不再需要的解码的视频数据。例如,当解码压缩数据片段506时,并且一旦图像帧410被解码,处理电路系统214可以确定不再需要解码的图像帧410的数据,并且删除解码的图像帧410以释放存储器空间。一旦对应于选定图像帧的压缩数据片段被解码,只有针对选定图像帧的解码的数据可以被保持在存储器212或分配的存储器区域中,而其他解码的图像帧(非选定图像帧)可以被删除以释放存储器空间。

[0113] 如上面所讨论的,GPU 210可以包括可以并行操作的多个视频解码器。GPU 210可以以各种方式执行并行视频解码。例如,处理电路系统214可以将不同的压缩数据片段(无论对应的比特流如何)指派给不同的视频解码器216,以保持所有视频解码器216持续忙碌并且加快对片段的视频解码。在一些实现中,处理电路系统214可以将不同的比特流(或其压缩数据片段)指派给不同的视频解码器216。在一些实现中,GPU 210可以同时接收针对多个会话的视频比特流(例如,由不同的自体140捕获的比特流或以不同的时间间隔捕获的比特流)。处理电路系统214可以指派不同的视频解码器216来解码不同会话的视频数据。

[0114] 方法300可以包括GPU 210使用一个或多个图像帧来训练ML模型(步骤310)。一旦选定图像帧被解码,处理电路系统214可以从每个选定图像帧中提取一个或多个特征,并且将提取的特征馈送到被配置为训练ML模型的训练模块。作为响应,训练模块可以修改或更新ML模型的一个或多个参数。

[0115] 虽然方法300在上面被描述为由GPU 210执行或实行,但是一般地,方法300可以由包括存储器和一个或多个处理器的任何计算系统执行或实行。此外,可以使用另一类型的处理器来代替GPU 210。

[0116] 结合本文中公开的实施例而描述的各种说明性逻辑块、模块、电路和算法步骤可以被实现为电子硬件、计算机软件或这两者的组合。为了清楚地说明硬件和软件的这种可互换性,已经在其功能方面总体描述了各种说明性组件、块、模块、电路和步骤。这种功能是被实现为硬件还是软件取决于特定应用和施加在整个系统上的设计约束。本领域技术人员可以针对每个特定应用以不同方式实现所描述的功能,但这种实现决策不应当被解释为导致偏离本公开或权利要求的范围。

[0117] 在计算机软件著作权中实现的实施例可以在软件、固件、中间件、微码、硬件描述语言或

其任何组合中被实现。代码段或机器可执行指令可以表示过程、函数、子程序、程序、例程、子例程、模块、软件包、类,或指令、数据结构或程序语句的任何组合。通过传递和/或接收信息、数据、自变量、参数或存储器内容,代码段可以被耦合到另一代码段或硬件电路。信息、自变量、参数、数据等可以经由任何适合的手段(包括内存共享、消息传递、令牌传递、网络传输等)来被传递、转发或传输。

[0118] 被用于实现这些系统和方法的实际软件代码或专用控制硬件并不限制所要求保护的征或本公开。因此,在不参考特定软件代码的情况下描述系统和方法的操作和行为,可以理解,软件和控制硬件可以被设计为基于本文中的描述来实现系统和方法。

[0119] 当在软件中被实现时,功能可以作为一个或多个指令或代码被存储在非暂态的计算机可读或处理器可读的存储介质上。本文中公开的方法或算法的步骤可以体现在处理器可执行软件模块中,该处理器可执行软件模块可以驻留在计算机可读或处理器可读存储介质上。非暂态计算机可读或处理器可读介质包括计算机存储介质和有形存储介质两者,有形存储介质促进将计算机程序从一个地方传送到另一地方。非暂态的处理器可读存储介质可以是由计算机可以访问的任何可用介质。作为示例而非限制,这种非暂态的处理器可读介质可以包括RAM、ROM、EEPROM、CD-ROM或其他光盘存储、磁盘存储或其他磁存储设备、或者任何其他有形存储介质,该任何其他有形存储介质可以被用于以指令或数据结构的形式存储期望的程序代码并且可以由计算机或处理器访问。本文中使用的磁盘和光盘包括光碟(CD)、激光光盘、光盘、数字多功能光盘(DVD)、蓝光盘和软盘,其中“磁盘”通常磁性地再现数据,而“光盘”则用激光光学地再现数据。上面各项的组合也应当被包括在计算机可读介质的范围内。附加地,方法或算法的操作可以作为代码和/或指令的一个或任何组合或集合驻留在非暂态的处理器可读介质和/或计算机可读介质上,该非暂态的处理器可读介质和/或计算机可读介质可以并入计算机程序产品中。

[0120] 提供所公开的实施例的前述描述是为了使得本领域技术人员能够制造或使用本文中描述的实施例和其变型。对这些实施例的各种修改对于本领域技术人员来说将是很明显的,并且本文中定义的原理可以被应用于其他实施例,而不脱离本文中公开的主题内容的精神或范围。因此,本公开并不旨在限于本文所示的实施例,而是应当被赋予与以下权利要求以及本文中公开的原理和新颖特征一致的最宽范围。

[0121] 虽然已经公开了各种方面和实施例,但也可以考虑其他方面和实施例。所公开的各个方面和实施例是出于说明的目的,并不旨在进行限制,真实的范围和精神由以下权利要求书指示。

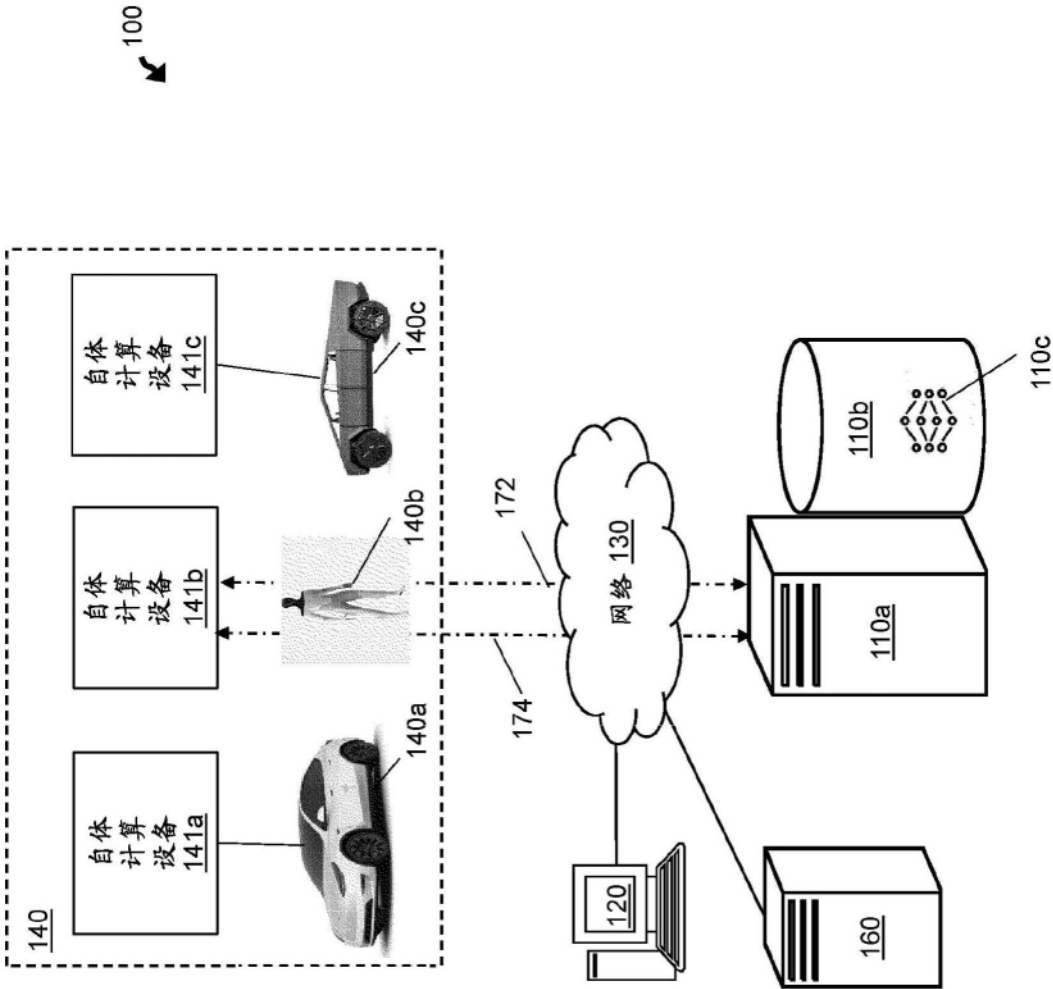


图1A

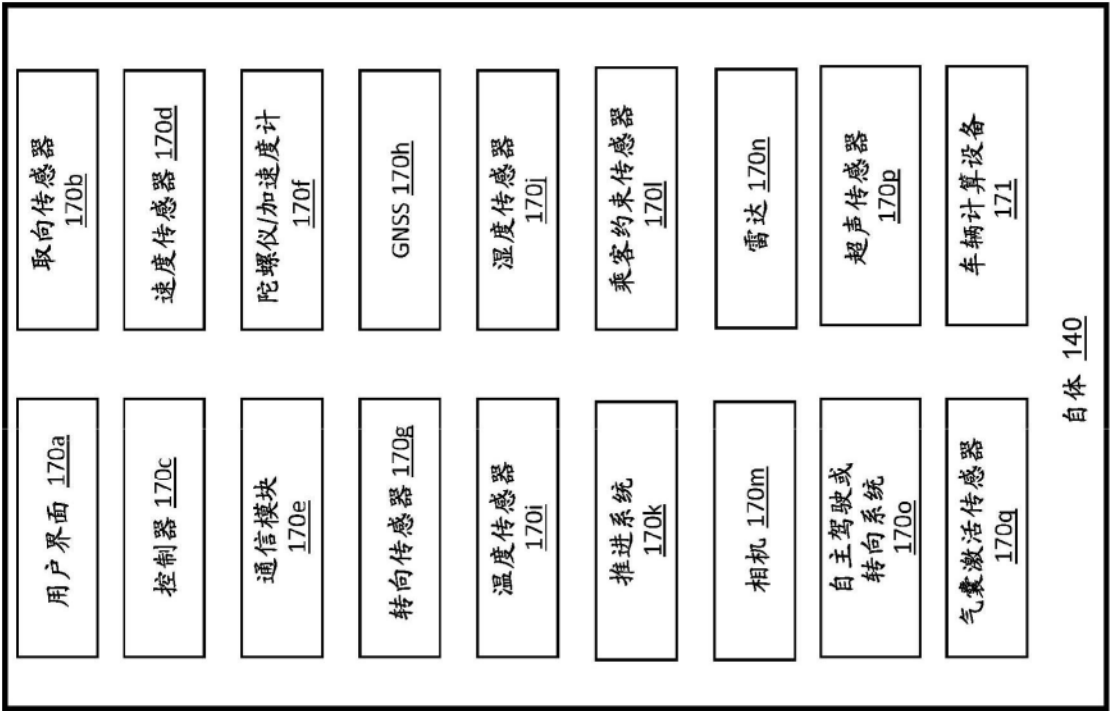


图1B

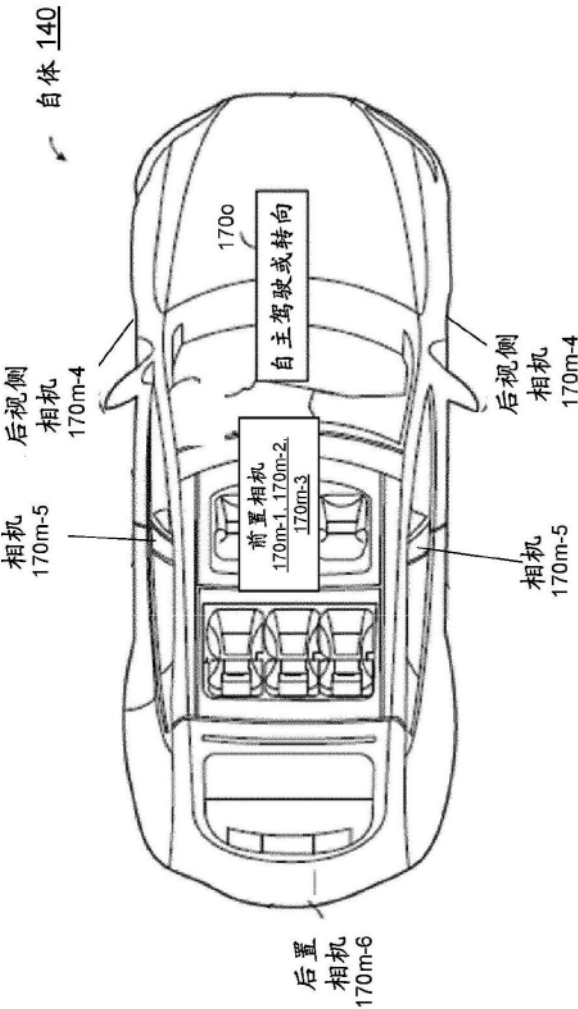


图1C

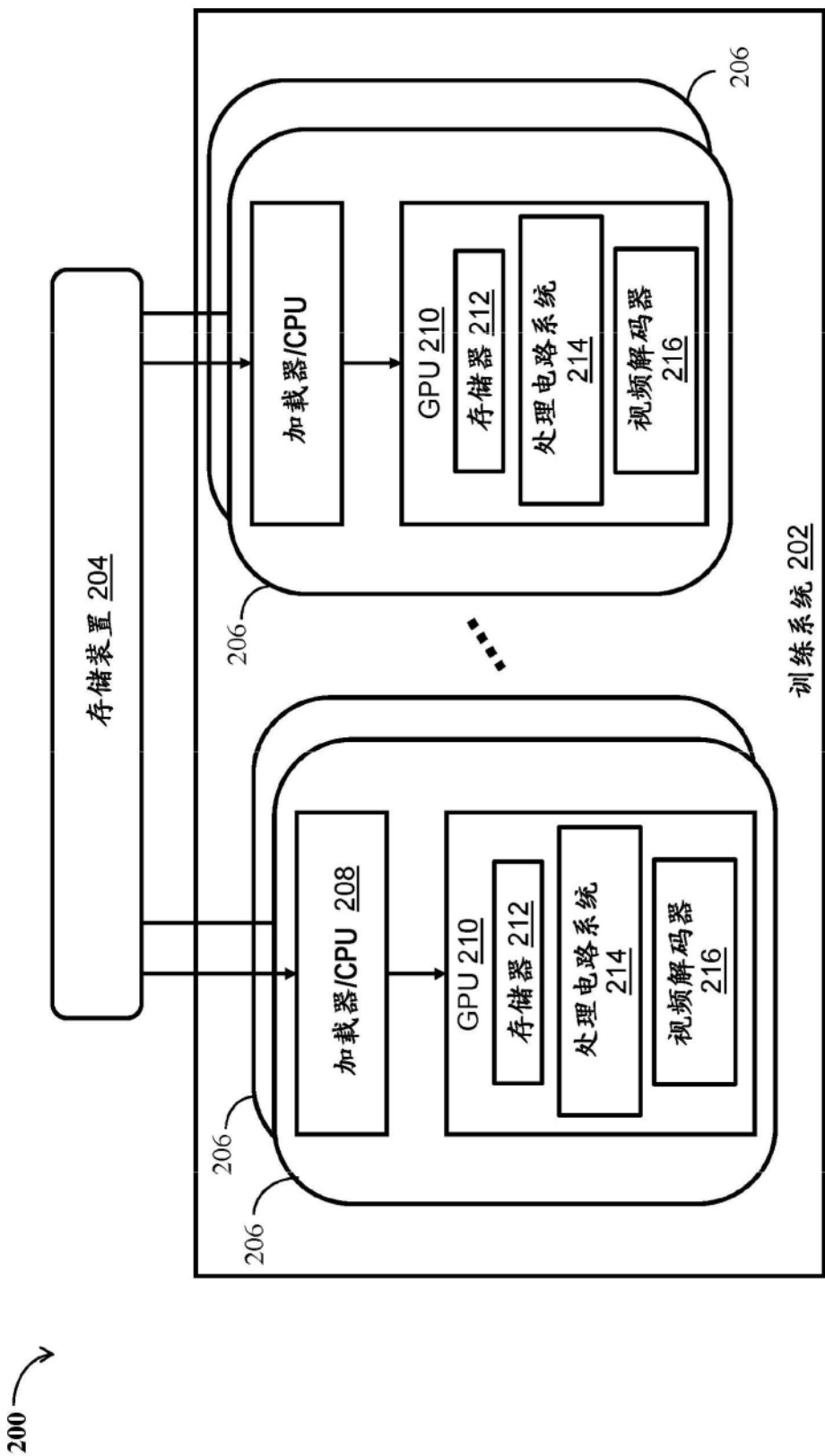


图2

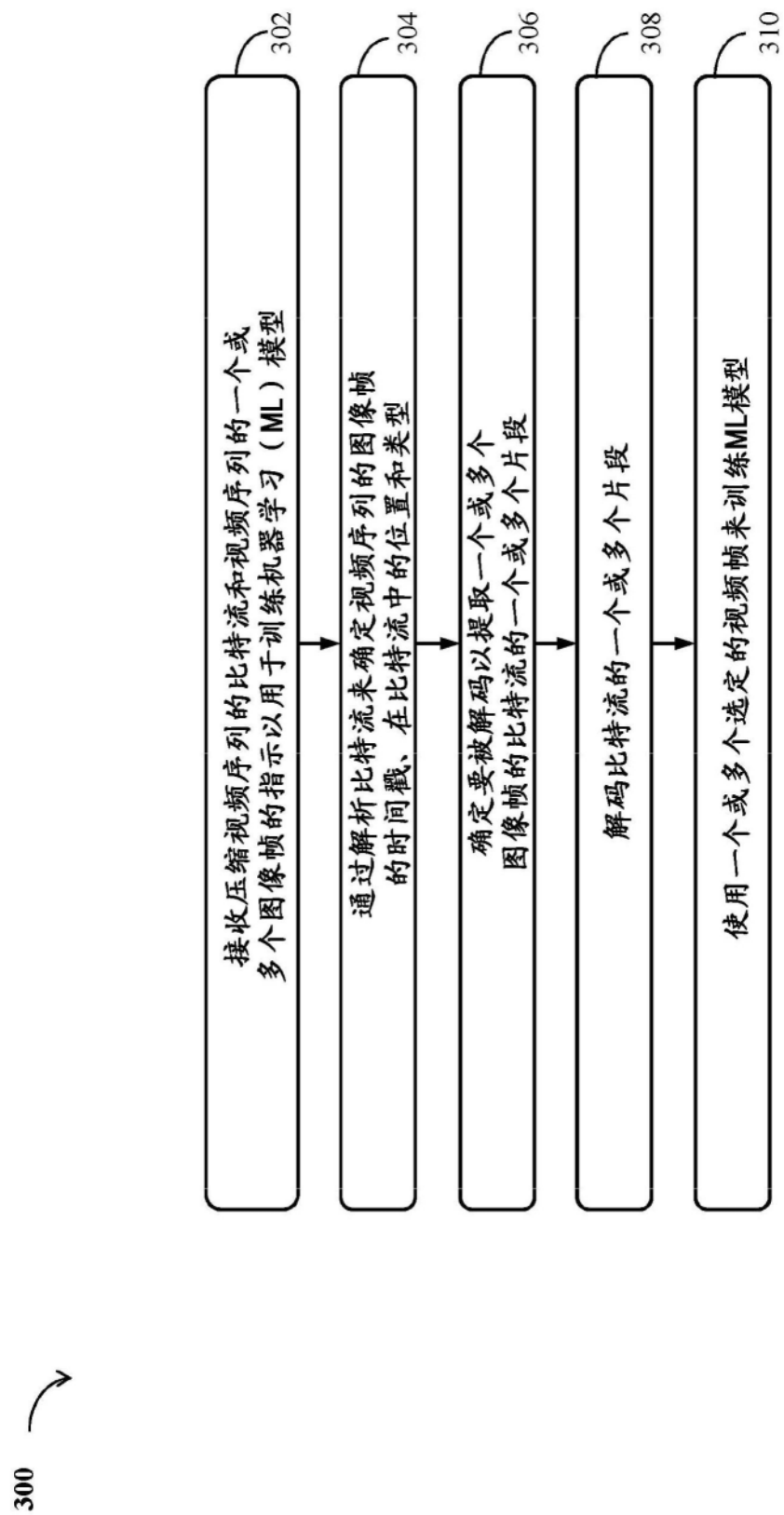


图3

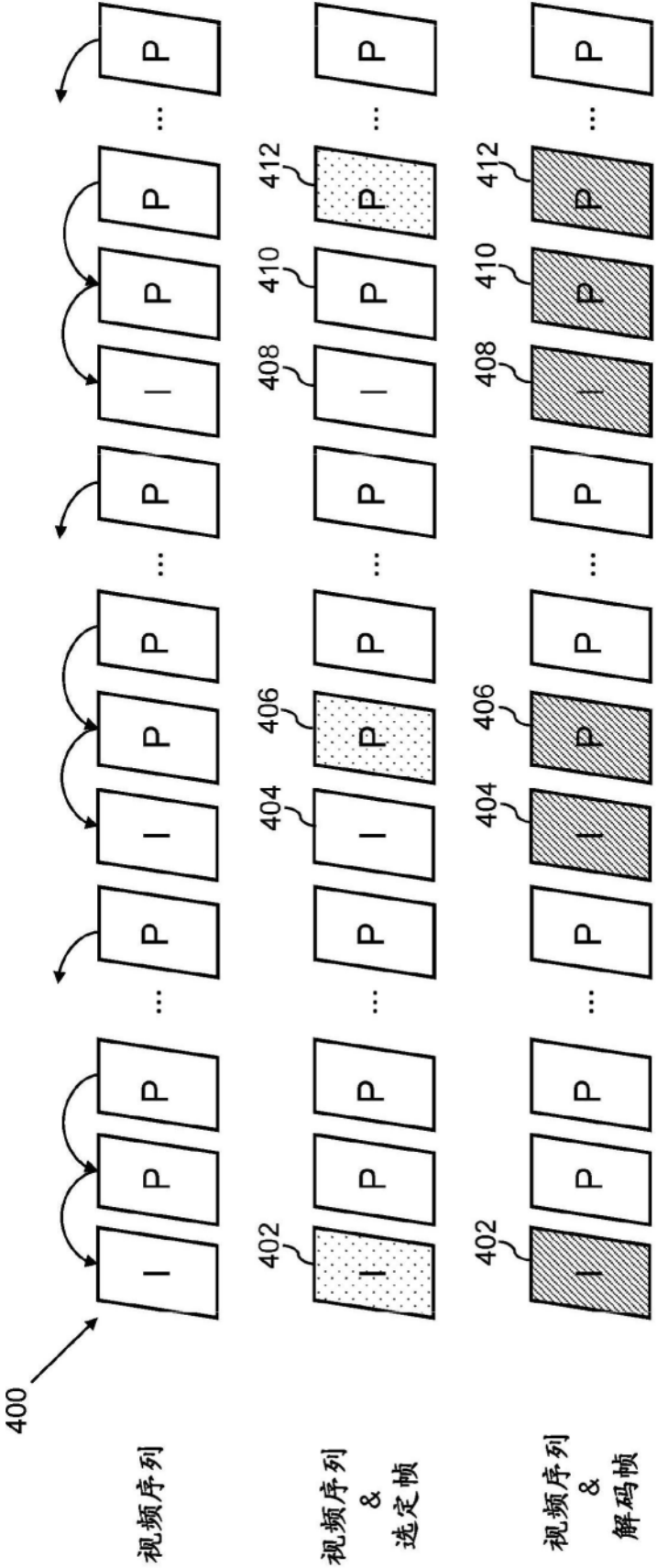


图4

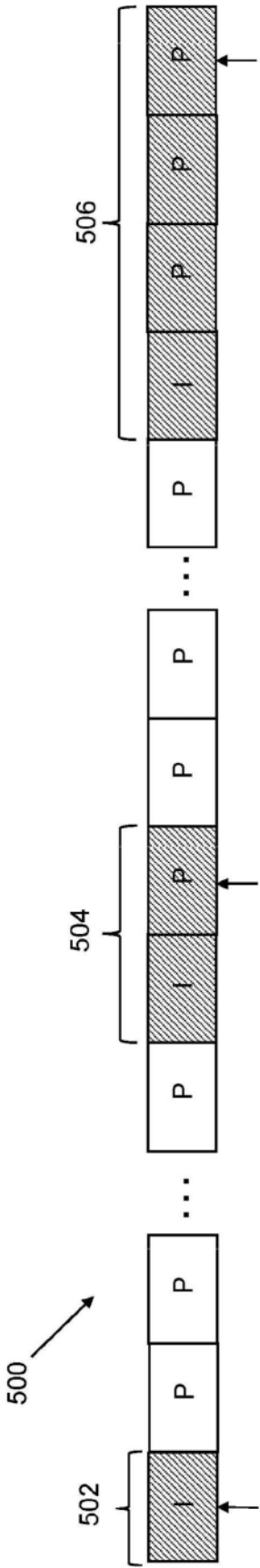


图5