



(12) 发明专利申请

(10) 申请公布号 CN 112424769 A

(43) 申请公布日 2021.02.26

(21) 申请号 201980046711.3

(22) 申请日 2019.01.10

(30) 优先权数据

62/781,276 2018.12.18 US

(85) PCT国际申请进入国家阶段日

2021.01.12

(86) PCT国际申请的申请数据

PCT/US2019/013024 2019.01.10

(87) PCT国际申请的公布数据

WO2020/131140 EN 2020.06.25

(71) 申请人 谷歌有限责任公司

地址 美国加利福尼亚州

(72) 发明人 A.戈班 Y.吴

(74) 专利代理机构 北京市柳沈律师事务所
11105

代理人 金玉洁

(51) Int.Cl.

G06F 16/783 (2006.01)

G06K 9/00 (2006.01)

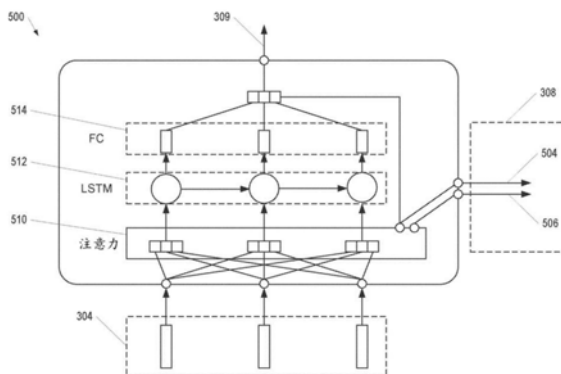
权利要求书3页 说明书16页 附图12页

(54) 发明名称

用于地理位置预测的系统和方法

(57) 摘要

在一个示例实施例中,用于从影像中提取信息的计算机实施的方法包括获得表示图像序列的数据,图像序列中的至少一个图像描绘对象。该方法包括将图像序列输入到被训练成从图像序列中提取位置信息的机器学习信息提取模型。该方法包括响应于输入图像序列,获得作为信息提取模型的输出的表示与在图像序列中描绘的对象相关联的真实世界位置的数据。



1. 一种用于从影像中提取信息的计算机实施的方法,所述方法包括:

在包括一个或多个处理器的计算系统处,获得表示图像序列的数据,所述图像序列中的至少一个图像描绘对象;

由所述计算系统将所述图像序列输入到被训练成从所述图像序列中提取位置信息的机器学习信息提取模型中;以及

由所述计算系统获得作为信息提取模型响应于输入所述图像序列的输出的、表示与所述图像序列中描绘的所述对象相关联的真实世界位置的数据。

2. 根据权利要求1所述的计算机实施的方法,其中获得表示与所述图像序列中描绘的所述对象相关联的所述真实世界位置的所述数据包括:

由所述计算系统确定表示与图像帧序列中描绘的所述对象相关联的类的的数据;

由所述计算系统至少部分地基于表示从所述图像序列中提取的图像特征序列的数据,来确定与所述图像序列相关联的时间注意力值和空间注意力值;以及

由所述计算系统至少部分地基于所述图像特征序列、时间注意力值和空间注意力值来预测与所述对象相关联的所述真实世界位置。

3. 根据权利要求2所述的计算机实施的方法,其中确定与所述图像序列相关联的所述时间注意力值和所述空间注意力值包括:

由所述计算系统将所述图像特征序列输入到弱监督对象分类模型中,其中对象分类模型包括至少一个长短期记忆块;以及

由所述计算系统获得作为所述对象分类模型响应于输入所述图像特征序列的输出的所述时间注意力值和所述空间注意力值。

4. 根据权利要求2或3所述的计算机实施的方法,其中确定与所述图像帧序列中描绘的所述对象相关联的所述分类包括:

由所述计算系统将所述图像特征序列输入到弱监督对象分类模型中;以及

由所述计算系统获得作为对象分类模型响应于输入所述图像特征序列的输出的与所述对象相关联的所述分类。

5. 根据权利要求4所述的计算机实施的方法,其中表示所述图像帧序列的所述数据包括与所述图像帧序列相关联的至少一个分类标签,并且所述方法还包括:

由所述计算系统至少部分地基于与所述图像序列相关联的所述至少一个分类标签,来确定与由所述对象分类模型输出的所述分类相关联的损失;以及

由所述计算系统至少部分地基于所确定的损失来训练所述对象分类模型。

6. 根据权利要求2、3、4或5所述的计算机实施的方法,其中预测与所述对象相关联的所述真实世界位置包括:

由所述计算系统将所述图像特征序列、所述时间注意力值和所述空间注意力值输入到帧级位置特征提取模型中;

由所述计算系统获得作为所述帧级位置特征提取模型响应于输入所述图像特征序列、所述时间注意力值和所述空间注意力值的输出的、表示包括与所述对象相关联的一个或多个位置特征的位置特征序列的数据;

由所述计算系统将所述位置特征序列输入到帧级位置预测模型中;

由所述计算系统获得作为所述帧级位置预测模型响应于输入所述位置特征序列的输

出的、表示相机坐标空间中与所述对象相关联的坐标的数据；以及

由所述计算系统至少部分地基于所述相机坐标空间中的所述坐标和与所述对象相关联的相机姿态数据来确定与所述对象相关联的真实世界坐标。

7. 根据权利要求6所述的计算机实施的方法,还包括:

由所述计算系统至少部分地基于与跨描绘所述对象的所述图像序列中的多个图像的所述对象相关联的坐标之间的方差来确定位置一致性损失;以及

由所述计算系统至少部分地基于所述位置一致性损失来训练所述帧级位置预测模型。

8. 根据权利要求6或7所述的计算机实施的方法,还包括:

由所述计算系统至少部分地基于跨描绘所述对象的所述图像序列中的多个图像而确定的外观特征之间的方差来确定外观一致性损失;以及

由所述计算系统至少部分地基于所述外观一致性损失来训练所述帧级位置预测模型。

9. 根据权利要求6、7或8所述的计算机实施的方法,还包括:

由所述计算系统至少部分地基于所述相机坐标空间中与所述对象相关联的所述坐标和与跨描绘所述对象的所述图像序列中的多个图像的所述对象相关联的空间注意力来确定瞄准损失;以及

由所述计算系统至少部分地基于所述瞄准损失来训练所述帧级位置预测模型。

10. 根据权利要求6、7、8或9所述的计算机实施的方法,还包括:

由所述计算系统至少部分地基于与所述对象相关联的所述真实世界坐标和与用于捕获描绘所述对象的所述图像序列的相机相关联的视场来确定视场损失;以及

由所述计算系统至少部分地基于所述视场损失来训练所述帧级位置预测模型。

11. 根据任一前述权利要求所述的计算机实施的方法,其中所述图像序列描绘跨所述图像序列中的多个图像的多个对象,并且所述信息提取模型的所述输出包括表示与所述图像序列中描绘的所述多个对象相关联的真实世界位置的数据。

12. 一种计算机实施的方法,用于训练信息提取模型以确定表示与图像序列中描绘的对象相关联的真实世界位置的数据,所述信息提取模型包括:

图像特征提取模型;

弱监督对象分类模型;

地理位置预测模型;

所述方法包括在包括一个或多个处理器的计算系统处:

获得表示具有噪声分类的图像序列的数据,所述图像序列中的至少一个图像描绘所述对象;

由所述图像特征提取模型响应于输入所述图像序列而输出图像特征序列;

由对象分类模型响应于输入所述图像特征序列而输出包括与所述图像序列相关联的一个或多个分类标签的分类数据,其中所述分类数据至少部分地基于与所述图像特征序列相关联的一个或多个时间注意力值和一个或多个空间注意力值来确定,所述一个或多个时间注意力值和一个或多个空间注意力值由所述对象分类模型确定;

至少部分地基于与所述图像序列相关联的所述分类数据和所述噪声分类来训练所述对象分类模型;

由所述地理位置预测模型响应于输入所述图像特征序列、所述一个或多个时间注意力

值和所述一个或多个空间注意力值而输出与所述图像序列中描绘的所述对象相关联的真实世界位置;以及

至少使用所述图像特征序列、所述时间注意力值和所述空间注意力值来训练所述地理位置预测模型。

13. 一种用于从影像中提取信息的计算机实施的方法,所述方法包括:

在包括一个或多个处理器的计算系统处,获得表示一个或多个图像的数据,所述一个或多个图像中的至少一个图像描绘对象;

由所述计算系统将所述一个或多个图像输入到被训练成从所述一个或多个图像中提取位置信息的机器学习信息提取模型中;以及

由所述计算系统获得作为所述信息提取模型响应于输入所述一个或多个图像的输出的、表示与所述一个或多个图像中描绘的所述对象相关联的真实世界位置的数据。

14. 一种计算系统,所述系统包括:

一个或多个处理器;

一个或多个机器学习信息提取模型;和

计算机可读介质,具有存储在其上的指令,所述指令当由所述一个或多个处理器执行时使得所述系统执行根据权利要求1至13所述的方法。

15. 一个或多个有形的、非暂时性的计算机可读介质,存储一个或多个机器学习信息提取模型和计算机可读指令,所述计算机可读指令当由一个或多个处理器执行时使得所述一个或多个处理器执行根据权利要求1至13中任一项的操作。

用于地理位置预测的系统和方法

技术领域

[0001] 本公开一般涉及预测一个或多个对象的真实世界位置(例如,地理位置)。更具体地,本公开涉及一种信息提取模型,该信息提取模型可以以无监督的方式训练以预测(多个)对象的(多个)真实世界位置。

背景技术

[0002] 使用机器学习模型从影像(imagery)中提取数据的主要瓶颈之一是获得足够的数据来训练这种模型的高成本。获得任何对象的包括真实世界坐标的地面真实数据是非常耗时的过程。在大多数情况下,获得足够的地面真实数据来训练神经网络以预测真实世界位置是不切实际的。另一个问题是需要从图像中提取的各种类型的数据快速增长。

发明内容

[0003] 本公开的实施例的方面和优点将在以下描述中部分阐述,或者可以从描述中获知,或者可以通过实施例的实践获知。

[0004] 本公开的一个示例方面针对一种用于从影像中提取信息的计算机实施的方法。该方法包括在包括一个或多个处理器的计算系统处获得表示图像序列的数据,该图像序列中的至少一个图像描绘对象。该方法包括由计算系统将该图像序列输入到被训练成从该图像序列中提取位置信息的机器学习信息提取模型中。该方法包括由计算系统获得作为信息提取模型响应于输入该图像序列的输出的、表示与在图像序列中描绘的对象相关联的真实世界位置的数据。

[0005] 本公开的另一示例方面针对计算系统。该计算系统包括一个或多个处理器、一个或多个机器学习信息提取模型、以及一个或多个有形的、非暂时性的计算机可读介质,该一个或多个有形的、非暂时性的计算机可读介质共同存储指令,该指令当由所述一个或多个处理器执行时使得该计算系统执行操作。这些操作包括获得表示图像序列的数据,该图像序列中的至少一个图像描绘对象。这些操作包括将该图像序列输入到被训练成从该图像序列中提取位置信息的机器学习信息提取模型。这些操作包括获得作为信息提取模型响应于输入该图像序列的输出的、表示与在该图像序列中描绘的对象相关联的真实世界位置的数据。

[0006] 本公开的另一示例方面针对一个或多个有形的、非暂时性的计算机可读介质,存储了一个或多个机器学习信息提取模型和计算机可读指令,该计算机可读指令当由所述一个或多个处理器执行时使得所述一个或多个处理器执行操作。这些操作包括获得表示图像序列的数据,该图像序列中的至少一个图像描绘对象。这些操作包括将该图像序列输入到被训练成从该图像序列中提取位置信息的机器学习信息提取模型。这些操作包括获得作为信息提取模型响应于输入该图像序列的输出的、表示与在该图像序列中描绘的对象相关联的真实世界位置的数据。

[0007] 本公开的其他方面针对各种系统、装置、非暂时性计算机可读介质、用户界面和电

子设备。

[0008] 参考以下描述和所附权利要求,本公开的各种实施例的这些和其他特征、方面和优点将变得更好理解。并入本说明书并且构成其一部分的附图示出了本公开的示例实施例,并且与说明书一起用于解释相关原理。

附图说明

[0009] 参考附图,在说明书中阐述了针对本领域普通技术人员的实施例的详细讨论,其中:

[0010] 图1A-1C描绘了根据本公开的示例实施例的可以从图像序列中提取位置信息的示例计算系统/设备的框图;

[0011] 图2描绘了根据本公开的示例实施例的示例信息提取模型的框图;

[0012] 图3描绘了根据本公开的示例实施例的示例信息提取模型的框图;

[0013] 图4描绘了根据本公开的示例实施例的示例地理位置预测模型的框图;

[0014] 图5描绘了根据本公开的示例实施例的示例对象分类模型的框图;

[0015] 图6描绘了根据本公开的示例实施例的示例位置特征提取模型的框图;

[0016] 图7描绘了根据本公开的示例实施例的示例位置预测模型的框图;

[0017] 图8描绘了根据本公开的示例实施例的从图像序列中提取位置信息的示例方法的流程图;

[0018] 图9描绘了根据本公开的示例实施例的训练对象分类模型的示例方法的流程图;以及

[0019] 图10描绘了根据本公开的示例实施例的训练位置预测模型的示例方法的流程图。

[0020] 在多个附图中重复的附图标记旨在标识各种实施方式中的相同特征。

具体实施方式

[0021] 概观

[0022] 与本公开一致的系统和方法可以包括信息提取模型,该信息提取模型可以用于确定多个图像中描绘的一个或多个对象(例如,街道标志)的预测真实世界位置(例如,纬度和经度)。信息提取模型可以被训练成使用具有噪声分类的图像数据来预测所述一个或多个对象的真实世界位置。通过能够利用大量的有噪声的加标签的数据和非常大量的未加标签的数据的能力,本公开能够更快更便宜地开发更多应用的模型。本公开还能够实施多个新的应用,其中对于这些应用,没有足够的地面真实位置数据。附加地,由于信息提取模型可以在不需要地面真实位置数据(例如,作为回归的目标值)的情况下被训练,信息提取模型可以以无监督的方式被训练为端到端模型,以预测对象的真实世界位置,仅需要弱监督来对对象进行分类。大量实验已经证明,本公开的信息提取模型是有效的,并且能够达到与使用相同训练数据和测试数据集的传统完全监督模型相当的准确度水平,但是在模型训练时不需要地面真实标签、对象边界框和强监督。

[0023] 根据本公开的方面,包括一个或多个处理器的计算系统可以用于帮助实施包括信息提取模型的所公开技术的方面。在一些实施方式中,计算系统可以获得图像数据。图像数据可以包括多个图像,例如图像帧序列。图像帧序列可以描绘场景中的一个或多个对象(例

如,街道标志)。作为示例,图像帧序列中的一个或多个图像帧可以描绘一个或多个对象。作为另一示例,图像帧序列中的多个图像帧可以描绘相同对象。描绘相同对象的多个图像帧可以是图像帧序列中的连续图像帧或非连续图像帧。在一些实施方式中,图像帧序列可以从穿过街道的车辆的角度描绘接近街道的场景。在一些实施方式中,图像帧序列中的一个或多个图像帧可以对应于视频的一个或多个帧,或者其他类型的运动捕捉。

[0024] 在一些实施方式中,图像数据可以包括与图像帧序列相关联的分类。例如,图像数据可以包括与图像帧序列相关联的单个分类标签。可替代的,图像数据可以包括与图像帧序列中的一个或多个图像帧相关联的一个以上的分类标签。如下文进一步描述的,根据本公开的方面,系统可以使用信息提取模型来至少部分地基于具有噪声分类(例如,与图像帧序列相关联的单个分类标签)的图像帧序列来获得表示对象的真实世界位置的数据。

[0025] 在一些实施方式中,图像数据可以包括相机姿态数据。相机姿态数据可以表示用于捕获图像帧序列中的一个或多个图像帧的相机的真实世界位置和/或方位。例如,相机姿态数据可以包括4x4的、相机到世界的投影矩阵。

[0026] 根据本公开的方面,该系统可以生成地理位置数据。地理位置数据可以包括图像帧序列中描绘的一个或多个对象(例如,街道标志)的预测真实世界位置(例如,纬度和经度)。信息提取模型可以被配置为接收图像数据,并且响应于接收到图像数据而输出地理位置数据。

[0027] 该系统可以将图像数据输入到信息提取模型中,并且获得作为信息提取模型响应于输入该图像数据的输出的地理位置数据。该系统可以使用地理位置数据来例如识别对应于街道标志的路段(例如,在街道标志的纬度和经度处或附近的路段)。例如,图像数据可以包括描绘限速标志的图像帧序列。该系统可以将图像数据输入到信息提取模型中,并且获得作为信息提取模型的输出的包括限速标志的预测真实世界位置的地理位置数据。该系统可以使用地理位置数据来识别对应于该限速标志的路段(例如,在限速标志的预测真实世界坐标处或附近的路段)。

[0028] 根据本公开的方面,信息提取模型可以包括多个机器学习模型,诸如图像特征提取模型、对象分类模型和地理位置预测模型。在一些实施方式中,信息提取模型和/或包括在信息提取模型中的多个机器学习模型(例如,图像特征提取模型、对象分类模型、地理位置预测模型等)可以是或者可以以其他方式包括诸如神经网络(例如,深度神经网络)的各种机器学习模型或其他类型的机器学习模型,包括非线性模型和/或线性模型。神经网络可以包括前馈神经网络、递归神经网络(例如,长短期记忆递归神经网络)、卷积神经网络或其他形式的神经网络。

[0029] 在一些实施方式中,图像特征提取模型可以生成包括从图像帧序列中的一个或多个图像帧提取的一个或多个图像特征的图像特征数据。图像特征提取模型可以被配置为接收表示图像帧序列的数据(例如,图像数据),并且响应于接收到图像帧序列而输出图像特征数据。如下文进一步描述的,根据本公开的方面,图像特征数据中的一个或多个图像特征可以用于识别和/或分类图像帧序列中描绘的一个或多个对象。在一些实施方式中,图像特征数据可以包括图像特征序列。例如,图像特征数据中的一个或多个图像特征可以被组织成图像特征嵌入序列。图像特征嵌入序列中的每个图像特征嵌入可以对应于图像帧序列中的一图像帧,并且每个图像特征嵌入可以表示从对应的图像帧中提取的一个或多个图像特

征。在一些实施方式中,图像特征提取模型可以包括卷积神经网络(convolutional neural network,CNN),诸如Inception v2、任何SoTA图像分类网络(或其底部)等。

[0030] 系统可以将表示图像帧序列的数据(例如,图像数据)输入到图像特征提取模型中,并且获得作为图像特征提取模型响应于输入图像帧序列输出的图像特征序列(例如,图像特征数据)。如下文进一步描述的,系统可以使用信息提取模型以至少部分地基于图像特征数据来确定地理位置数据。

[0031] 在一些实施方式中,对象分类模型可以生成分类数据和注意力值数据。对象分类模型可以被配置为接收表示图像特征序列的数据(例如,由图像特征提取模型输出的图像特征数据),并且响应于接收到图像特征序列和相关联的图像特征嵌入,输出分类数据和注意力值数据。在一些实施方式中,对象分类模型可以包括弱监督的递归神经网络(recurrent neural network,RNN)。

[0032] 分类数据可以表示与图像帧序列中描绘的一个或多个对象相关联的分类。例如,对象分类模型可以至少部分地基于图像特征序列来识别所述一个或多个对象,并且确定与一个或多个识别到的对象相关联的一个或多个分类标签。可以在图像帧序列中的一些或所有图像帧中描绘所述一个或多个对象。作为另一示例,对象分类模型可以接收表示从图像数据中提取的图像特征序列的数据,其中该图像数据包括描绘限速标志的图像帧序列。响应于接收到图像特征序列,对象分类模型可以输出分类数据,该分类数据包括指示与限速标志相对应的限速值的分类标签。

[0033] 注意力值数据可以表示例如经分类的对象出现在特定帧中的特定像素处的概率。注意力值数据可以包括与图像特征序列相关联的一个或多个时间注意力值和一个或多个空间注意力值。作为示例,图像特征数据可以包括表示图像特征序列的图像特征嵌入序列。对象分类模型可以确定图像特征嵌入序列中的每个图像特征嵌入的时间注意力值和空间注意力值。每个图像特征嵌入的时间注意力值和空间注意力值可以表示经分类的对象出现在与图像特征嵌入相对应的图像帧中的特定像素处的概率。附加地或可替代地,对象分类模型可以确定图像特征序列(例如,图像特征嵌入序列)的单个时间注意力值和单个空间注意力值。

[0034] 该系统可以将表示图像特征序列的数据(例如,图像特征数据)输入到对象分类模型中,并且获得作为该对象分类模型响应于输入图像特征序列的输出的分类数据和注意力值数据。如下文进一步描述的,系统可以使用信息提取模型以至少部分地基于分类数据和注意力值数据来确定地理位置数据。

[0035] 在一些实施方式中,对象分类模型可以包括具有空间-时间注意力机制的长短期记忆(long-short term memory,LSTM)。空间-时间注意力机制可以用于确定注意力值数据,并且使得仅使用弱监督就能有效地训练对象分类模型。例如,对象分类模型可以包括多个LSTM块,每个块基于输入到对象分类模型中的图像特征序列(例如,图像特征数据)来输出每帧(per-frame)嵌入。对象分类模型可以使用时间注意力来加权由LSTM块产生的每帧嵌入,以便确定对象分类模型的输出。以这种方式,来自该输出的梯度一起在相同时间步在每个LSTM块之间成比例地分布到时间注意力的对应权重。

[0036] 在一些实施方式中,可以至少部分地基于与由对象分类模型输出的分类数据相关联的损失来训练对象分类模型。例如,系统可以至少部分地基于分类数据中的一个或多个

分类标签和与图像数据中的图像帧序列相关联的分类(例如,与图像帧序列相关联的单个分类标签),来确定柔性最大值(softmax)交叉熵损失。系统可以使用所确定的柔性最大值交叉熵来训练对象分类模型。

[0037] 在一些实施方式中,信息提取模型可以包括可以生成地理位置数据的地理位置预测模型。地理位置预测模型可以被配置为接收表示图像特征序列的数据(例如,由图像特征提取模型输出的图像特征数据)、表示与用于捕获与图像特征序列相对应的图像帧序列的一个或多个相机相关联的位置和/或方位的数据(例如,图像数据中的相机姿态数据)和表示与图像特征序列相关联的注意力值的数据(例如,由对象分类模型生成的注意力值数据)。地理位置预测模型可以被配置为响应于接收到图像特征序列、相机位置和/或方位信息以及与图像特征序列相关联的注意力值,输出地理位置数据。该系统可以将图像特征数据、相机姿态数据和注意力值数据输入到地理位置预测模型中,并且,获得作为地理位置预测模型响应于输入图像特征数据、相机姿态数据和注意力值数据的输出的地理位置数据。

[0038] 在一些实施方式中,地理位置预测模型可以生成与一个或多个经分类的对象中的每一个经分类的对象相关联的单个嵌入向量。例如,单个嵌入向量可以编码来自图像特征序列的与相关联的经分类的对象相关的所有数据。地理位置预测模型可以使用单个嵌入向量来预测与相关联的经分类的对象相关联的真实世界位置。

[0039] 在一些实施方式中,地理位置预测模型可以包括帧级位置特征提取模型和帧级位置预测模型。帧级位置特征提取模型可以被配置为接收表示图像特征序列的数据(例如,由图像特征提取模型输出的图像特征数据),并且响应于接收到该图像特征序列,输出包括与一个或多个经分类的对象相关联的一个或多个位置特征的位置特征数据。如下文将进一步描述的,根据本公开的方面,位置特征数据中的所述一个或多个位置特征可用于预测所述一个或多个经分类的对象的真实世界位置。在一些实施方式中,位置特征数据可以包括位置特征序列。例如,位置特征数据中的所述一个或多个位置特征可以被组织成位置特征嵌入序列。位置特征嵌入序列中的每个位置特征嵌入可以对应于图像帧序列中的图像帧,并且每个位置特征嵌入可以表示与对应图像帧中描绘的一个或多个经分类的对象相关联的一个或多个位置特征。

[0040] 该系统可以将图像特征数据输入到帧级位置特征提取模型中,并且获得作为帧级位置特征提取模型的输出的位置特征数据。例如,图像特征数据可以包括表示图像特征序列并且对应于图像帧序列的图像特征嵌入序列。该系统可以将表示与图像帧相关联的图像特征的数据(例如,图像特征嵌入序列中的图像特征嵌入)输入到帧级位置特征提取模型中,并且获得作为帧级位置特征提取模型的输出的包括位置特征嵌入的位置特征数据,其中该位置特征嵌入表示与图像帧相关联(例如,与图像帧中描绘的一个或多个经分类的对象相关联)的一个或多个位置特征。以这种方式,系统可以将图像特征嵌入序列中的每个图像特征嵌入输入到帧级位置特征提取模型中,并且获得位置特征数据,其中该位置特征数据包括表示位置特征序列并且对应于图像帧序列的位置特征嵌入序列。

[0041] 帧级位置预测模型可以被配置为接收表示位置特征序列的数据(例如,由帧级位置特征提取模型输出的位置特征数据),并且响应于接收到位置特征序列,输出包括与一个或多个经分类的对象相关联的坐标的坐标数据。在一些实施方式中,坐标数据可以包括对应于图像帧序列的坐标嵌入序列。每个坐标嵌入可以表示与对应图像帧中描绘的一个或多

个经分类的对象相关联的坐标。与图像帧中描绘的经分类的对象相关联的坐标可以指示经分类的对象在与图像帧相关联的相机坐标空间中的三维位置。

[0042] 该系统可以将位置特征数据输入到帧级位置预测模型中,并且获得作为该帧级位置预测模型的输出的坐标数据。例如,位置特征数据可以包括表示位置特征序列并且对应于图像帧序列的位置特征嵌入序列。该系统可以将表示与图像帧相关联的位置特征(例如,位置特征嵌入序列中的位置特征嵌入)的数据输入到帧级位置预测模型中,并且获得作为帧级位置预测模型的输出的包括坐标嵌入的坐标数据,其中该坐标嵌入表示与图像帧中描绘的一个或多个经分类的对象相关联的坐标。以这种方式,系统可以将位置特征嵌入序列中的每个位置特征嵌入输入到帧级位置预测模型中,并且获得坐标数据,其中该坐标数据包括表示坐标序列并且对应于图像帧序列的坐标嵌入序列。

[0043] 地理位置预测模型可以被配置为至少部分地基于由帧级位置预测模型输出的坐标数据和相机姿态数据来确定与一个或多个经分类的对象相关联的预测真实世界位置。在一些实施方式中,地理位置预测模型可以被配置为通过将坐标值数据中与经分类的对象相关联的坐标从相机坐标空间转换成真实世界坐标(例如,纬度和经度),来确定一个或多个经分类的对象的预测真实位置。例如,地理位置预测模型可以确定图像帧序列中的多个图像帧中描绘了经分类的对象。地理位置预测模型可以至少部分地基于坐标数据获得与所述多个图像帧中的每一个图像帧的经分类的对象相关联的坐标,并且至少部分地基于相机姿态数据将与经分类的对象相关联的所述多个坐标转换成真实世界坐标(例如,纬度和经度)。例如,地理位置预测模型可以基于用于捕获图像帧的相机的位置和/或方位,将经分类的对象在与图像帧相关联的相机坐标空间中的三维位置转换成经分类的对象的真实世界坐标。以这种方式,系统可以确定与经分类的对象相关联的多个真实世界坐标的时间加权平均值,以确定经分类的对象的预测真实世界位置。

[0044] 在一些实施方式中,地理位置预测模型可以被配置为验证由帧级位置预测模型输出的坐标数据,并且基于该验证来确定预测真实世界坐标。作为示例,地理位置预测模型可以验证与识别到的对象相关联的坐标是准确的。作为另一示例,地理位置预测模型可以验证与跨多个图像帧的识别到的对象相关联的坐标对应于相同的识别到的对象。

[0045] 在一些实施方式中,可以至少部分地基于多个损失值中的一个或多个来训练地理位置预测模型,以便确保预测真实世界位置是准确的并且对应于感兴趣的经分类的对象。所述多个损失值可以包括位置一致性损失、外观一致性损失、瞄准(aiming)损失和视场(field-of-view,F0V)损失。作为示例,该系统可以至少部分地基于与跨多个图像帧的识别到的对象相关联的坐标之间的方差来确定位置一致性损失。该系统可以使用所确定的位置一致性损失来训练地理位置预测模型,使得由地理位置预测模型确定的坐标跨经分类的对象的多个图像帧一致。

[0046] 作为另一示例,系统可以至少部分地基于图像特征数据(例如,由图像特征提取模型输出)和注意力值数据(例如,由对象分类模型输出),确定外观一致性损失。特别地,该系统可以用包括在注意力值数据中的空间注意力值来加权对应于图像帧的图像特征,以确定多个图像帧的外观特征,并且该系统可以至少部分地基于跨多个图像帧的所确定的外观特征之间的方差来确定外观一致性损失。该系统可以使用所确定的外观一致性损失来训练地理位置预测模型,使得由地理位置预测模型分类的一个或多个对象在该对象可见的每个图

像帧中具有类似的视觉外观。

[0047] 作为另一示例,该系统可以至少部分地基于坐标数据(例如,由帧级位置预测模型输出)和注意力值数据(例如,由对象分类模型输出),来确定瞄准损失。该系统可以使用瞄准损失来训练地理位置预测模型,使得坐标数据中与图像帧中描绘的经分类的对象相关联的坐标被投影到与其中与经分类的对象相关联的空间注意力最高的区域中的图像帧相关联的相机坐标空间中。

[0048] 作为另一示例,该系统可以确定FOV损失,以将预测真实世界坐标约束在用于捕获图像帧的相机的实际可能FOV内,其中预测真实世界坐标是基于该图像帧来确定的。系统可以使用所确定的FOV损失来训练地理位置预测模型,以在预测真实世界坐标的范围上包括有意义的限制(例如,合理的空间)。

[0049] 本文描述的系统和方法可以提供多个技术效果和益处。例如,计算系统可以包括可以从表示具有噪声分类的影像的数据中提取位置信息的一个或多个信息提取模型。(多个)信息提取模型可以被端到端地训练,以预测在具有噪声分类的图像中描绘的一个或多个对象的真实世界位置。例如,(多个)信息提取模型可以用于预测各种类型的街道标志、门牌号、转弯限制、街道名称等的真实世界位置。更特别地,(多个)信息提取模型可以包括对象分类模型和/或地理位置预测模型。可以使用具有弱分类标签(例如,与图像帧序列相关联的单个分类标签)的图像数据作为监督信号来训练对象分类模型,并且可以以无监督的方式(例如,基于所确定的位置一致性损失、外观一致性损失、瞄准损失和/或FOV损失)来训练地理位置预测模型。通过能够利用大量的有噪声的加标签的数据和非常大量的未加标签的数据,本公开能够更快更便宜地开发用于更多应用的模型。附加地,由信息提取模型提取的位置信息可以用于开发许多新的应用,其中对于这些应用以前没有足够的地面真实位置数据。例如,(多个)信息提取模型可以被包括在车辆上的计算系统中,或者作为dashcam应用的一部分,并且可以用于检测真实世界中的对象,而不需要发送数据进行离线处理。此外,(多个)信息提取模型的一个或多个组件(例如,图像特征提取模型、对象分类模型、地理位置预测模型等)可以集成到一个或多个其他机器学习模型中。例如,具有空间-时间注意力的对象分类模型可以用于确定与具有噪声分类的各种视频内容相关联的分类(例如,以对视频共享平台上的暴力或攻击性内容进行分类)。

[0050] 示例设备和系统

[0051] 图1A描绘了根据本公开的示例实施例的执行信息提取的示例地理位置系统100的框图。特别地,地理位置系统100可以预测在多个图像中描绘的一个或多个对象的真实世界位置。地理位置系统100可以对应于包括通过网络180通信耦合的用户计算设备102、服务器计算系统130和训练计算系统150的计算系统。

[0052] 用户计算设备102可以是任何类型的计算设备,诸如个人计算设备(例如,膝上型或台式)、移动计算设备(例如,智能手机或平板计算机)、游戏控制台或控制器、可穿戴计算设备、嵌入式计算设备或任何其他类型的计算设备。

[0053] 用户计算设备102包括一个或多个处理器112和存储器114。所述一个或多个处理器112可以是任何合适的处理设备(例如,处理器核心、微处理器、ASIC、FPGA、控制器、微控制器等),并且可以是一个处理器或者可操作地连接的多个处理器。存储器114可以包括一个或多个非暂时性计算机可读存储介质,诸如RAM、ROM、EEPROM、EPROM、闪存设备、磁盘等、

以及它们的组合。存储器114可以存储数据116和由处理器112执行以使用户计算设备102执行操作的指令118。

[0054] 在一些实施方式中,用户计算设备102可以存储或包括一个或多个信息提取模型120。例如,信息提取模型120可以是或者可以以其他方式包括各种机器学习模型,诸如神经网络(例如,深度神经网络)或其他类型的机器学习模型,包括非线性模型和/或线性模型。神经网络可以包括前馈神经网络、递归神经网络(例如,长短期记忆递归神经网络)、卷积神经网络或其他形式的神经网络。示例信息提取模型120参考图2-7进行了讨论。

[0055] 在一些实施方式中,一个或多个信息提取模型120可以通过网络180从服务器计算系统130接收,被存储在用户计算设备存储器114中,并且然后由所述一个或多个处理器112使用或实施。在一些实施方式中,用户计算设备102可以实施单个信息提取模型120的多个并行实例(例如,以执行跨多个实例的并行信息提取)。

[0056] 更特别地,信息提取模型120可以被配置为接收图像数据,并且响应于接收到该图像数据而输出地理位置数据。地理位置系统100可以将图像数据输入到信息提取模型120中,并且获得作为信息提取模型120响应于输入该图像数据的输出的地理位置数据。

[0057] 附加地或可替代地,一个或多个信息提取模型140可以被包括在服务器计算系统130中,或者以其他方式由服务器计算系统130存储和实施,其中服务器计算系统130根据客户端-服务器关系与用户计算设备102通信。例如,(多个)信息提取模型140可以由服务器计算系统130实施为网络服务(例如,地理位置信息提取服务)的一部分。因此,一个或多个模型120可以在用户计算设备102处存储和实施,和/或一个或多个模型140可以在服务器计算系统130处存储和实施。

[0058] 用户计算设备102还可以包括接收用户输入的一个或多个用户输入组件122。例如,用户输入组件122可以是触敏组件(例如,触敏显示屏或触摸板),其中该触敏组件对用户输入对象(例如,手指或触笔)的触摸敏感。触敏组件可以用作实施虚拟键盘。其他示例用户输入组件包括麦克风、传统键盘或用户可以通过其提供用户输入的其他手段。

[0059] 服务器计算系统130包括一个或多个处理器132和存储器134。一个或多个处理器132可以是任何合适的处理设备(例如,处理器核心、微处理器、ASIC、FPGA、控制器、微控制器等),并且可以是一个处理器或者可操作地连接的多个处理器。存储器134可以包括一个或多个非暂时性计算机可读存储介质,诸如RAM、ROM、EEPROM、EPROM、闪存设备、磁盘等、以及它们的组合。存储器134可以存储数据136和由处理器132执行以使服务器计算系统130执行操作的指令138。

[0060] 在一些实施方式中,服务器计算系统130包括一个或多个服务器计算设备或者由一个或多个服务器计算设备来实施。在其中服务器计算系统130包括多个服务器计算设备的实例中,这种服务器计算设备可以根据顺序计算体系结构、并行计算体系结构或它们的一些组合来操作。

[0061] 如上所述,服务器计算系统130可以存储或以其他方式包括一个或多个机器学习信息提取模型140。例如,(多个)模型140可以是或者可以以其他方式包括各种机器学习模型。示例机器学习模型包括神经网络或其他多层非线性模型。示例神经网络包括前馈神经网络、深度神经网络、递归神经网络和卷积神经网络。示例模型140参考图2-7进行讨论。

[0062] 用户计算设备102和/或服务器计算系统130可以经由与通过网络180通信耦合的

训练计算系统150的交互来训练模型120和/或140。训练计算系统150可以与服务器计算系统130分离,或者可以是服务器计算系统130的一部分。

[0063] 训练计算系统150包括一个或多个处理器152和存储器154。一个或多个处理器152可以是任何合适的处理设备(例如,处理器核心、微处理器、ASIC、FPGA、控制器、微控制器等),并且可以是一个处理器或者可操作地连接的多个处理器。存储器154可以包括一个或多个非暂时性计算机可读存储介质,诸如RAM、ROM、EEPROM、EPROM、闪存设备、磁盘等、以及它们的组合。存储器154可以存储数据156和由处理器152执行以使训练计算系统150执行操作的指令158。在一些实施方式中,训练计算系统150包括一个或多个服务器计算设备或者以其他方式由一个或多个服务器计算设备实施。

[0064] 训练计算系统150可以包括模型训练器160,其中该模型训练器160使用各种训练或学习技术(诸如误差反向传播)来训练存储在用户计算设备102和/或服务器计算系统130中的机器学习模型120和/或140。在一些实施方式中,执行误差反向传播可以包括随着时间执行截断反向传播。模型训练器160可以执行多种泛化技术(例如,权重衰减、丢弃等)以提高被训练的模型的泛化能力。

[0065] 特别地,模型训练器160可以基于训练数据集162来训练信息提取模型120和/或140。作为示例,训练数据162可以包括经弱分类的图像数据,诸如包括与图像帧序列相关联的单个分类标签的图像数据。模型训练器160可以通过使用具有弱分类的图像数据来训练包括在信息提取模型120和/或140中的对象分类模型。作为另一示例,训练数据162可以包括与图像帧序列相关联的多于一个的分类标签,并且模型训练器160可以通过使用具有多于一个分类标签的图像数据来训练包括在信息提取模型120和/或140中的对象分类模型。作为另一示例,训练数据162可以包括作为信息提取模型120和/或140的输入而提供的数据、以及作为信息提取模型120和/或140响应于输入数据的输出而提供的数据。模型训练器160可以通过使用输入数据和输出数据,以无监督的方式训练包括在信息提取模型120和/或140中的地理位置预测模型。

[0066] 在一些实施方式中,如果用户已经提供许可,则训练示例可以由用户计算设备102提供。因此,在这种实施方式中,提供给用户计算设备102的模型120可以由训练计算系统150基于从用户计算设备102接收的用户特定数据来训练。在一些情况下,这个过程可以称为个性化(personalizing)模型。

[0067] 模型训练器160包括用于提供期望功能的计算机逻辑。模型训练器160可以在控制通用处理器的硬件、固件和/或软件中实施。例如,在一些实施方式中,模型训练器160包括存储在存储设备上、加载到存储器中并且由一个或多个处理器执行的程序文件。在其他实施方式中,模型训练器160包括存储在有形的计算机可读存储介质(诸如RAM硬盘或光学或磁性介质)中的一个或多个计算机可执行指令集。

[0068] 网络180可以是任何类型的通信网络(诸如局域网(例如,内部网)、广域网(例如,互联网)、或者它们的一些组合),并且可以包括任意数量的有线或无线链路。一般地,通过网络180的通信可以使用各种通信协议(例如,TCP/IP、HTTP、SMTP、FTP)、编码或格式(例如,HTML、XML)和/或保护方案(例如,VPN,安全HTTP,SSL),来经由各种类型的有线和/无线连接承载。

[0069] 图1A示出了可用于实施本公开的一个示例计算系统。也可以使用其他计算系统。

例如,在一些实施方式中,用户计算设备102可以包括模型训练器160和训练数据集162。在这种实施方式中,模型120可以在用户计算设备102处被本地训练和使用。在这种实施方式的一些中,用户计算设备102可以实施模型训练器160,以基于用户特定数据来个性化模型120。

[0070] 图1B描绘了根据本公开的示例实施例执行信息提取的示例计算设备10的框图。计算设备10可以是用户计算设备或服务器计算设备。

[0071] 计算设备10包括多个应用(例如,应用1到N)。每个应用都包括其自己的机器学习库和(多个)机器学习模型。例如,每个应用可以包括机器学习模型。示例应用包括文本消息应用、电子邮件应用、听写应用、虚拟键盘应用、浏览器应用等。

[0072] 如图1B所示,每个应用可以与计算设备的多个其他组件(诸如一个或多个传感器、上下文管理器、设备状态组件和/或附加组件)通信。在一些实施方式中,每个应用可以使用API(例如,公共API)来与每个设备组件进行通信。在一些实施方式中,每个应用使用的API特定于该应用。

[0073] 图1C描绘了根据本公开的示例实施例的执行信息提取的示例计算设备50的框图。计算设备50可以是用户计算设备或服务器计算设备。

[0074] 计算设备50包括多个应用(例如,应用1到N)。每个应用与一个中央智能层进行通信。示例应用包括文本消息应用、电子邮件应用、听写应用、虚拟键盘应用、浏览器应用等。在一些实施方式中,每个应用可以使用API(例如,跨所有应用的公共API)来与中央智能层进行通信。

[0075] 中央智能层包括多个机器学习模型。例如,如图1C所示,可以由中央智能层为各个应用提供相应的机器学习模型(例如,模型)并且由中央智能层管理。在其他实施中,两个或多个应用可以共享单个机器学习模型。例如,在一些实施方式中,中央智能层可以为所有应用提供单个模型(例如,单个模型)。在一些实施方式中,中央智能层被包括在计算设备50的操作系统中或者以其他方式由计算设备50的操作系统来实施。

[0076] 中央智能层可以与中央设备数据层进行通信。中央设备数据层可以是计算设备50的集中式数据仓库。如图1C所示,中央设备数据层可以与计算设备的多个其他组件(诸如一个或多个传感器、上下文管理器、设备状态组件和/或附加组件)进行通信。在一些实施方式中,中央设备数据层可以使用API(例如,私有API)与每个设备组件进行通信。

[0077] 图2描绘了根据本公开的示例实施例的示例信息提取模型200的框图。在一些实施方式中,信息提取模型200被训练成接收描述多个图像的输入数据204(例如,图像数据)(诸如图像帧序列)的集合,并且作为接收输入数据204的结果,提供包括图像帧序列中描绘的一个或多个对象的预测真实世界位置的输出数据206(例如,地理位置数据)。在一些实施方式中,输入数据204可以包括指示用于捕获图像帧序列中的一个或多个图像帧的相机的真实世界位置和/或方位的相机姿态数据。

[0078] 图3描绘了根据本公开的示例实施例的示例信息提取模型300的框图。信息提取模型300与图2的信息提取模型200类似,除了信息提取模型300还包括图像特征提取模型302、对象分类模型306和地理位置预测模型310。

[0079] 在一些实施方式中,图像特征提取模型302被训练成接收输入数据204(例如,表示图像帧序列的数据)或它的一部分,并且作为接收输入数据204的结果,提供包括从输入数

据204中的一个或多个图像帧提取的一个或多个图像特征的图像特征数据304。在一些实施方式中,图像特征数据304可以包括表示从所述一个或多个图像帧提取的图像特征序列的图像特征嵌入序列。

[0080] 在一些实施方式中,对象分类模型306被训练成接收图像特征数据304(例如,表示图像特征序列的数据),并且作为接收图像特征数据304的结果,提供注意力值数据308和分类数据309。注意力值数据308可以包括与图像特征数据304中的图像特征序列相关联的一个或多个时间注意力值和一个或多个空间注意力值。分类数据309可以包括与输入数据204的一个或多个图像帧中描绘的一个或多个经分类的对象(例如,对应于图像帧中描绘的限速标志的限速值)相关联的一个或多个分类标签。在一些实施方式中,注意力值数据308可以包括表示所述一个或多个时间注意力值的时间注意力数据(例如,如图5中所示的时间注意力数据504)和表示所述一个或多个空间注意力值的空间注意力数据(例如,如图5所示的空间注意力数据506)。

[0081] 在一些实施方式中,地理位置预测模型310被训练成接收图像特征数据304(例如,表示图像特征序列的数据)、输入数据204(例如,表示相机姿态数据的数据)或它的一部分和注意力值数据308,并且作为接收数据的结果,提供输出数据206(例如,地理位置数据),其中输出数据206包括在输入数据204中描绘的(例如,在输入数据204中的一个或多个图像帧中描绘的)一个或多个对象的预测真实世界位置。信息提取模型300可以至少部分地基于分类数据309,来将对象的预测真实世界位置与对应于该对象的分类标签相关联。

[0082] 图4描绘了根据本公开的示例实施例的示例地理位置预测模型400的框图。地理位置预测模型400与图3的地理位置预测模型310类似,除了地理位置预测模型400还包括位置特征提取模型402、位置预测模型406和坐标转换模型410。

[0083] 在一些实施方式中,位置特征提取模型402被训练成接收图像特征数据304和注意力值数据308,并且作为接收数据的结果,提供包括与在输入数据204中描绘的一个或多个经分类的对象相关联的一个或多个位置特征的位置特征数据404。在一些实施方式中,位置特征数据可以包括对应于输入数据204中的图像帧序列的位置特征嵌入序列。位置特征嵌入序列中的每个位置特征嵌入可以表示与对应图像帧中描绘的一个或多个经分类的对象相关联的位置特征。

[0084] 在一些实施方式中,位置预测模型406被训练成接收位置特征数据404,并且作为接收位置特征数据404的结果,提供包括与在输入数据204中描绘的一个或多个经分类的对象相关联的坐标的坐标数据408。在一些实施方式中,坐标数据408可以包括对应于输入数据204中的图像帧序列的坐标嵌入序列。坐标嵌入序列中的每个坐标嵌入可以表示与对应图像帧中描绘的一个或多个经分类的对象相关联的坐标。在一些实施方式中,坐标数据408可以包括与经分类的对象相关联的坐标,其中该坐标指示经分类的对象在与描绘经分类的对象的图像帧相关联的相机坐标空间中的三维位置。

[0085] 在一些实施方式中,坐标转换模型410被训练成接收坐标数据408和输入数据204(例如,相机姿态数据)的至少一部分,并且作为接收该数据的结果,提供输出数据206(例如,地理位置数据)。特别地,坐标转换模型410可以将相机坐标空间中与经分类的对象相关联的坐标转换成真实世界坐标(例如,纬度和经度值)。

[0086] 在一些实施方式中,可以至少部分地基于多个损失值中的一个或多个来训练地理

位置预测模型400,以确保预测真实世界位置是准确的并且对应于感兴趣的经分类的对象。作为示例,地理位置系统100可以至少部分地基于与跨多个图像帧的识别到的对象相关联的坐标之间的方差来确定位置一致性损失。地理位置系统100可以使用所确定的位置一致性损失来训练地理位置预测模型400,使得由地理位置预测模型确定的坐标跨经分类的对象的多个图像帧一致。作为另一示例,地理位置系统100可以至少部分地基于图像特征数据304和注意力值数据308来确定外观一致性损失。特别地,地理位置系统100可以加权对应于具有包括在注意力值数据308中的空间注意力值的图像帧的图像特征,以确定多个图像帧的外观特征,并且地理位置系统100可以至少部分地基于跨多个图像帧的所确定的外观特征之间的方差来确定外观一致性损失。地理位置系统100可以使用所确定的外观一致性损失来训练地理位置预测模型400,使得由地理位置预测模型分类的一个或多个对象在该对象可见的每个图像帧中具有类似的视觉外观。作为另一示例,地理位置系统100可以至少部分地基于坐标数据408和注意力值数据308来确定瞄准损失。地理位置系统100可以使用瞄准损失来训练地理位置预测模型400,使得坐标数据408中与图像帧中描绘的经分类的对象相关联的坐标被投影到与其中经分类的对象相关联的空间注意力最高的区域中的图像帧相关联的相机坐标空间中。作为另一示例,地理位置系统100可以确定视场(FOV)损失,以将预测真实世界坐标约束在用于捕获图像帧的相机的实际可能FOV内,其中预测真实世界坐标是基于该图像帧来确定的。地理位置系统100可以使用所确定的FOV损失来训练地理位置预测模型400,以便在预测真实世界坐标的范围上包括有意义的限制(例如,合理的空间)。

[0087] 图5描绘了根据本公开的示例实施例的示例对象分类模型500的框图。对象分类模型500与图3的对象分类模型306类似,除了对象分类模型500输出分类数据309以及注意力值数据308。

[0088] 在一些实施方式中,对象分类模型500被训练成接收表示图像特征序列的数据(例如,图像特征数据304),并且作为接收图像特征数据304的结果,提供分类数据309、时间注意力数据504和空间注意力数据506。分类数据309可以包括与图像帧序列中描绘的一个或多个对象相关联的分类。时间注意力数据504可以包括与图像特征序列相关联的一个或多个时间注意力值,并且空间注意力数据506可以包括与图像特征序列相关联的一个或多个空间注意力值。在一些实施方式中,可以至少部分地基于分类数据309来训练对象分类模型500。例如,地理位置系统100可以至少部分地基于分类数据中的一个或多个分类标签以及与输入数据204中的图像帧序列相关联的分类来确定柔性最大值交叉熵损失。地理位置系统100可以至少部分地基于所确定的柔性最大值交叉熵损失来训练对象分类模型500。

[0089] 在一些实施方式中,对象分类模型500可以包括空间-时间注意力机制层510、包括多个LSTM块的长短期记忆(LSTM)层512和完全连接(fully connected,FC)层514。空间-时间注意力机制层510可以至少部分地基于图像特征数据304来确定时间注意力数据504和空间注意力数据506。LSTM层512中的每个LSTM块可以至少部分地基于图像特征数据304来确定每帧嵌入,并且将每帧嵌入提供给FC层514,以确定跨多个图像帧持续的一个或多个对象。对象分类模型500可以基于时间注意力数据504来加权每帧嵌入,以确定分类数据309。

[0090] 图6描绘了根据本公开的示例实施例的示例位置特征提取模型600的框图。位置特征提取模型600与图4的位置特征提取模型402类似,除了位置特征提取模型600被训练成接收表示对应于单个图像帧的一个或多个图像特征的数据(例如,图像特征数据304)。作为接

收到所述一个或多个图像特征和注意力值数据308的结果,位置特征提取模型600提供包括对应于单个图像帧的一个或多个位置特征的位置特征数据404。在一些实施方式中,对于输入数据204中的每个图像帧,地理位置系统100可以顺序输入表示对应于图像帧的一个或多个图像特征的数据。在一些实施方式中,信息提取模型300可以包括多个位置特征提取模型600,并且地理位置系统100可以并行输入表示一个或多个图像特征的数据。例如,如果信息提取模型300包括第一位置特征提取模型和第二位置特征提取模型600,则地理位置系统100然后可以同时表示与第一图像帧相对应的一个或多个图像特征的数据输入到第一位置特征提取模型600,并且将表示与第二图像帧相对应的一个或多个图像特征的数据输入到第二位置特征提取模型600。以这种方式,位置特征提取模型600可以提供对应于输入数据204中的图像帧序列的位置特征序列(例如,位置特征数据404)。

[0091] 图7描绘了根据本公开的示例实施例的示例位置预测模型700的框图。位置预测模型700与图4的位置预测模型406类似,除了位置预测模型700包括长短期记忆(LSTM)层712(包括多个LSTM块)和完全连接(FC)层714。位置预测模型700被训练成接收表示与输入数据204中的图像帧序列相对应的位置特征序列的数据(例如,位置特征数据404),并且作为接收位置特征序列的结果,位置预测模型700提供表示与图像帧序列中描绘的经分类的对象的图像帧序列相对应的坐标序列的数据。坐标序列可以包括例如与在描绘经分类的对象的每个图像帧中的经分类的对象相关联的坐标。例如,位置预测模型700可以接收包括位置特征嵌入序列的位置特征数据404,每个位置特征嵌入表示对应于图像帧序列中的图像帧的一个或多个位置特征。位置预测模型700可以将每个位置特征嵌入提供给LSTM层712中的对应LSTM块。来自每个LSTM块的输出可以表示对象在对应图像帧中的预测位置。以这种方式,LSTM层712可以输出对象的预测位置序列,其中该预测位置序列对应于该对象在描绘该对象的图像帧序列中的每个图像帧中的预测位置。LSTM层712的输出可以被提供给FC层714,以确定包括对象的坐标序列的坐标数据408。

[0092] 在一些实施方式中,地理位置系统100可以使用位置预测模型700来顺序地确定在输入数据204中的图像帧序列中描绘的多个经分类的对象的坐标序列。例如,位置预测模型700的每次迭代可以输出与图像帧序列中描绘的不同对象相关联的坐标序列。在一些实施方式中,信息提取模型300可以包括多个位置预测模型700,并且地理位置系统100可以将位置特征数据404并行输入到所述多个位置预测模型700中的每一个位置预测模型700。例如,如果信息提取模型300包括第一和第二位置预测模型700,则地理位置系统100然后可以同时表示位置特征数据404输入到第一和第二位置预测模型700中,并且获得作为第一位置预测模型700的输出的与第一经分类的对象相关联的第一坐标序列、以及作为第二位置预测模型700的输出的与第二经分类的对象相关联的第二坐标序列。

[0093] 示例方法

[0094] 图8描绘了根据本公开的示例实施例的执行信息提取的示例方法的流程图。虽然图8描绘了出于说明和讨论的目的以特定顺序执行的步骤,但是本公开的方法不限于特定说明的顺序或布置。方法800的各种步骤可以在不脱离本公开的范围情况下以各种方式省略、重新排列、组合和/或调整。

[0095] 在802,计算系统可以获得表示图像序列的数据。例如,地理位置系统100可以获得包括表示图像序列的数据的输入数据204。地理位置系统100可以将图像序列输入到训练成

从图像序列中提取位置信息的机器学习信息提取模型120/140中。在一些实施方式中,图像序列可以描绘跨图像序列中的多个图像的多个对象,并且信息提取模型120/140的输出可以包括表示与在图像序列中描绘的所述多个对象相关联的真实世界位置的数据。

[0096] 在804,计算系统可以至少部分地基于从图像序列提取的图像特征序列来确定与图像序列相关联的分类标签和注意力值。例如,地理位置系统100可以至少部分地基于表示从图像序列中提取(例如,通过图像特征提取模型302)的图像特征序列的数据(例如,图像特征数据304),确定与图像序列相关联的分类数据309、包括时间注意力值的时间注意力数据504、以及包括空间注意力值的空间注意力数据506。特别地,地理位置系统100可以将图像特征序列、时间注意力数据504和空间注意力数据506输入到弱监督对象分类模型306中,并且响应于输入图像特征序列,获得作为对象分类模型306的输出的分类数据309。地理位置系统100可以至少部分地基于图像特征序列、时间注意力数据504和空间注意力数据506来预测与对象相关联的真实世界位置。

[0097] 在806,计算系统可以至少部分地基于图像特征序列和注意力值来确定位置特征序列。例如,地理位置系统100可以将表示图像特征序列的数据、时间注意力数据504和空间注意力数据506输入到帧级位置特征提取模型600中,并且获得作为帧级位置特征提取模型600响应于输入图像特征序列、时间注意力数据504和空间注意力数据506的输出的表示包括与该对象相关联的一个或多个位置特征的位置特征序列的位置特征数据404。

[0098] 在808,计算系统可以至少部分地基于位置特征序列和注意力值来确定与在图像序列中描绘的一个或多个对象相关联的坐标。例如,地理位置系统100可以将位置特征数据404输入到帧级位置预测模型406中,并且获得作为帧级位置预测模型406响应于输入位置特征数据404的输出的表示相机坐标空间中与该对象相关联的坐标的坐标数据408。地理位置系统100可以至少部分地基于与输入数据204中的该对象相关联的坐标数据408和相机姿态数据来确定与该对象相关联的真实世界坐标。

[0099] 在810,计算系统可以至少部分地基于所确定的坐标来预测与所述一个或多个对象相关联的真实世界位置。例如,地理位置系统100可以获得作为信息提取模型120/140响应于输入输入数据204的输出的表示与在图像序列中描绘的对象相关联的真实世界位置的输出数据206。地理位置系统100可以至少部分地基于分类数据309,来将该对象的预测真实世界位置与对应于该对象的分类标签相关联。

[0100] 图9描绘了根据本公开的示例实施例的训练信息提取模型的示例方法的流程图。虽然图9描绘了出于说明和讨论的目的以特定顺序执行的步骤,本公开的方法不限于特定说明的顺序或布置。方法900的各种步骤可以在不脱离本公开的范围的情况下以各种方式省略、重新排列、组合和/或调整。

[0101] 在902,计算系统(例如,训练计算系统150或地理位置系统100的其他部分)可以获得表示从具有噪声分类的图像序列中提取的图像特征序列的数据。例如,地理位置系统100可以获得表示从具有与图像序列相关联的单个分类标签的图像序列中提取的图像特征序列的图像数据(例如,图像特征数据304)。地理位置系统100可以将图像数据输入到图像特征提取模型302中,并且获得作为图像特征提取模型302响应于输入图像数据的输出的图像特征数据304。

[0102] 在904,计算系统可以至少部分地基于图像特征序列来确定与在图像序列中描绘

的一个或多个对象相关联的分类。例如,地理位置系统100可以将图像特征数据304输入到弱监督对象分类模型306中,并且获得作为对象分类模型306响应于输入图像特征序列的输出的表示与图像帧序列中描绘的对象相关联分类的数据(例如,分类数据309)。

[0103] 在906,计算系统可以确定与所确定的分类相关联的损失。例如,地理位置系统100可以至少部分地基于与图像序列相关联的噪声分类,来确定与由对象分类模型306输出的分类数据309相关联的损失。

[0104] 在908,计算系统可以至少部分地基于与所确定的分类相关联的损失来训练对象分类模型。例如,地理位置系统100可以至少部分地基于所确定的损失来训练对象分类模型306。

[0105] 图10描绘了根据本公开的示例实施例的训练信息提取模型的示例方法的流程图。虽然图10描绘了出于说明和讨论的目的以特定顺序执行的步骤,本公开的方法不限于特定说明的顺序或布置。方法1000的各种步骤可以在不脱离本公开的范围情况下以各种方式省略、重新排列、组合和/或调整。

[0106] 在1002,计算系统(例如,训练计算系统150或地理位置系统100的其他部分)可以获得表示从具有噪声分类的图像序列中提取的图像特征序列的数据、以及表示与图像序列相关联的注意力值的数据。例如,地理位置系统100可以获得表示从图像序列(例如,输入数据204)提取的图像特征序列的数据(例如,图像特征数据304)。地理位置系统100可以将输入数据204输入到图像特征提取模型302中,并且获得作为图像特征提取模型302响应于输入该输入数据204的输出的图像特征数据304。

[0107] 在1004,计算系统可以至少部分地基于位置特征序列和注意力值来预测与在图像序列中描绘的一个或多个对象相关联的真实世界位置。例如,地理位置系统100可以获得作为信息提取模型120/140响应于输入输入数据204的输出的输出数据206。输出数据206可以表示与在图像序列中描绘的对象相关联的真实世界位置。特别地,地理位置系统100可以将图像特征数据304和注意力值数据308输入到位置特征提取模型402中,并且结果,获得包括与在输入数据204中描绘的一个或多个经分类的对象相关联的一个或多个位置特征的位置特征数据404。地理位置系统100可以将位置特征数据404输入到位置预测模型406中,并且结果,获得包括与在输入数据204中描绘的一个或多个经分类的对象相关联的坐标的坐标数据408。地理位置系统100可以将坐标数据408和输入数据204(例如,相机姿态数据)的至少一部分输入到坐标转换模型410,并且结果,获得包括输入数据204中描绘的一个或多个对象(例如,街道标志)的预测真实世界位置(例如,纬度和经度)的输出数据206(例如,地理位置数据)。

[0108] 在1006,计算系统可以至少部分地基于与所述一个或多个对象相关联的预测真实世界位置来确定位置一致性损失。例如,地理位置系统100可以至少部分地基于与跨描绘该对象的图像序列中的多个图像的对象相关联的坐标之间的方差来确定位置一致性损失。

[0109] 在1008,计算系统可以至少部分地基于与所述一个或多个对象相关联的预测真实世界位置来确定外观一致性损失。例如,地理位置系统100可以至少部分地基于跨描绘对象的图像序列中的多个图像而确定的外观特征之间的方差来确定外观一致性损失。

[0110] 在1010,计算系统可以至少部分地基于与所述一个或多个对象相关联的预测真实世界位置来确定瞄准损失。例如,地理位置系统100可以至少部分地基于相机坐标空间中与

该对象相关联的坐标和与跨描绘对象的图像序列中的多个图像的该对象相关联的空间注意力来确定瞄准损失。

[0111] 在1012,计算系统可以至少部分地基于与所述一个或多个对象相关联的预测真实世界位置来确定视场损失。例如,地理位置系统100可以至少部分地基于与对象相关联的真实世界坐标和与用于捕获描绘对象的图像序列的相机相关联的视场来确定视场损失。

[0112] 在1014,计算系统可以至少部分地基于所确定的损失来训练位置预测模型。例如,地理位置系统100可以至少部分地基于位置一致性损失、外观一致性损失、瞄准损失和/或视场损失来训练位置预测模型406。

[0113] 附加公开

[0114] 本文讨论的技术涉及服务器、数据库、软件应用和其他基于计算机的系统,以及采取的动作和发送到这些系统和从这些系统发送的信息。基于计算机的系统的固有灵活性允许组件之间的任务和功能的各种可能的配置、组合以及划分。例如,本文讨论的过程可以使用单个设备或组件或者组合工作的多个设备或组件来实施。数据库和应用可以在单个系统上实施,也可以分布在多个系统上。分布式组件可以顺序或并行操作。

[0115] 虽然已经针对本主题的各种具体实施例详细描述了本主题,但是每个示例都是通过解释的方式提供的,而不是对本公开的限制。本领域的技术人员在理解了前述内容后,可以容易地对这些实施例进行变更、变化和等效。因此,本主题公开不排除包括对本主题的这种修改、变化和/或添加,这对于本领域普通技术人员来说是明显的。例如,作为一个实施例的一部分示出或描述的特征可以与另一个实施例一起使用,以产生又一个实施例。因此,本公开旨在覆盖这些变更、变化和等同物。

[0116] 特别地,虽然图8-10分别描绘了出于说明和讨论的目的以特定顺序执行的步骤,但是本公开的方法不限于特定示出的顺序或布置,方法800、900和1000的各种步骤可以在不脱离本公开的范围的情况下以各种方式省略、重新排列、组合和/或调整。

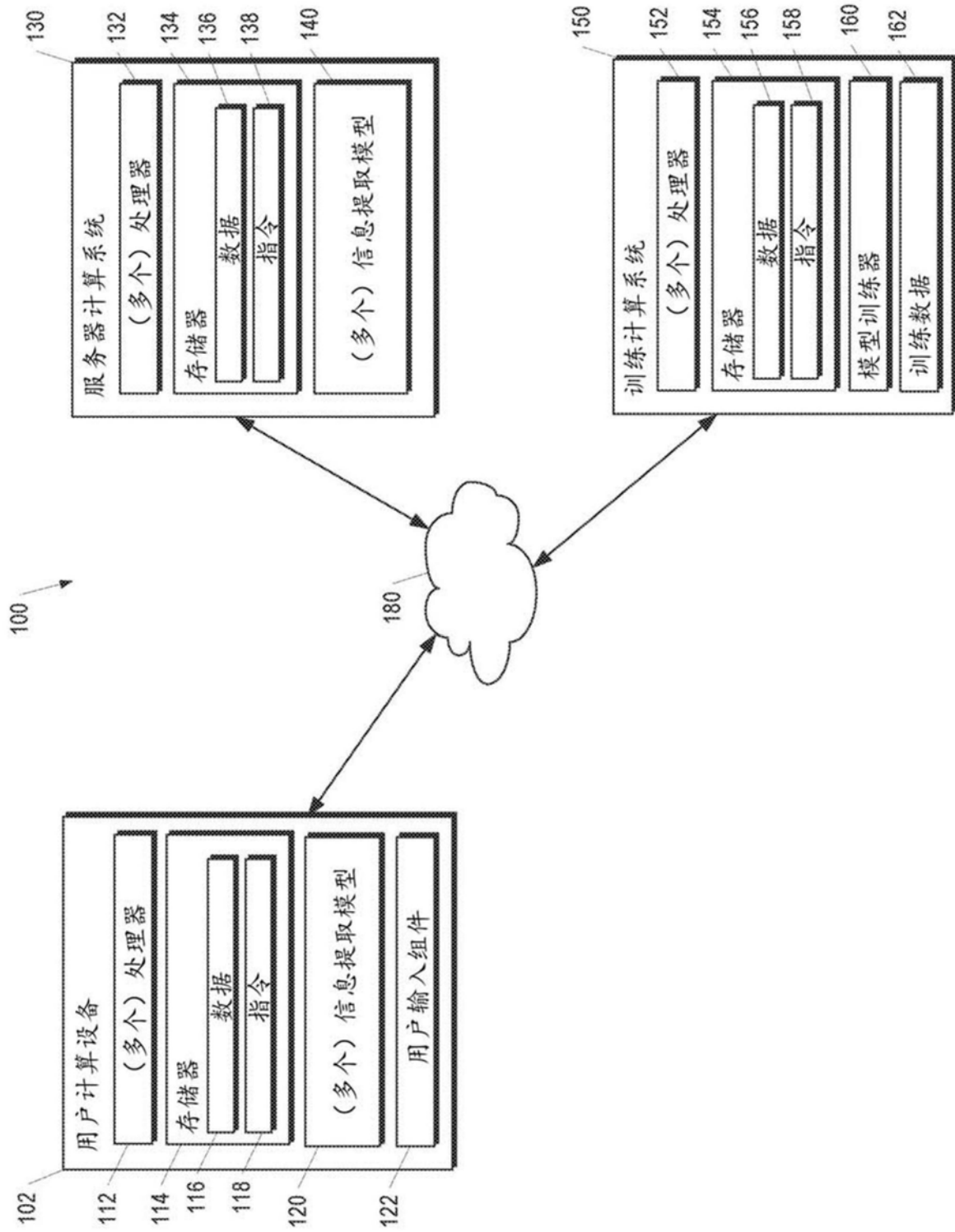


图1A

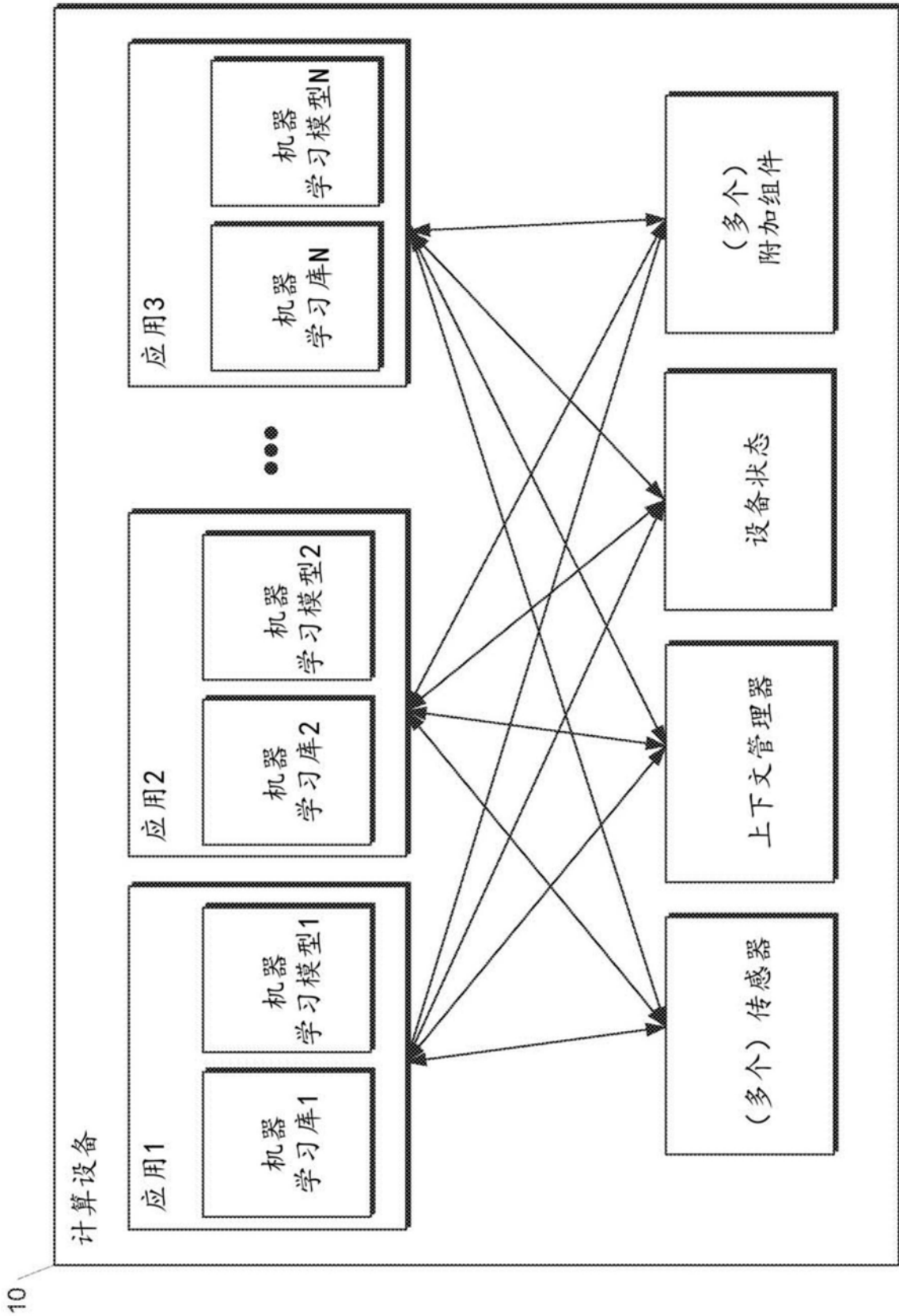


图1B

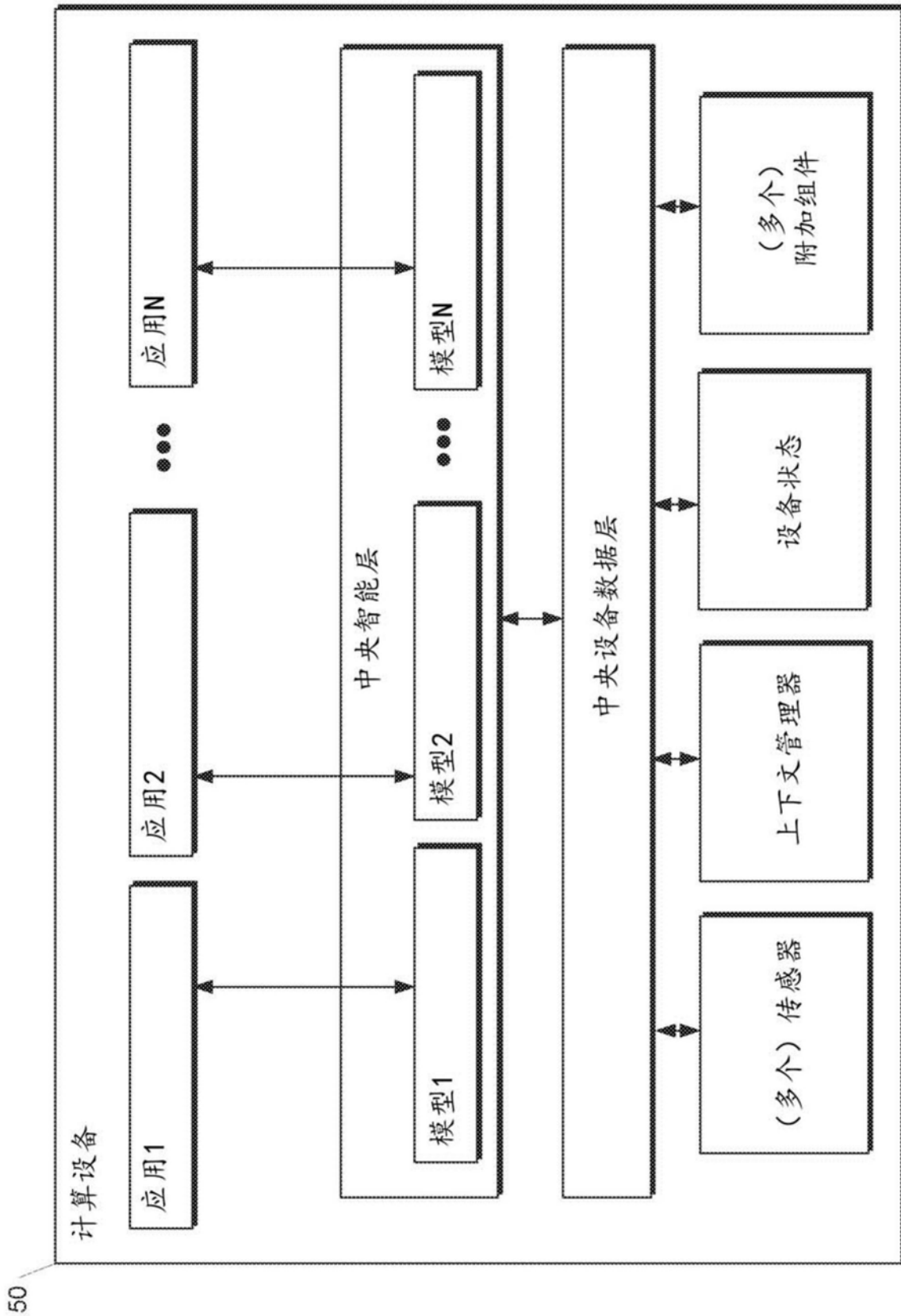


图1C

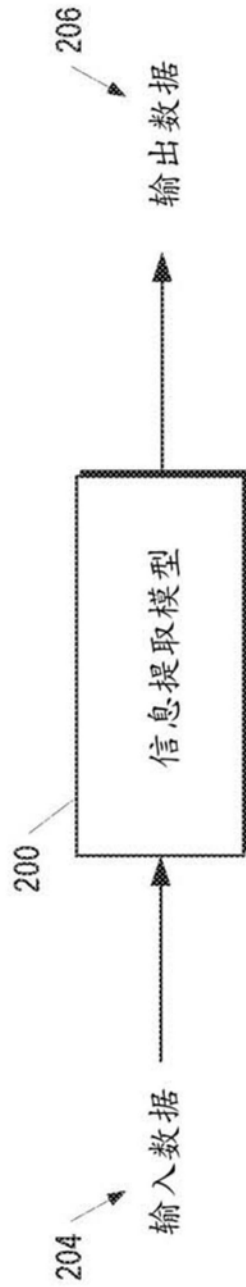


图2

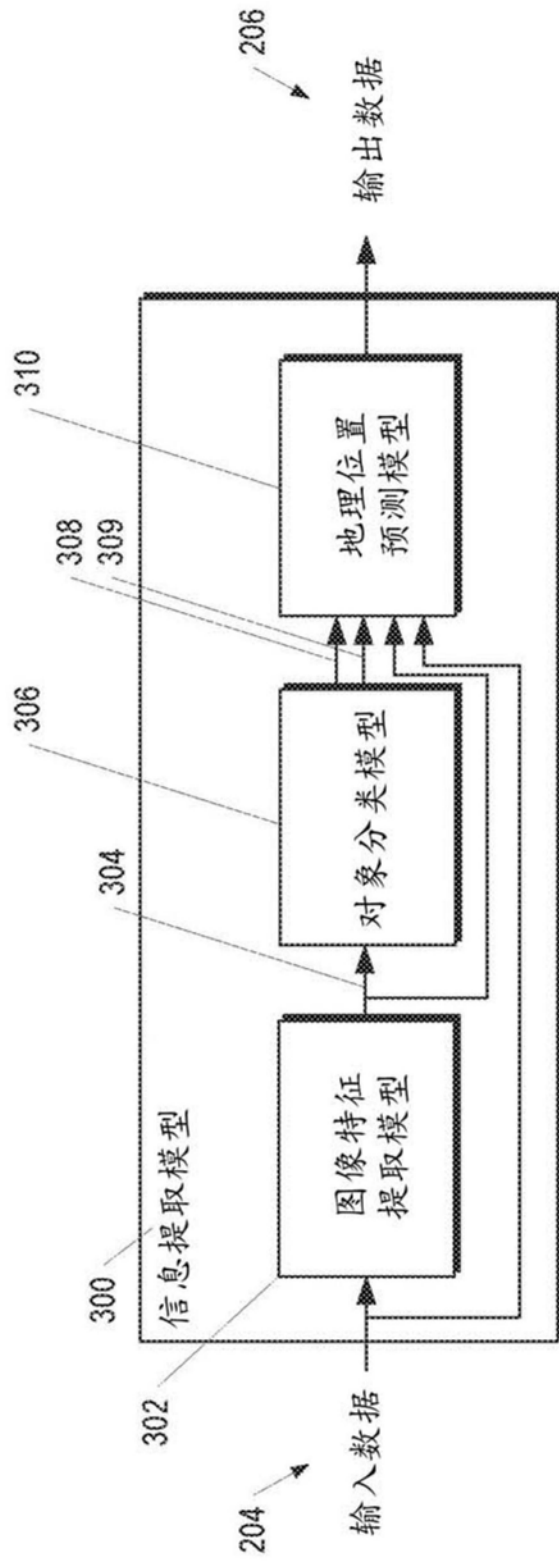


图3

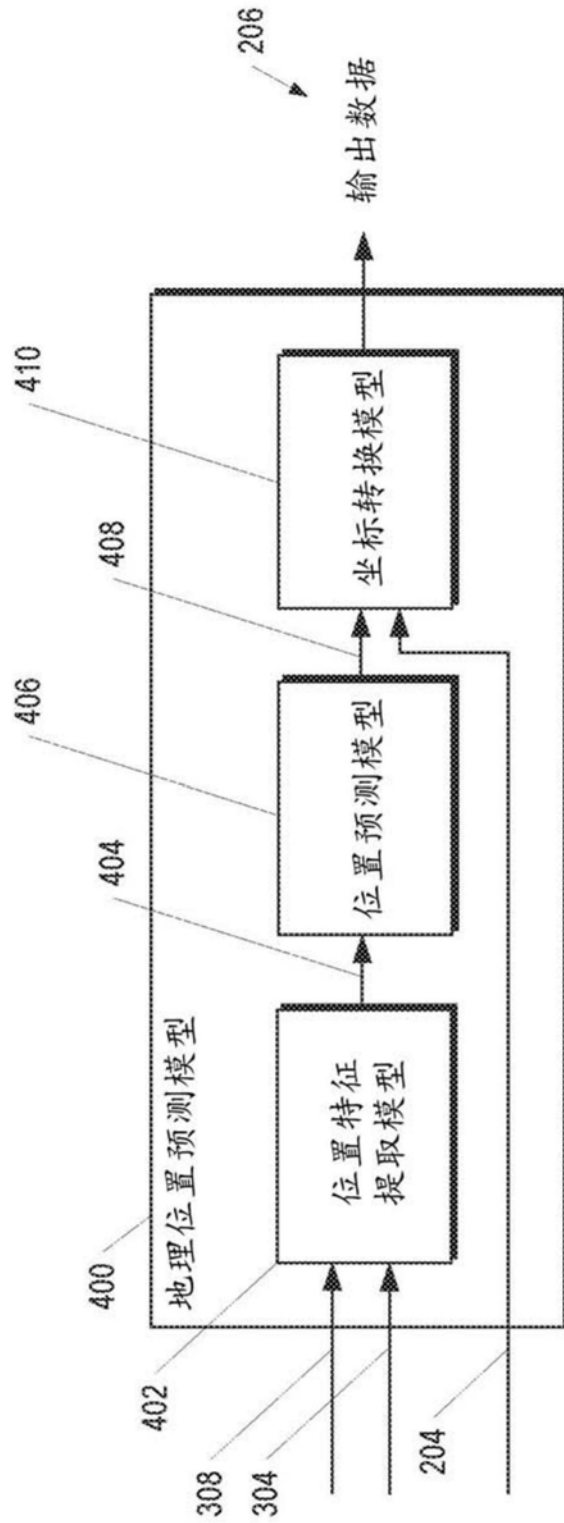


图4

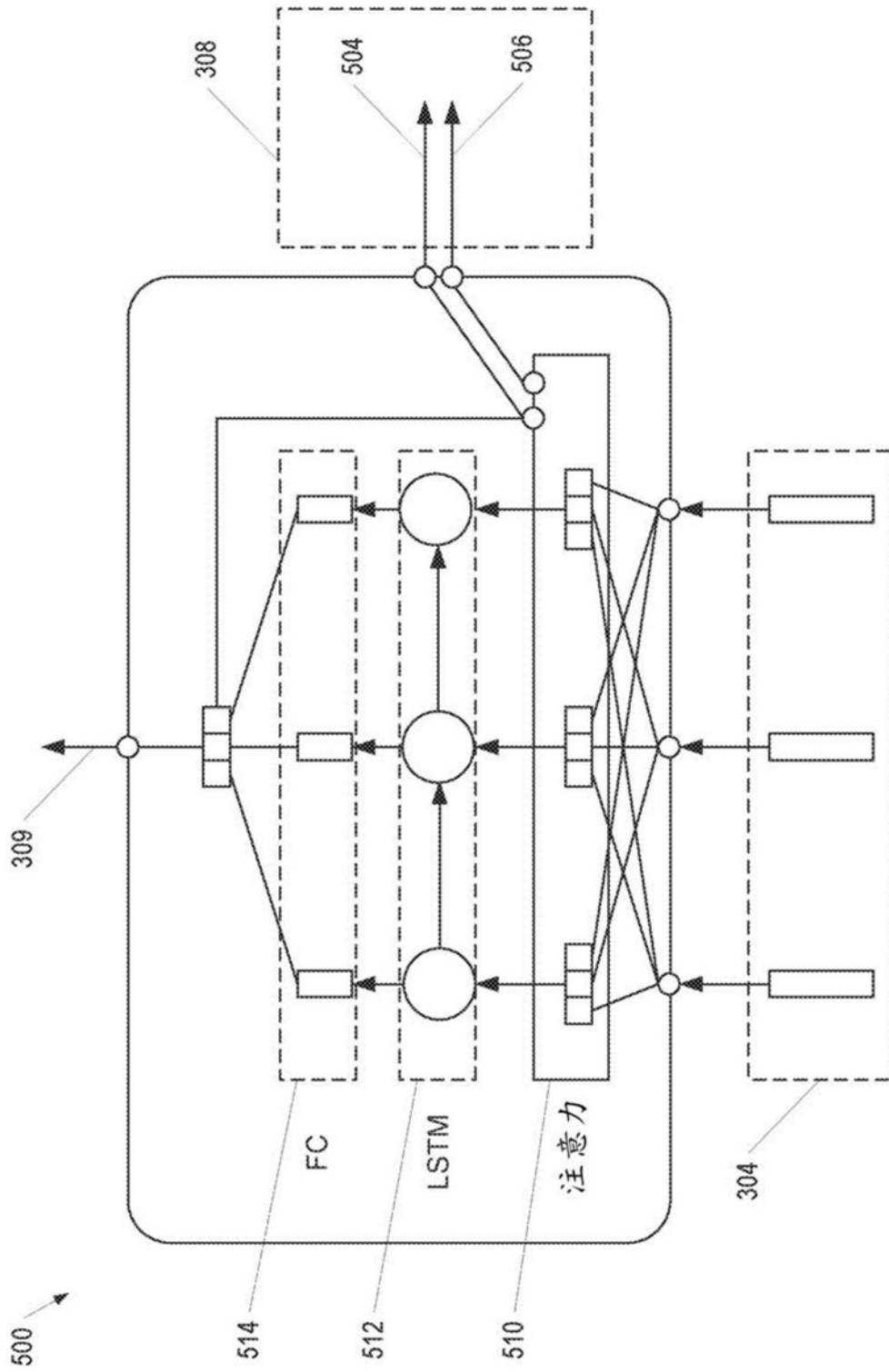


图5

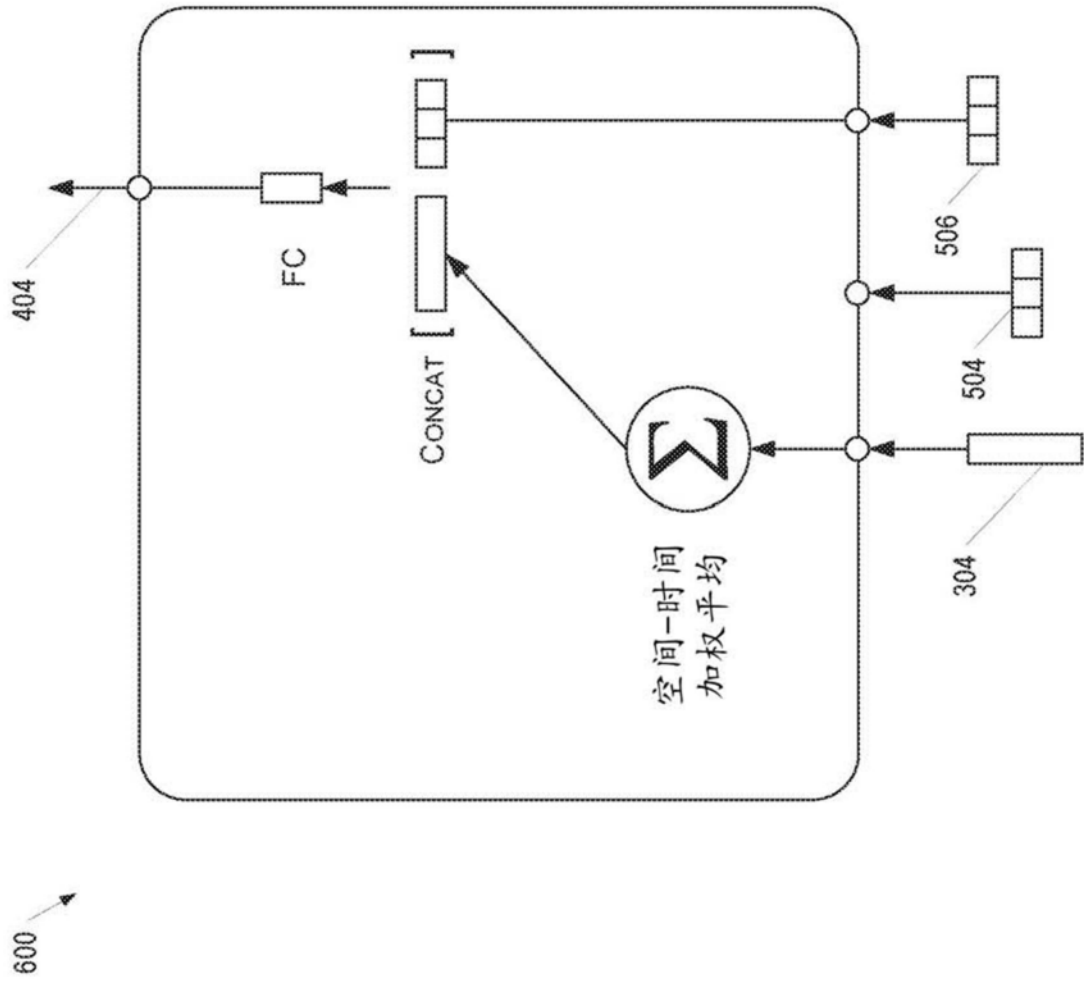


图6

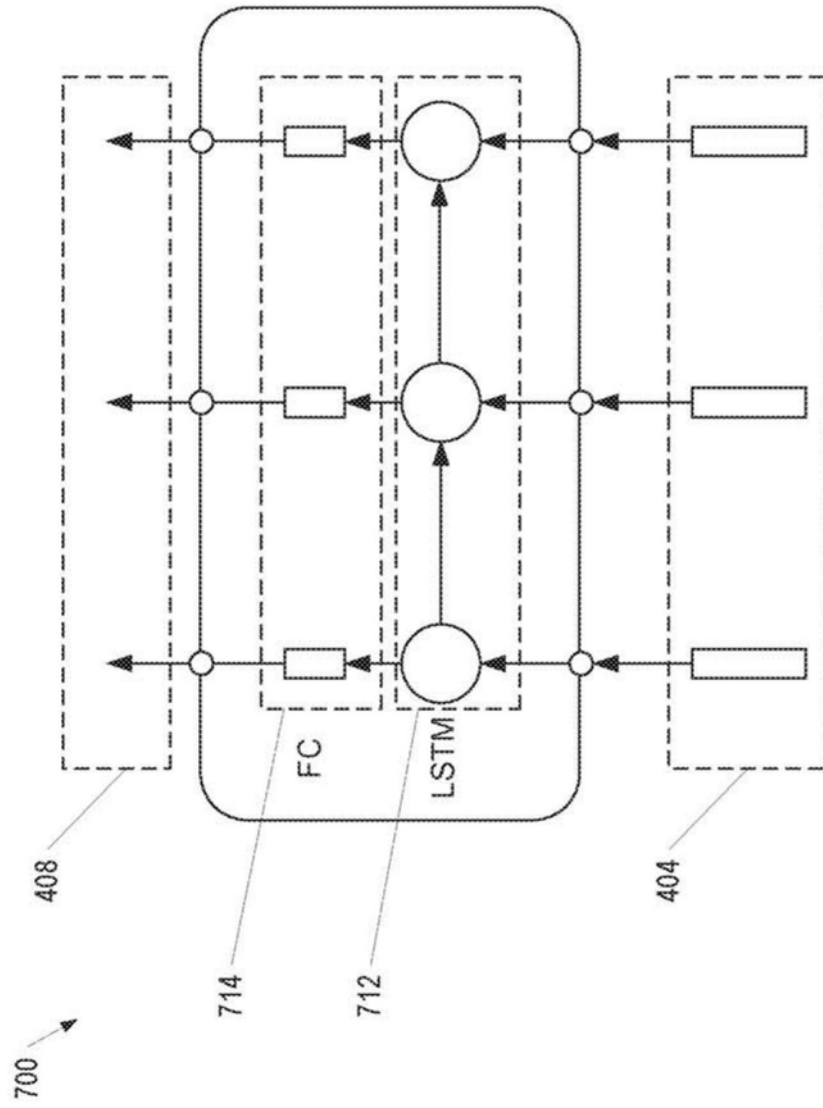


图7

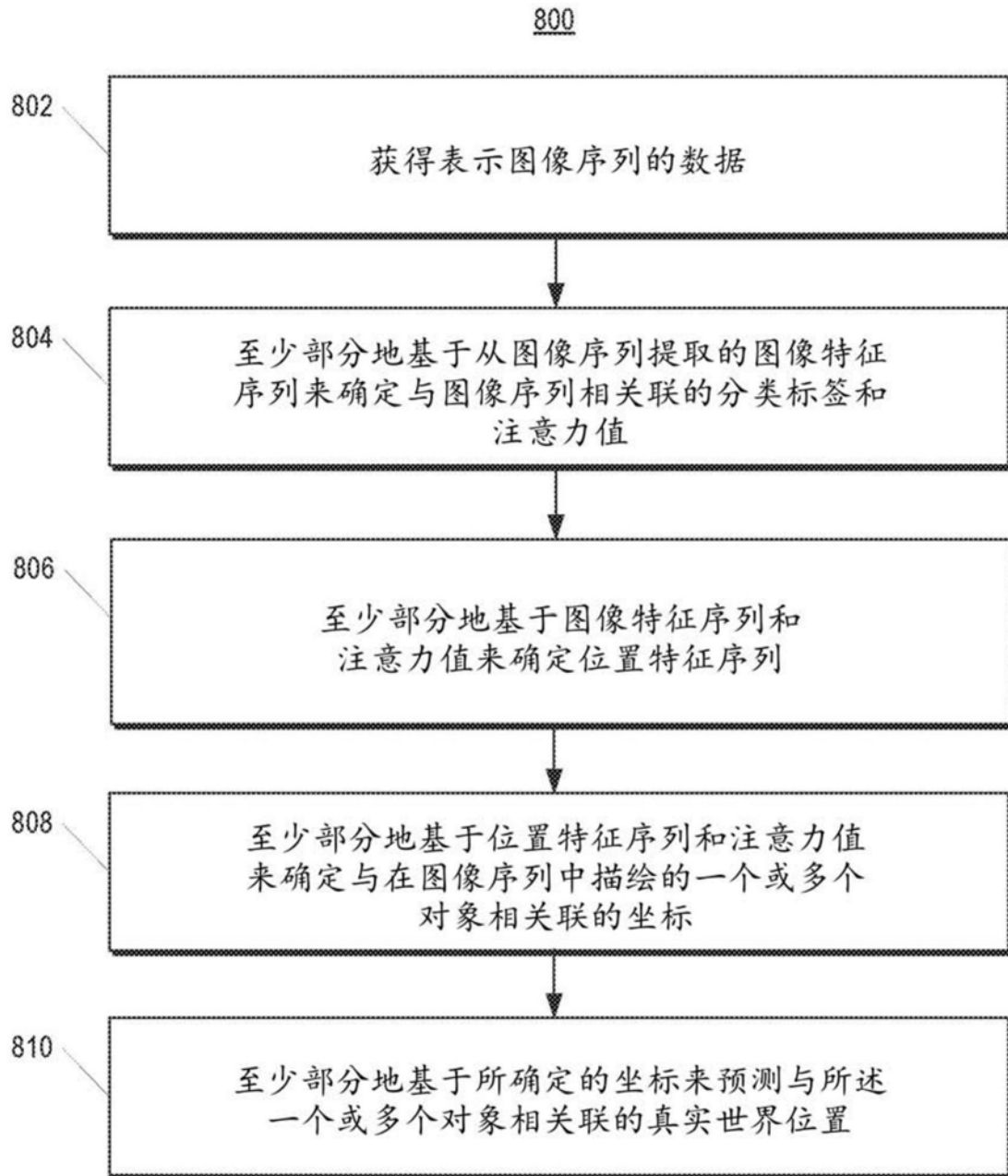


图8

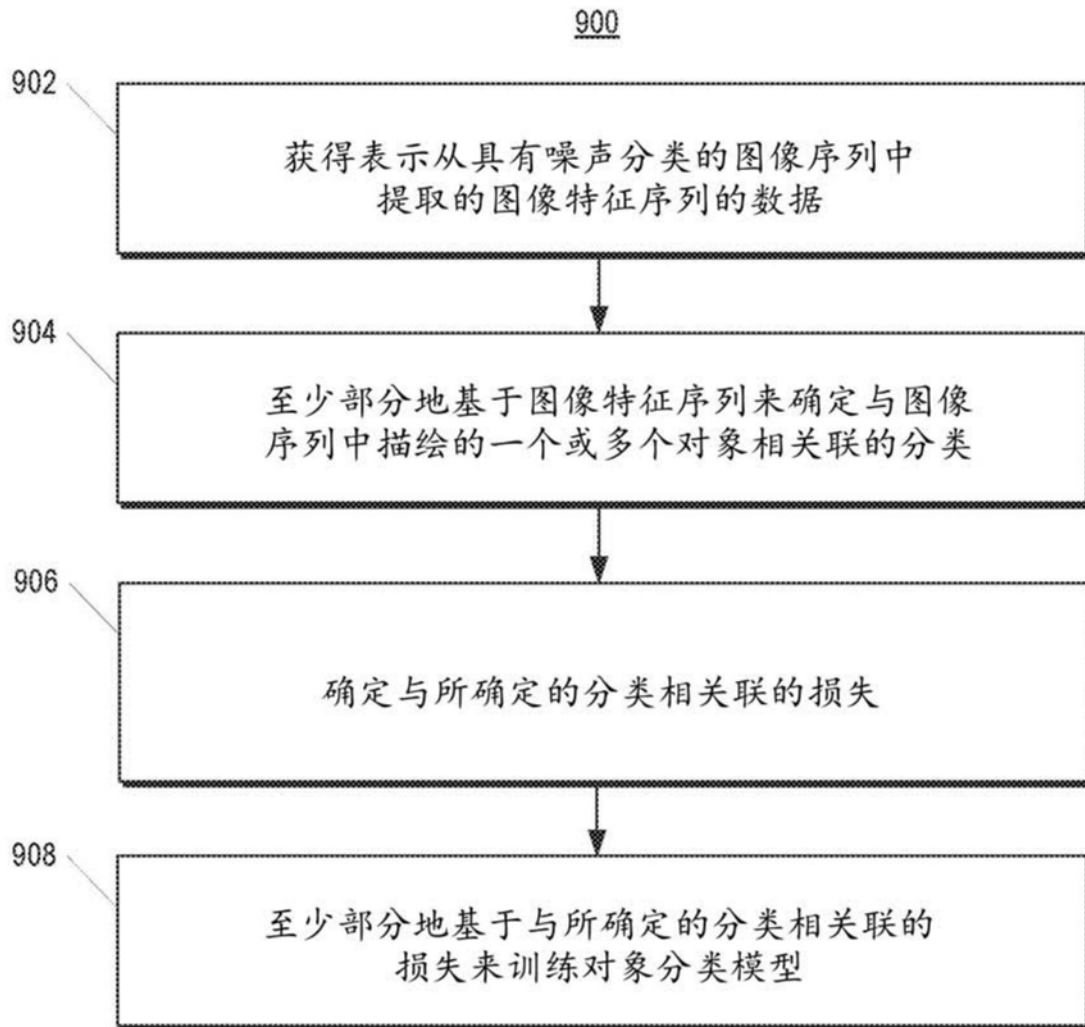


图9

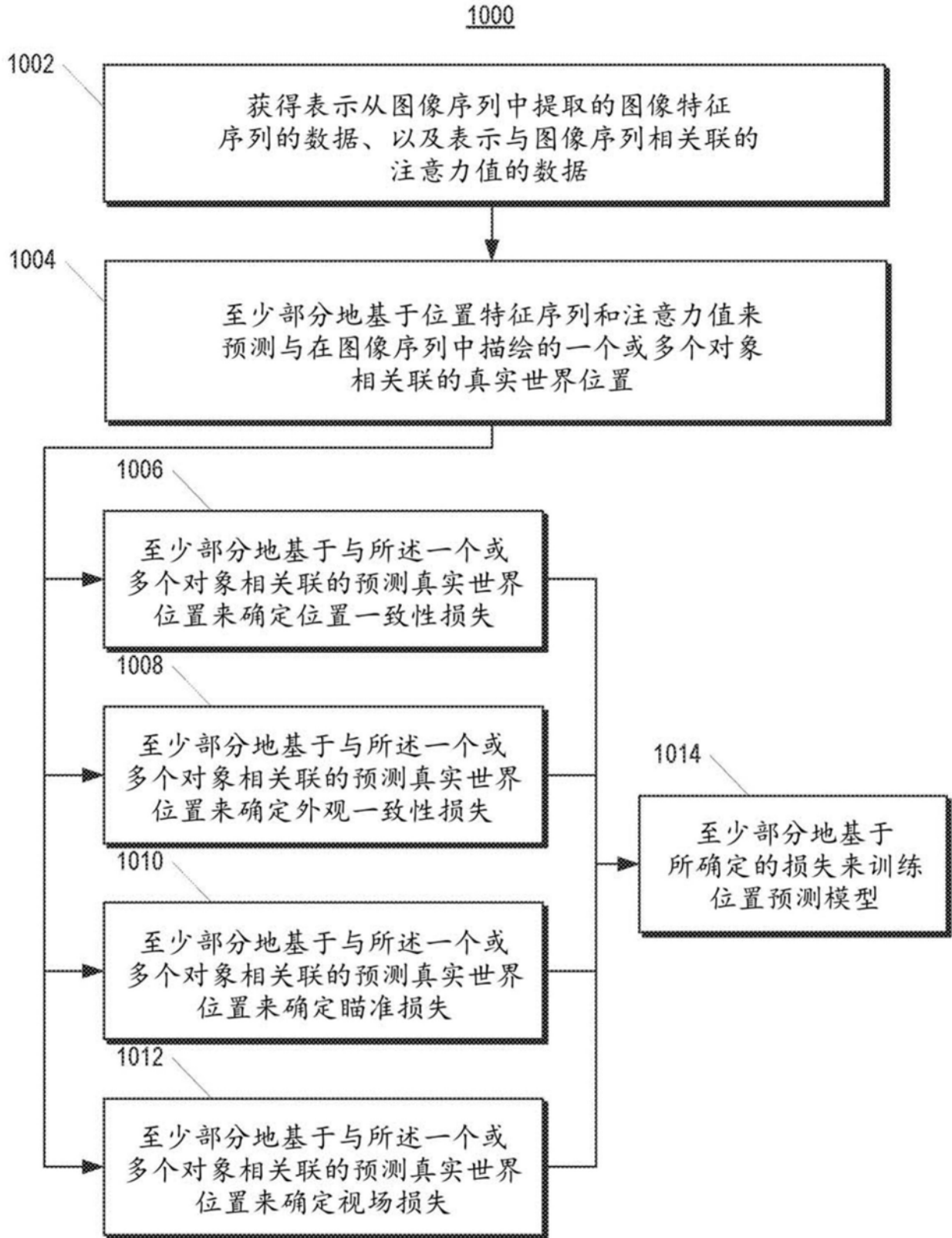


图10