(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117677973 A
(43)申请公布日 2024.03.08

(21)申请号 202280006658.6
(22)申请日 2022.06.30
(85)PCT国际申请进入国家阶段日
2023.03.28
(86)PCT国际申请的申请数据
PCT/US2022/035793 2022.06.30
(87)PCT国际申请的公布数据
WO2024/005832 EN 2024.01.04
(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 奥利·利巴 佩德罗·韦莱兹 李思扬 陈惠中 马塞尔·普亚特 宝雅男
(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
专利代理师 邓聪惠 周亚荣
(51)Int.Cl.
G06T 7/194(2017.01)
G06T 7/136(2017.01)
G06T 7/11(2017.01)

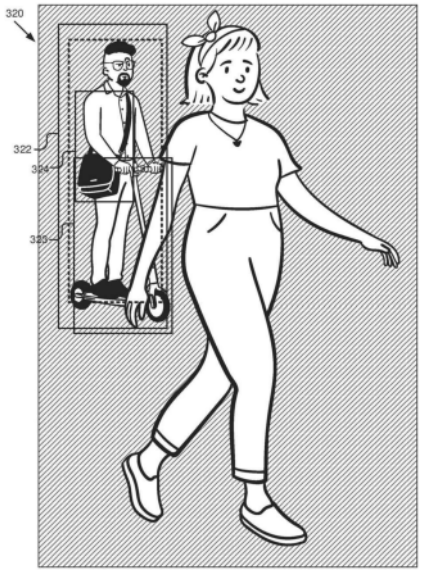
权利要求书3页 说明书13页 附图11页

(54)发明名称

旁观者和附着对象移除

(57)摘要

媒体应用检测初始图像中的旁观者。媒体应用生成包括旁观者的旁观者框,其中旁观者的所有像素在旁观者框内。媒体应用生成包含旁观者和附着到旁观者的一个或多个对象的定位器框。媒体应用聚集旁观者框和定位器框中的一个或多个定位器框以形成聚集框。媒体应用基于聚集框将分割器应用于初始图像,以从初始图像分割旁观者和一个或多个对象以生成旁观者掩模,其中旁观者掩模包括聚集框内的像素的子集。媒体应用生成用与初始图像中的背景匹配的像素替换旁观者掩模内的所有像素的修复图像。



CN 117677973 A

1.一种计算机实现的方法,包括:

检测初始图像中的旁观者;

生成包括所述旁观者的旁观者框,其中所述旁观者的所有像素在所述旁观者框内;

生成包含所述旁观者和附着到所述旁观者的一个或多个对象的定位器框;

聚集所述旁观者框和所述定位器框中的一个或多个定位器框以形成聚集框;

基于所述聚集框将分割器应用于所述初始图像,以从所述初始图像分割所述旁观者和所述一个或多个对象以生成旁观者掩模,其中所述旁观者掩模包括所述聚集框内的像素的子集;以及

生成修复图像,其中所述旁观者掩模内的所有像素在所述修复图像中用与所述初始图像中的背景匹配的像素被替换。

2.根据权利要求1所述的方法,进一步包括:

在将所述分割器应用于所述初始图像之前,

将所述聚集框定义为所述旁观者框,以及

对于所述定位器框中的每一个,计算所述聚集框与定位器框的交并比IoU比率,以及

响应于所述交并比比率超过阈值或响应于所述聚集框与所述定位器框之间的重叠面积超过占用阈值,增加所述聚集框的大小。

3.根据权利要求2所述的方法,其中所述IoU比率是所述聚集框与所述定位器框之间的所述重叠面积除以所述聚集框与所述定位器框之间的并集面积,并且所述方法进一步包括:

填充所述聚集框。

4.根据权利要求1所述的方法,其中将所述分割器应用于所述初始图像包括将前景与背景分割,以将附着到所述旁观者的所述一个或多个对象与所述背景区分开。

5.根据权利要求1所述的方法,进一步包括:

确定所述初始图像的主体是否遮挡所述旁观者;以及

响应于确定所述初始图像的所述主体遮挡所述旁观者,

生成主体掩模并且更新所述旁观者掩模以移除所述主体掩模内的像素。

6.根据权利要求1所述的方法,其中所述分割器是经过训练的机器学习模型。

7.根据权利要求1所述的方法,其中所述定位器框包括所述初始图像中的主体、所述旁观者或对象中的一个或多个,并且所述方法进一步包括:

移除与所述主体相关联的对应定位器框。

8.一种计算机实现的方法,包括:

获得合成训练数据集,所述合成训练数据集被生成为包括训练图像,每个训练图像包括将对象叠加到背景图像上;以及

通过以下方式获得经过训练的机器学习模型:

提供所述合成训练数据集作为原始机器学习模型的输入;

输出用于训练图像的预测旁观者框;

将每个预测旁观者框与对应于训练图像的地面真实旁观者掩模进行比较;以及

基于所述比较来调整所述原始机器学习模型的一个或多个参数;

其中所述经过训练的机器学习模型被训练以接收初始图像并且输出旁观者掩模,所述

旁观者掩模包括聚集框内的像素的子集。

9.根据权利要求8所述的方法,其中所述经过训练的机器学习模型是包括多个层的神经网络,并且用所述经过训练的机器学习模型输出所述旁观者掩模包括:

第一层,所述第一层通过对于每个像素将所述像素分类为前景的一部分或背景的一部分中的一个来将图像分割成所述前景和所述背景;以及

输出层,所述输出层对于每个像素输出所述像素是旁观者掩模的一部分、主体掩模的一部分,还是两者都不是,其中在所述神经网络中所述第一层在所述输出层之前。

10.根据权利要求8所述的方法,进一步包括:

接收所述初始图像作为所述经过训练的机器学习模型的输入,所述初始图像具有包含旁观者和附着到所述旁观者的一个或多个对象的聚集框;以及

用所述经过训练的机器学习模型输出旁观者掩模,所述旁观者掩模包括所述聚集框内的像素的子集;以及

生成修复图像,其中所述旁观者掩模内的所有像素在所述修复图像中用与所述初始图像中的背景匹配的像素被替换。

11.根据权利要求10所述的方法,进一步包括在接收所述初始图像作为所述经过训练的机器学习模型的输入之前:

检测所述初始图像中的所述旁观者;

生成包括所述旁观者的旁观者框,其中所述旁观者的所有像素在所述旁观者框内;

生成包含所述旁观者和附着到所述旁观者的所述一个或多个对象的定位器框;以及

聚集所述旁观者框和所述定位器框中的一个或多个定位器框以形成所述聚集框。

12.根据权利要求11所述的方法,进一步包括:

将所述聚集框定义为所述旁观者框,以及

对于所述定位器框中的每一个,

计算所述聚集框与定位器框的交并比IoU比率,以及

响应于所述交并比比率超过阈值或响应于所述聚集框与所述定位器框之间的重叠面积超过占用阈值,增加所述聚集框的大小。

13.根据权利要求12所述的方法,其中所述IoU比率是所述聚集框与所述定位器框之间的所述重叠面积除以所述聚集框与所述定位器框之间的并集面积,并且所述方法进一步包括:

填充所述聚集框。

14.根据权利要求10所述的方法,其中输出所述旁观者掩模包括输出对应置信度值;以及响应于所述对应置信度值未能满足置信度阈值,在所述旁观者的边缘周围添加模糊。

15.一种计算装置,包括:

处理器;以及

耦合到所述处理器的存储器,其中指令存储在所述存储器上,所述指令在由所述处理器执行时使所述处理器执行操作,所述操作包括:

检测初始图像中的旁观者;

生成包括所述旁观者的旁观者框,其中所述旁观者的所有像素在所述旁观者框内;

生成包含所述旁观者和附着到所述旁观者的一个或多个对象的定位器框;

聚集所述旁观者框和所述定位器框中的一个或多个定位器框以形成聚集框;

基于所述聚集框将分割器应用于所述初始图像,以从所述初始图像分割所述旁观者和所述一个或多个对象以生成旁观者掩模,其中所述旁观者掩模包括所述聚集框内的像素的子集;以及

生成修复图像,其中所述旁观者掩模内的所有像素在所述修复图像中用与所述初始图像中的背景匹配的像素被替换。

16.根据权利要求15所述的计算装置,其中所述操作进一步包括:

在将所述分割器应用于所述初始图像之前,

将所述聚集框定义为所述旁观者框,以及

对于所述定位器框中的每一个,计算所述聚集框与定位器框的交并比IoU比率,以及

响应于所述交并比比率超过阈值或响应于所述聚集框与所述定位器框之间的重叠面积超过占用阈值,增加所述聚集框的大小。

17.根据权利要求16所述的计算装置,其中所述IoU比率是所述聚集框与所述定位器框之间的所述重叠面积除以所述聚集框与所述定位器框之间的并集面积,并且所述方法进一步包括:

填充所述聚集框。

18.根据权利要求15所述的计算装置,其中将所述分割器应用于所述初始图像包括将前景与背景分割,以将附着到所述旁观者的所述一个或多个对象与所述背景区分开。

19.根据权利要求15所述的计算装置,其中所述操作进一步包括:

确定所述初始图像的主体是否遮挡所述旁观者;以及

响应于确定所述初始图像的所述主体遮挡所述旁观者,

生成主体掩模并且更新所述旁观者掩模以移除所述主体掩模内的像素。

20.根据权利要求15所述的计算装置,其中所述分割器是经过训练的机器学习模型。

旁观者和附着对象移除

背景技术

[0001] 诸如图像(静态图像、具有选择性运动的图像等)和视频的视觉媒体项中的对象分类可以通过移除分散媒体项的焦点的旁观者来改进。然而,当旁观者附着到对象(例如,骑自行车/玩滑板、拿着包等的人)时,可能识别与旁观者相关联的像素,并且可能不能识别与对象相关联的像素。当这种分类用于图像编辑目的(例如,用匹配背景的像素替换旁观者)时,由于存在与对象相关联的像素,因此所得图像可能不令人满意。

[0002] 本文提供的背景描述是出于大体地呈现本公开的场境的目的。在本背景技术部分描述的范围内本发明人的工作以及在递交时可能没有资格作为现有技术的描述的方面既不明确地也不暗示地被承认为本公开的现有技术。

发明内容

[0003] 一种计算机实现的方法包括检测初始图像中的旁观者。所述方法进一步包括生成包括旁观者的旁观者框,其中旁观者的所有像素在旁观者框内。所述方法进一步包括生成包括旁观者的旁观者框,其中旁观者的所有像素在旁观者框内。所述方法进一步包括生成定位器框,所述定位器框包含旁观者和附着到旁观者的一个或多个对象。所述方法进一步包括聚集旁观者框和定位器框中的一个或多个定位器框以形成聚集框。所述方法进一步包括基于聚集框将分割器应用于初始图像,以从初始图像分割旁观者和一个或多个对象以生成旁观者掩模,其中旁观者掩模包括聚集框内的像素的子集。所述方法进一步包括修复图像,其中用与初始图像中的背景匹配的像素替换修复图像中旁观者掩模内的所有像素。在一个示例中,所述方法或所述方法的至少一个或若干步骤可以由机器学习模型执行或提供。机器学习模型可以是例如通过下面描述的训练方法获得的经过训练的机器学习模型。因此,机器学习模型可以执行以下步骤中的至少一个:生成定位器框、形成聚集框、应用分割器,以及可选地还生成修复图像。机器学习模型可以进一步执行下面描述的任何方法步骤。代替使用一个单一模型,也可以使用多个模型。

[0004] 在一些实施例中,所述方法进一步包括在将分割器应用于初始图像之前,将聚集框定义为旁观者框;对于每个定位器框,计算聚集框与定位器框的交并比(IoU)比率;以及响应于交并比比率超过阈值或响应于聚集框与定位器框之间的重叠面积超过占用阈值,增加聚集框的大小。在一些实施例中,其中IoU比率是聚集框与定位器框之间的重叠面积除以聚集框与定位器框之间的并集面积,并且所述方法进一步包括填充聚集框。在一些实施例中,将分割器应用于初始图像包括将前景与背景分割,以将附着到旁观者的一个或多个对象与背景区分开。在一些实施例中,所述方法进一步包括确定初始图像的主体是否遮挡旁观者,并响应于确定初始图像的主体遮挡旁观者,生成主体掩模并更新旁观者掩模以移除主体掩模内的像素。在一些实施例中,分割器是经过训练的机器学习模型。在一些实施例中,定位器框包括初始图像中的主体、旁观者或对象中的一个或多个,并且所述方法进一步包括移除与主体相关联的对应定位器框。

[0005] 在一些实施例中,一种计算机实现的方法(例如,还称为训练方法)包括:获得合成

训练数据集,所述合成训练数据集被生成为包括训练图像,每个训练图像包括将对象叠加到背景图像上;以及通过以下方式获得经过训练的机器学习模型:提供合成训练数据集作为原始机器学习模型的输入;输出用于训练图像的预测旁观者框;将每个预测旁观者框与对应于训练图像的地面真实旁观者掩模进行比较;以及基于比较来调整原始机器学习模型的一个或多个参数,其中训练经过训练的机器学习模型以接收初始图像并且输出旁观者掩模,所述旁观者掩模包括聚集框内的像素的子集。

[0006] 在一些实施例中,经过训练的机器学习模型是包括多个层的神经网络,并且用经过训练的机器学习模型输出旁观者掩模包括:第一层,所述第一层通过对于每个像素将像素分类为以下中的一个来将图像分割成前景和背景:前景的一部分或背景的一部分;以及输出层,所述输出层对于每个像素输出该像素是旁观者掩模的一部分、主体掩模的一部分,还是两者都不是,其中在神经网络中,第一层在输出层之前。在一些实施例中,所述方法进一步包括接收具有聚集框的初始图像作为经过训练的机器学习模型的输入,所述聚集框包含旁观者和附着到旁观者的一个或多个对象;用经过训练的机器学习模型输出包括聚集框内的像素的子集的旁观者掩模;以及生成修复图像,其中用与初始图像中的背景匹配的像素替换修复图像中旁观者掩模内的所有像素。在一些实施例中,所述方法进一步包括在接收初始图像作为经过训练的机器学习模型的输入之前:检测初始图像中的旁观者;生成包括旁观者的旁观者框,其中旁观者的所有像素在旁观者框内;生成包含旁观者和附着到旁观者的一个或多个对象的定位器框;以及聚集旁观者框和定位器框中的一个或多个定位器框以形成聚集框。在一些实施例中,所述方法进一步包括将聚集框定义为旁观者框;对于定位器框中的每一个,计算聚集框与定位器框的交并比(IoU)比率;以及响应于交并比比率超过阈值或响应于聚集框与定位器框之间的重叠面积超过占用阈值,增加聚集框的大小。在一些实施例中,IoU比率是聚集框与定位器框之间的重叠面积除以聚集框与定位器框之间的并集面积,并且所述方法进一步包括填充聚集框。在一些实施例中,输出旁观者掩模包括输出对应置信度值;以及响应于对应置信度值未能满足置信度阈值,在旁观者的边缘周围添加模糊。

[0007] 在一些实施例中,一种计算装置包括一个或多个处理器和耦合到一个或多个处理器的存储器,其中指令存储在存储器上,所述指令在由处理器执行时使处理器执行操作。所述操作可以包括检测初始图像中的旁观者;生成包括旁观者的旁观者框,其中旁观者的所有像素在旁观者框内;生成包含旁观者和附着到旁观者的一个或多个对象的定位器框;聚集旁观者框和定位器框中的一个或多个定位器框以形成聚集框;基于聚集框将分割器应用于初始图像,以从初始图像分割旁观者和一个或多个对象以生成旁观者掩模,其中旁观者掩模包括聚集框内的像素的子集;以及生成修复图像,其中用与初始图像中的背景匹配的像素替换修复图像中旁观者掩模内的所有像素。

[0008] 在一些实施例中,所述操作进一步包括在将分割器应用于初始图像之前,将聚集框定义为旁观者框;对于定位器框中的每一个,计算聚集框与定位器框的交并比(IoU)比率;以及响应于交并比比率超过阈值或响应于聚集框与定位器框之间的重叠面积超过占用阈值,增加聚集框的大小。在一些实施例中,IoU比率是聚集框与定位器框之间的重叠面积除以聚集框与定位器框之间的并集面积,并且所述操作进一步包括填充聚集框。在一些实施例中,将分割器应用于初始图像包括将前景与背景分割,以将附着到旁观者的一个或多

个对象与背景区分开。在一些实施例中,所述操作进一步包括确定初始图像的主体是否遮挡旁观者,并响应于确定初始图像的主体遮挡旁观者,生成主体掩模并更新旁观者掩模以移除主体掩模内的像素。在一些实施例中,分割器是经过训练的机器学习模型。

[0009] 本说明书中描述的技术通过以下方式有利地解决当对象附着到旁观者时识别对象以从图像移除的问题:生成包括旁观者和附着到旁观者的一个或多个对象的聚集框、包括聚集框内的像素的子集的旁观者掩模,以及生成用与初始图像中的背景匹配的像素替换旁观者掩模内的像素的修复图像。在一些实施例中,在初始图像的主体遮挡旁观者的情况下,本申请描述用于生成主体掩模的技术,其中作为主体掩模的一部分的像素保留在最终图像中,所述最终图像基于不包括旁观者掩模中的主体像素的像素掩模生成。所描述的技术可以自动地执行图像编辑以移除旁观者和附着对象,或者执行诸如模糊旁观者和附着对象等的其它动作以产生最终图像。

[0010] 因此,本公开的方法解决各种技术问题。例如,所述方法可以提供移除了旁观者的图像,这可以改善对初始图像的主体的更可靠和/或精确的检测(例如,通过使用具有相应能力的相同模型或另一模型)。在另一示例中,所提供的移除了旁观者的图像可以用作用于训练另一机器学习模型(例如,用于人检测或图像分类的机器学习模型)的优化数据集。可能会优化数据集,因为可能会在训练任务中混淆模型的任何令人困惑的旁观者都会被移除。

附图说明

[0011] 图1是根据本文所描述的一些实施例的示例网络环境的框图。

[0012] 图2是根据本文所描述的一些实施例的示例计算装置的框图。

[0013] 图3A图示根据本文所描述的一些实施例的示例初始图像。

[0014] 图3B图示根据本文所描述的一些实施例的具有旁观者框和定位器框的示例初始图像。

[0015] 图3C图示根据本文所描述的一些实施例的具有框聚集器输出的示例初始图像。

[0016] 图3D图示根据本文所描述的一些实施例的具有聚集框和旁观者框的示例初始图像。

[0017] 图3E图示根据一些实施例的具有包含旁观者和附着对象的旁观者掩模的示例初始图像。

[0018] 图3F图示根据本文所描述的一些实施例的其中移除了旁观者和对象的示例最终图像。

[0019] 图4图示根据本文所描述的一些实施例的用于生成移除了旁观者和附着对象的修复图像的示例流程图。

[0020] 图5图示根据本文所描述的一些实施例的用于生成旁观者掩模的示例流程图。

[0021] 图6图示根据本文所描述的一些实施例的示例神经网络的操作。

具体实施方式

[0022] 示例环境100

[0023] 图1是图示示例环境100的框图。在一些实施例中,环境100包括耦合到网络105的

媒体服务器101、用户装置115a和用户装置115n。用户125a、125n可以与相应的用户装置115a、115n相关联。在一些实施例中,环境100可以包括图1中未示出的其它服务器或装置。在图1和其余图中,在附图标号之后有字母(例如“115a”)表示对具有该特定附图标号的元件的引用。文本中没有后面字母的附图标号(例如“115”)表示对标有该附图标号的元件的实施例的一般引用。

[0024] 媒体服务器101可以包括处理器、存储器和网络通信硬件。在一些实施例中,媒体服务器101是硬件服务器。媒体服务器101经由信号线102通信地耦合到网络105。信号线102可以是有线连接,诸如以太网、同轴线缆、光纤线缆等,或无线连接,诸如Wi-Fi®、Bluetooth®或其它无线技术。在一些实施例中,媒体服务器101经由网络105向用户装置115a、115n中的一个或多个发送数据并从用户装置115a、115n中的一个或多个接收数据。媒体服务器101可以包括媒体应用103a和数据库199。

[0025] 数据库199可以存储机器学习模型、训练数据集、图像等。数据库199还可以存储与用户125相关联的社交网络数据、用户125的用户偏好等。

[0026] 用户装置115可以是包括耦合到硬件处理器的存储器的计算装置。例如,用户装置115可以包括移动装置、平板计算机、移动电话、可穿戴装置、头戴式显示器、移动电子邮件装置、便携式游戏机、便携式音乐播放器、读取器装置或能够访问网络105的另一电子装置。

[0027] 在所图示的实施方案中,用户装置115a经由信号线108耦合到网络105,并且用户装置115n经由信号线110耦合到网络105。媒体应用103可以被存储为用户装置115a上的媒体应用103b和/或用户装置115n上的媒体应用103c。信号线108和110可以是有线连接,诸如以太网、同轴线缆、光纤线缆等,或无线连接,诸如Wi-Fi®、Bluetooth®或其它无线技术。用户装置115a、115n分别由用户125a、125n访问。作为示例,使用图1中的用户装置115a、115n。尽管图1图示了两个用户装置115a和115n,但是本公开适用于具有一个或多个用户装置115的系统架构。

[0028] 媒体应用103可以被存储在媒体服务器101或用户装置115上。在一些实施例中,本文所描述的操作在媒体服务器101或用户装置115上执行。在一些实施例中,一些操作可以在媒体服务器101上执行或一些可以在用户装置115上执行。操作的执行是根据用户设置。例如,用户125a可以指定要在其相应装置115a上而不在服务器101上执行操作的设置。通过此类设置,本文所描述的操作(例如,参考图4到6)完全在用户装置115a上执行,并且在媒体服务器101上不执行任何操作。此外,用户125a可以指定用户的图像和/或其它数据要仅本地存储在用户装置115a上而不存储在媒体服务器101上。通过此类设置,没有用户数据被传输到媒体服务器101或存储在媒体服务器101上。向媒体服务器101传输用户数据、由媒体服务器101对此类数据的任何临时或永久存储以及由媒体服务器101对此类数据执行操作仅当用户已经同意传输、存储和由媒体服务器101执行操作时才被执行。向用户提供随时改变设置的选项,例如使得用户可以启用或停用媒体服务器101的使用。

[0029] 机器学习模型(例如,神经网络或其它类型的模型)如果用于一个或多个操作则在具有特定用户许可的情况下在用户装置115上本地存储和使用。只有在用户允许的情况下才使用服务器端模型。使用合成数据集来执行模型训练,如下面参考图5所描述。此外,可以提供经过训练的模型以供在用户装置115上使用。在此类使用期间,如果用户125允许,则可以执行模型的装置上训练。如果用户115允许,则可以将更新的模型参数传输到媒体服务器

101,例如,以实现联邦学习。模型参数不包括任何用户数据。

[0030] 媒体应用103接收图像。例如,媒体应用103从作为用户装置115的一部分的相机接收图像,或媒体应用103通过网络105接收图像。媒体应用103检测初始图像中的一个或多个旁观者,并且生成包括旁观者的旁观者框,其中旁观者的所有像素在旁观者框内。媒体应用103进一步生成包含附着到旁观者的一个或多个对象的一个或多个定位器框,并且聚集旁观者框和一个或多个定位器框以形成聚集框。

[0031] 媒体应用103基于聚集框将分割器应用于初始图像,以从初始图像分割旁观者和一个或多个对象以生成旁观者掩模,其中旁观者掩模包括聚集框内的像素的子集。媒体应用103生成用与初始图像中的背景匹配的像素替换旁观者掩模内的所有像素的修复图像。

[0032] 在一些实施例中,可以使用包括中央处理单元(CPU)、现场可编程门阵列(FPGA)、专用集成电路(ASIC)、机器学习处理器/协处理器、任何其它类型的处理器或其组合的硬件来实现媒体应用103。在一些实施例中,可以使用硬件和软件的组合来实施媒体应用103a。

[0033] 示例计算装置200

[0034] 图2是可以用于实现本文中所描述的一个或多个特征的示例计算装置200的框图。计算装置200可以是任何合适的计算机系统、服务器或其它电子或硬件装置。在一个示例中,计算装置200是用于实现媒体应用103a的媒体服务器101。在另一示例中,计算装置200是用户装置115。

[0035] 在一些实施例中,计算装置200包括全部经由总线218耦合的处理器235、存储器237、输入/输出(I/O)接口239、显示器241、相机243和存储装置245。处理器235可以经由信号线222耦合到总线218,存储器237可以经由信号线224耦合到总线218,I/O接口239可以经由信号线226耦合到总线218,显示器241可以经由信号线228耦合到总线218,相机243可以经由信号线230耦合到总线218,并且存储装置245可以经由信号线232耦合到总线218。

[0036] 处理器235可以是一个或多个处理器和/或处理电路,以执行程序代码并控制计算装置200的基本操作。“处理器”包括处理数据、信号或其它信息的任何合适的硬件系统、机制或组件。处理器可以包括具有带有一个或多个核(例如,在单核、双核或多核配置中)的通用中央处理单元(CPU)、多个处理单元(例如,在多处理器配置中)、图形处理单元(GPU)、现场可编程门阵列(FPGA)、专用集成电路(ASIC)、复杂可编程逻辑器件(CPLD)、用于实现功能的专用电路系统、用于实现基于神经网络模型的处理的专用处理器、神经回路、针对矩阵计算(例如,矩阵乘法)优化的处理器的系统或其它系统。在一些实施例中,处理器235可以包括实现神经网络处理的一个或多个协处理器。在一些实施例中,处理器235可以是处理数据以产生概率输出的处理器,例如,由处理器235产生的输出可能不精确,或可能在距预期输出的范围内准确。处理不需要限于特定地理位置或具有时间限制。例如,处理器可以实时、离线、以批处理模式等执行其功能。处理的部分可以由不同(或相同)处理系统在不同时间处和在不同位置处执行。计算机可以是与存储器通信的任何处理器。

[0037] 存储器237通常被设置在计算装置200中以供处理器235存取,并且可以是适合于存储指令以供处理器或处理器集合执行并且与处理器235分开定位和/或与处理器235集成的任何合适的处理器可读存储介质,诸如随机存取存储器(RAM)、只读存储器(ROM)、电可擦除只读存储器(EEPROM)、快闪存储器等。存储器237可以存储由处理器235在计算装置200上操作的软件,包括媒体应用103。

[0038] 存储器237可以包括操作系统262、其它应用264和应用数据266。其它应用264可以包括例如图像库应用、图像管理应用、图像图库应用、通信应用、web托管引擎或应用、媒体共享应用等。本文所公开的一种或多种方法可以在若干环境和平台中运行,例如,作为可以在任何类型的计算装置上运行的独立计算机程序、作为具有web页的web应用、作为在移动计算装置上运行的移动应用(“应用”)等。

[0039] 应用数据266可以是由计算装置200的其它应用264或硬件生成的数据。例如,应用数据266可以包括由图像库应用使用的图像以及由其它应用264(例如,社交网络应用)识别的用户动作等。

[0040] I/O接口239可以提供用于使计算装置200能够与其它系统和装置接口连接的功能。接口连接装置可以被包括作为计算装置200的一部分,或可以是独立的并且与计算装置200通信。例如,网络通信装置、存储装置(例如,存储器237和/或存储装置245)和输入/输出装置可以经由I/O接口239通信。在一些实施例中,I/O接口239可以连接到接口装置,诸如输入装置(键盘、定点装置、触摸屏、麦克风、扫描仪、传感器等)和/或输出装置(显示装置、扬声器装置、打印机、监视器等)。

[0041] 可以连接到I/O接口239的接口连接装置的一些示例可以包括显示器241,所述显示器可以用于显示内容,例如图像、视频和/或如本文所述的输出应用的用户界面,并且接收来自用户的触摸(或手势)输入。例如,显示器241可以用于显示包括取景器上的图形指南的用户界面。显示器241可以包括任何合适的显示装置,诸如液晶显示器(LCD)、发光二极管(LED)或等离子显示屏、阴极射线管(CRT)、电视机、监视器、触摸屏、三维显示屏或其它视觉显示装置。例如,显示器241可以是设置在移动装置上的平面显示屏、嵌入在眼镜形状因数或头戴式耳机装置中的多个显示屏,或用于计算机装置的监视器屏幕。

[0042] 相机243可以是可以捕获图像和/或视频的任何类型的图像捕获装置。在一些实施例中,相机243捕获I/O接口239传输到媒体应用103的图像或视频。

[0043] 存储装置245存储与媒体应用103相关的数据。例如,存储装置245可以存储训练数据集,所述训练数据集包括标记图像、机器学习模型、来自机器学习模型的输出等。

[0044] 图2图示存储在存储器237中的示例媒体应用103,所述媒体应用包括边界框模块202、分割器204、机器学习模块206、修复器模块208和用户界面模块210。

[0045] 边界框模块202生成边界框。在一些实施例中,边界框模块202包括可由处理器235执行以生成边界框的一组指令。在一些实施例中,边界框模块202存储在计算装置200的存储器237中并且可由处理器235存取和执行。

[0046] 在一些实施例中,边界框模块202接收初始图像。可以经由I/O接口239从计算装置200的相机243或从媒体服务器101接收初始图像。初始图像包括主体,诸如人。边界框模块202检测初始图像中的旁观者。旁观者为不是初始图像的主体的人,诸如步行、跑步、骑自行车、站在主体后面或以其他方式在初始图像内的人。在不同示例中,旁观者可以在前景中(例如,一个人在相机前面穿过)、在与主体相同的深度处(例如,站在主体旁边的人),或在背景中。在一些示例中,初始图像中可能有多于一个旁观者。旁观者可以是处于任意姿势,例如站着、坐着、蹲下、躺下、跳跃等的人。旁观者可以面对相机,可以与相机成一定角度,或可以背对相机。例如,图3A图示主体和旁观者的示例初始图像300。主体是一名行走的女子,并且旁观者是一名骑着滑板车的男子。

[0047] 边界框模块202生成包括旁观者的旁观者框。在一些实施例中,旁观者框是包含旁观者的所有像素的矩形边界框。边界框模块202可以通过执行对象识别,将对象与人的对象先验相比较并且丢弃不是人的对象来检测旁观者。在一些实施例中,边界框模块202使用诸如神经网络或更具体地卷积神经网络的机器学习算法来识别旁观者并且生成旁观者框。旁观者框与边界框的x和y坐标相关联。

[0048] 初始图像中的旁观者附着到对象。例如,旁观者可以骑着自行车、站在包旁边或拿着包、拿着伞、骑着滑板车等。边界框模块202生成初始图像中的对象的边界框。对象可以包括非人,诸如钱包、婴儿车、滑板车等,并且也可以包括人和/或人的部分,诸如人的脸和人的裤子。对象还可以包括宠物、固定对象(例如灯柱、垃圾桶等)或其它类型的对象。当边界框与方位(例如,边界框的x和y坐标、边界框的中心的x和y坐标,以及用于指示到边界框的边缘的距离的y变化和x变化等)和对象的类型的注释相关联时,边界框可以被称为定位器框。

[0049] 边界框模块202生成定位器框,所述定位器框包含旁观者和附着到初始图像中的旁观者的一个或多个对象。例如,图3B图示具有如由滑板车上的男子周围的虚线框所图示的旁观者框和定位器框的示例初始图像310。各种定位器框包含对象,诸如旁观者佩戴的斜挎包和衣物,诸如旁观者身上的裤子和主体身上的裤子312,以及初始图像的旁观者的脸和主体的脸314。

[0050] 边界框模块202移除与初始图像的主体相关联的定位器框,所述定位器框是图3B中的定位器框312和314。图3C图示其中移除与初始图像的主体相关联的定位器框的示例初始图像320。

[0051] 边界框模块202聚集旁观者框和定位器框以形成聚集框。边界框模块202可以确定聚集框是否足够大以包含旁观者和对象。在一些实施例中,边界框模块202将聚集框的区域与每个定位器框的区域相比较。例如,参考图3C,边界框模块202将旁观者周围的虚线框的区域与包含旁观者、斜跨包和滑板车的实线框322的区域相比较;将虚线框的区域与包含滑板车的实线框323的区域相比较;以及将虚线框的区域与包含斜跨包的实线框324的区域相比较。

[0052] 边界框模块202计算交并比(IoU)比率,所述IoU比率是聚集框与每个定位器框之间的重叠面积除以聚集框与每个定位器框之间的并集面积之间的比率。对于每个定位器框,如果IoU比率超过阈值或如果聚集框与每个定位器框之间的重叠面积超过占用阈值,编辑框模块202增加聚集框的大小,直到满足阈值或占用阈值。

[0053] 在一些实施例中,一旦最后确定聚集框的大小,边界框模块202就填充聚集框。例如,边界框模块202用聚集框外部的预定数量的像素(例如,10个像素)填充聚集框,以确保聚集框是包含性的。在一些实施方案中,填充可以在聚集框的所有四侧上等同。在一些实施方案中,可以在不同侧上使用不同数量的填充(例如,左侧上0个像素、顶部和底部上5个像素、右侧上10个像素等)。图3D图示具有聚集框332和旁观者框的示例初始图像330。聚集框332大于以虚线的旁观者框(其包含旁观者),因为它包含旁观者和附着对象——斜挎包和滑板车。

[0054] 在一些实施例中,媒体应用103包括分割器204,所述分割器生成旁观者掩模。在一些实施例中,媒体应用103包括机器学习模型206,所述机器学习模型生成分割初始图像并

且输出旁观者掩模的经过训练的机器学习模型。在一些实施例中,当旁观者与初始图像的主体之间不存在重叠时,使用分割器204,并且当旁观者与主体之间存在重叠时,使用机器学习模块206。在一些实施例中,分割器204还采用机器学习技术或与机器学习模块206结合工作。

[0055] 分割器204分割初始图像的聚集框内的旁观者和一个或多个对象,以生成旁观者掩模。在一些实施例中,分割器204包括可由处理器235执行以生成旁观者掩模的一组指令。在一些实施例中,分割器204被存储在计算装置200的存储器237中并且可由处理器235存取和执行。

[0056] 从初始图像生成包含旁观者和一个或多个对象的旁观者掩模,使得从初始图像移除旁观者和对象。当旁观者附着到一个或多个对象时,此过程更困难,因为难以识别附着到旁观者的对象。在一些实施例中,基于在输入图像上生成超像素并将超像素质心与深度图值匹配以基于深度进行聚类检测来生成旁观者掩模。更具体来说,可以使用掩蔽区域中的深度值来确定深度范围,并且可以识别落在深度范围内的超像素。然而,附着到旁观者的对象的深度与地面相同,这可能会导致过度包含的旁观者掩模。

[0057] 用于生成旁观者掩模的另一技术包括基于深度值与旁观者掩模的接近程度来加权深度值,其中权重由距离变换图表示。然而,旁观者掩模可能对地面过度包含,又对一些对象不够包含。

[0058] 在一些实施例中,分割器204通过接收包括聚集框的初始图像并将前景与背景分割以将附着到旁观者的对象与背景(诸如人行道)区分开来解决此类技术问题。分割器204使用聚集框内的旁观者和附着到旁观者的对象的像素生成旁观者掩模。用于旁观者掩模的像素的子集小于聚集框内的像素。

[0059] 在一些实施例中,生成旁观者掩模的任务因主体遮挡旁观者的一部分而变得更加复杂。例如,返回到图3A,主体的右臂与附着到旁观者的滑板车重叠。如果分割器204在不考虑主体的情况下生成旁观者掩模,则旁观者掩模可以包括主体的一部分并且修复器模块可能移除主体手臂的一部分。在一些实施例中,分割器204确定初始图像的主体是否遮挡旁观者,并响应于确定初始图像的主体遮挡旁观者,生成主体掩模并更新旁观者掩模以移除主体掩模内的像素。

[0060] 机器学习模块206生成经过训练的模型,所述经过训练的模型在本文中称为机器学习模型。在一些实施例中,机器学习模块206被配置成将机器学习模型应用于输入数据,诸如应用数据266(例如,由用户装置115捕获的初始图像)以输出旁观者掩模。在一些实施例中,机器学习模块206可以包括将由处理器235执行的软件代码。在一些实施例中,机器学习模块206被存储在计算装置200的存储器237中并且可由处理器235存取和执行。

[0061] 在一些实施例中,机器学习模块206可以指定使处理器235能够应用机器学习模型的电路配置(例如,用于可编程处理器、用于现场可编程门阵列(FPGA)等)。在一些实施例中,机器学习模块206可以包括软件指令、硬件指令或组合。在一些实施例中,机器学习模块206可以提供应用编程接口(API),所述API可以由操作系统262和/或其它应用264使用以调用机器学习模块206,例如以将机器学习模型应用于应用数据266以输出分割掩码。

[0062] 机器学习模块206使用训练数据来生成经过训练的机器学习模型。例如,训练数据可以包括具有主体、旁观者、附着到旁观者的一个或多个对象、聚集框的图像,并且其中旁

观者和一个或多个对象被分割并且与旁观者掩模相关联。在一些实施方案中,可以针对训练数据中的每个图像手动地生成聚集框。

[0063] 可以从任何来源获得训练数据,所述来源例如是专门标记用于训练的数据存储库、为其提供用作机器学习的训练数据的许可的数据等。在一些实施例中,训练可以发生在直接向用户装置115提供训练数据的媒体服务器101上,训练本地发生在用户装置115上,或两者的组合。

[0064] 在一些实施例中,训练数据可以包括出于训练目的生成的合成数据,诸如不基于正在训练的场境中的活动的数据,例如从模拟或计算机生成的图像/视频生成的数据等。训练数据可以包括部分地遮挡旁观者的主体的合成图像。在一些实施例中,通过将二维对象或三维对象叠加到背景图像上来生成合成图像。可以从特定视角渲染三维对象以将三维对象转换为二维对象。在一些实施例中,通过将多个人和对象叠加到背景上以模拟复合或遮挡效果来生成合成图像。可以向合成图像添加不同的掩模,诸如人掩模和对象掩模,以训练原始机器学习模型来在合成图像中的人与对象之间进行区分。

[0065] 在一些实施例中,机器学习模型206使用从另一应用获取并且未被编辑/传递的权重。例如,在这些实施例中,可以例如在不同装置上生成经过训练的模型,并将经过训练的模型提供为媒体应用103的一部分。在各种实施例中,可以将经过训练的模型提供为数据文件,所述数据文件包括模型结构或形式(例如,其定义神经网络节点的数量和类型、节点之间的连接性以及节点到多个层的组织)以及相关联的权重。机器学习模块206可以读取经过训练的模型的数据文件,并且基于经过训练的模型中指定的模型结构或形式来实现具有节点连接性、层和权重的神经网络。

[0066] 经过训练的机器学习模型可以包括一个或多个模型形式或结构。例如,模型形式或结构可以包括任何类型的神经网络,诸如线性网络、实现多个层(例如,输入层与输出层之间的“隐藏层”,其中每个层是线性网络)的深度学习神经网络、卷积神经网络(例如,将输入数据拆分或分割成多个部分或瓦片(tile)、使用一个或多个神经网络层单独地处理每个瓦片并且聚集来自每个瓦片的处理的结果的网络)、序列到序列神经网络(例如,接收诸如句子中的单词、视频中的帧等的序列数据作为输入并产生结果序列作为输出的网络)等。

[0067] 模型形式或结构可以指定各种节点之间的连接性以及节点到层的组织。例如,第一层(例如,输入层)的节点可以接收数据作为输入数据或应用数据。例如,当经过训练的模型用于例如初始图像的分析时,此类数据可以包括例如每节点的一个或多个像素。后续中间层可以根据模型形式或结构中指定的连接性接收先前层的节点的输出作为输入。这些层也可以称为隐藏层。例如,第一层可以输出前景与背景之间的分割。最终层(例如,输出层)产生机器学习模型的输出。例如,输出层可以接收初始图像到前景和背景的分割,并且输出像素是旁观者掩模的一部分、主体掩模的一部分,还是两者都不是。在一些实施方案中,模型形式或结构还指定每个层中的节点的数量和/或类型。

[0068] 在不同实施方案中,经过训练的模型可以包括一个或多个模型。模型中的一个或多个可以包括多个节点,这些节点按照模型结构或形式布置成层。在一些实施方案中,节点可以是没有存储器的计算节点,例如,被配置为处理一个输入单元以产生一个输出单元。由节点执行的计算可以包括例如将多个节点输入中的每一个乘以权重,获得加权和,并且用偏差或截距值调整加权和以产生节点输出。在一些实施方案中,由节点执行的计算还可以

包括将阶跃/激活函数应用于调整后的加权和。在一些实施方案中,阶跃/激活函数可以是非线性函数。在各种实施方案中,此类计算可以包括诸如矩阵乘法的运算。在一些实施方案中,由多个节点进行的计算可以例如使用多核处理器的多个处理器核、使用图形处理单元(GPU)的单个处理单元或专用神经回路并行执行。在一些实施方案中,节点可以包括存储器,例如,可以能够在处理后续输入时存储和使用一个或多个更早输入。例如,具有存储器的节点可以包括长短期记忆(LSTM)节点。LSTM节点可以使用存储器来维持允许节点像有限状态机(FSM)一样运行的"状态"。

[0069] 在一些实施方案中,经过训练的模型可以包括单个节点的嵌入或权重。例如,可以将模型初始化为多个节点,这些节点被组织成如由模型形式或结构指定的层。在初始化时,可以将相应的权重应用于根据模型形式连接的每一对节点(例如神经网络的连续层中的节点)之间的连接。例如,相应权重可以被随机指配或初始化为默认值。然后可以例如使用训练数据来训练模型以产生结果。

[0070] 训练可以包括应用监督学习技术。在监督学习中,训练数据可以包括多个输入(例如,图像、聚集框、人掩模等)和每个输入的对应地面真实输出(例如,正确地识别每个图像的旁观者和附着对象的地面真实旁观者掩模)。基于模型的输出与地面真实输出的比较,例如以增加模型产生图像的地面真实输出的概率的方式自动地调整权重值。

[0071] 在各种实施方案中,经过训练的模型包括与模型结构相对应的一组权重或嵌入。在一些实施方案中,经过训练的模型可以包括一组固定的权重,例如,从提供权重的服务器下载的一组权重。在各种实施方案中,经过训练的模型包括与模型结构相对应的一组权重或嵌入。在省略数据的实施方案中,机器学习模块206可以例如由机器学习模块206的开发者、由第三方等生成基于先前训练的经过训练的模型。在一些实施方案中,经过训练的模型可以包括一组固定的权重,例如,从提供权重的服务器下载的一组权重。

[0072] 在一些实施例中,机器学习模块206接收具有主体、与一个或多个对象相关联的旁观者以及围绕旁观者和一个或多个对象的聚集框的初始图像。机器学习模块206将初始图像作为输入提供给经过训练的机器学习模型。在一些实施例中,经过训练的机器学习模型输出旁观者掩模,所述旁观者掩模包括聚集框内的像素的子集。图3E图示具有包含旁观者和附着对象的旁观者掩模(以实心黑色图示)的示例初始图像340。

[0073] 在一些实施例中,机器学习模型输出由经过训练的机器学习模型输出的每个旁观者掩模的置信度值。置信度值可以表示为百分比、从0到1的数字等。例如,对于旁观者掩模正确地包含旁观者和附着对象并且不包括来自主体的像素的置信度,机器学习模型输出85%的置信度值。在一些实施例中,当置信度值低于阈值时,修复器模块208在旁观者的边缘周围施加模糊或其它效果以最小化任何误差。

[0074] 在一些实施例中,机器学习模块206从用户装置115上的用户接收反馈。反馈可以采取用户的发布修复图像、删除修复图像、共享修复图像等的形式。机器学习模块206基于反馈来修正机器学习模型的参数。

[0075] 修复器模块208生成修复图像,所述修复图像用与初始图像中的背景匹配的像素替换旁观者掩模内的所有像素。在一些实施例中,修复器模块208包括可由处理器235执行以生成修复图像的一组指令。在一些实施例中,修复器模块208被存储在计算装置200的存储器237中并且可由处理器235存取和执行。

[0076] 在一些实施例中,修复器模块208从机器学习模块206接收具有旁观者掩模的初始图像。修复器模块208用与初始图像中的背景匹配的像素替换旁观者掩模内的所有像素。与背景匹配的像素可以是基于没有主体和/或旁观者的相同位置的另一图像。或者,修复器模块208可以基于围绕旁观者掩模的像素来匹配从旁观者掩模移除的像素。例如,在旁观者站在地面上的情况下,修复器模块208用地面的像素替换所述像素。图3F图示其中移除了旁观者和对象的示例修复图像350。其它修复技术也是可能的,包括基于机器学习的修复器技术。

[0077] 用户界面模块210生成用户界面。在一些实施例中,用户界面模块210包括可由处理器235执行以生成用户界面的一组指令。在一些实施例中,用户界面模块210被存储在计算装置200的存储器237中并且可由处理器235存取和执行。

[0078] 用户界面模块210生成包括修复图像的用户界面。在一些实施例中,用户界面包括用于编辑修复图像、共享修复图像、将修复图像添加到相册等的选项。

[0079] 示例流程图

[0080] 图4图示根据本文所描述的一些实施例的用于生成移除了旁观者和附着对象的修复图像的示例流程图400。流程图400中所图示的方法可以由图2中的计算装置200执行。在一些实施例中,流程图400由用户装置115、媒体服务器101执行,或部分地在用户装置115上并且部分地在媒体服务器101上执行。

[0081] 图4的方法400可以开始于框402。在框402处,检测初始图像中的旁观者。框402之后可以是框404。

[0082] 在框404处,生成包括旁观者的旁观者框,其中旁观者的所有像素在旁观者框内。框404之后可以是框406。

[0083] 在框406处,生成包含旁观者和附着到旁观者的一个或多个对象的定位器框。框406之后可以是框408。

[0084] 在框408处,聚集旁观者框和定位器框以形成聚集框。在一些实施例中,首先将聚集框定义为与旁观者框相同,然后基于应用IoU计算来添加特定定位器框。框408之后可以是框410。

[0085] 在框410处,基于聚集框将分割器应用于初始图像,以从初始图像分割旁观者和一个或多个对象以生成旁观者掩模,其中旁观者掩模包括聚集框内的像素的子集。框410之后可以是框412。

[0086] 在框412处,生成用与初始图像中的背景匹配的像素替换旁观者掩模内的所有像素的修复图像。

[0087] 图5图示根据本文所描述的一些实施例的用于移除具有对象的旁观者的示例流程图500。流程图500中所图示的方法可以由图2中的计算装置200执行。在一些实施例中,流程图500由用户装置115、媒体服务器101执行,或部分地在用户装置115上并且部分地在媒体服务器101上执行。

[0088] 图5的方法500可以开始于框502。流程图500中所图示的方法可以由图2中的计算装置200执行。在一些实施例中,流程图500由用户装置115、媒体服务器101执行,或部分地在用户装置115上并且部分地在媒体服务器101上执行。

[0089] 在框502处,合成训练数据集包括训练图像,每个训练图像包括附着到对象的旁观

者和主体。在一些实施方案中,在训练图像的至少子集中的主体遮挡旁观者和/或附着对象。对于监督学习,训练数据集还包括每个训练图像的地面真实旁观者掩模。框502之后可以是框504。

[0090] 在框504处,通过使用合成训练数据集作为输入训练原始机器学习模型来获得经过训练的机器学习模型,其中使用监督学习执行训练。在监督学习下,所述模型输出训练图像的预测旁观者框。将预测旁观者框与对应于训练图像的地面真实旁观者掩模进行比较。基于比较,提供反馈以调整训练下的模型(原始模型)的一个或多个参数。在一些实施方案中,可以使用梯度下降来调整参数。所述过程使用足够大的训练数据集执行,使得减少预测旁观者框与地面真实旁观者框之间的误差。当误差下降到阈值以下时或当已经利用整个训练数据集时,可以完成训练。训练可以用附加训练图像重复任意次数。此外,基于用户对修复的输出图像的反馈,例如,如果用户反馈指示修复图像不令人满意(例如,没有完全移除旁观者和/或附着对象),可以发起重新训练。经过训练的模型可以作为媒体应用103的一部分供用户装置115本地使用和/或存储用于在媒体服务器101上使用。框504之后可以是框506。

[0091] 在框506处,接收初始图像作为经过训练的机器学习模型的输入,所述初始图像具有围绕旁观者和附着到旁观者的一个或多个对象的聚集框。框506之后可以是框508。

[0092] 在框508处,用经过训练的机器学习模型输出旁观者掩模,所述旁观者掩模包括聚集框内的像素的子集。

[0093] 图6图示示例神经网络的操作。神经网络机器学习模型600包括第一层604和分割输出层608。尽管图示两个层,但是神经网络机器学习模型600可以包括更多层或更少层。

[0094] 提供初始图像602作为输入层604的输入。例如,初始图像602包括主体和附着到一个或多个对象的旁观者。通过对于每个像素将该像素分类为以下中的一个:初始图像602的前景的一部分或背景的一部分,第一层604将初始图像602分割为前景和背景606。

[0095] 提供初始图像602到前景和背景的分割606,作为分割输出层608的输入。分割输出层608输出像素是旁观者掩模的一部分、主体掩模的一部分,还是两者都不是610。

[0096] 除了以上描述之外,用户可以被提供有控件,所述控件允许用户选择是否以及何时本文所描述的系统、程序或特征可以实现用户信息(例如,关于用户的社交网络、社交行为或活动、职业、用户偏好或用户当前位置的信息)的收集,以及用户是否从服务器发送内容或通信。另外,某些数据可以在它被存储或使用之前通过一种或多种方式处理,使得去除个人可识别信息。例如,可以处理用户的身份,使得无法确定用户的个人可识别信息,或者可以在获得位置信息的情况下将用户的地理位置进行概括(诸如到城市、邮政编码或州级别),使得无法确定用户的特定位置。因此,用户可以控制收集关于用户的什么信息、如何使用该信息,以及向用户提供什么信息。

[0097] 在以上描述中,出于解释的目的,阐述了许多具体细节以便提供对本说明书的透彻理解。然而,对本领域技术人员将显而易见的是,可以在不具有这些具体细节的情况下实践本公开。在一些情况下,以框图形式示出结构和装置,以避免使描述模糊不清。例如,上文主要参考用户界面和特定硬件描述实施例。然而,实施例可以应用于可以接收数据和命令的任何类型的计算装置,以及提供服务的任何外围装置。

[0098] 说明书中对“一些实施例”或“一些实例”的引用意味着结合该实施例或实例描述

的特定特征、结构或特性可以被包括在本说明书的至少一个实施方案中。在说明书中各个地方出现的短语“在一些实施例中”未必都指代相同实施例。

[0099] 就计算机存储器内的数据比特上的操作的算法和符号表示方面来呈现以上详细描述的一些部分。这些算法描述和表示是数据处理领域的技术人员用于将其工作实质最有效地传达给本领域其他技术人员的手段。算法在此处通常被认为是得到期望结果的自相一致的步骤序列。这些步骤是需要对物理量进行物理操纵的步骤。尽管不是必需的,但是这些量通常采用能够进行存储、传递、组合、比较以及以其它方式操纵的电数据或磁数据的形式。有时,原则上出于通常使用的原因,将这些数据称为比特、值、元素、符号、字符、术语、数字等被证明是方便的。

[0100] 然而,应当记住,所有这些和类似术语与适当物理量相关联,并且仅仅是应用于这些量的方便标记。除非特别指出,否则从下面的论述中明显看出,应理解,在整个说明书中,利用包括“处理”或“计算”或“运算”或“确定”或“显示”等的术语的论述可以指计算机系统或类似电子计算装置的动作和过程,所述计算机系统或类似电子计算装置操纵在计算机系统的寄存器和存储器内表示为物理(电子)量的数据并将所述数据转换成在计算机系统存储器或寄存器或其它此类信息存储、传输或显示装置内类似地表示为物理量的其它数据。

[0101] 本说明书的实施例还可以涉及用于执行上述方法的一个或多个步骤的处理器。处理器可以是由存储在计算机中的计算机程序选择性地激活或重新配置的专用处理器。此类计算机程序可以被存储在非暂时性计算机可读存储介质中,非暂时性计算机可读存储介质包括但不限于任何类型的盘,包括光盘、ROM、CD-ROM、磁盘、RAM、EPROM、EEPROM、磁卡或光卡、包括具有非易失性存储器的USB密钥的快闪存储器,或适合于存储电子指令的任何类型的介质,每一者耦合到计算机系统总线。

[0102] 本说明书可以采用一些完全硬件实施例、一些完全软件实施例或包含硬件和软件元件两者的一些实施例的形式。在一些实施方案中,本说明书在软件中实现,所述软件包括但不限于固件、驻存软件、微代码等。

[0103] 此外,本说明书可以采用计算机程序产品的形式,所述计算机程序产品可从计算机可用或计算机可读介质存取,所述计算机可用或计算机可读介质提供用于由计算机或任何指令执行系统使用或与计算机或任何指令执行系统结合使用的程序代码。出于本说明书的目的,计算机可用或计算机可读介质可以是可以包含、存储、传送、传播或传输程序以供指令执行系统、设备或装置使用或结合指令执行系统、设备或装置使用的任何设备。

[0104] 适合于存储或执行程序代码的数据处理系统将包括直接地或通过系统总线间接地耦合到存储器元件的至少一个处理器。存储器元件可以包括在程序代码的实际执行期间采用的本地存储器、大容量存储装置以及提供至少一些程序代码的临时存储以便减少在执行期间必须从大容量存储装置检索代码的次数的高速缓冲存储器。

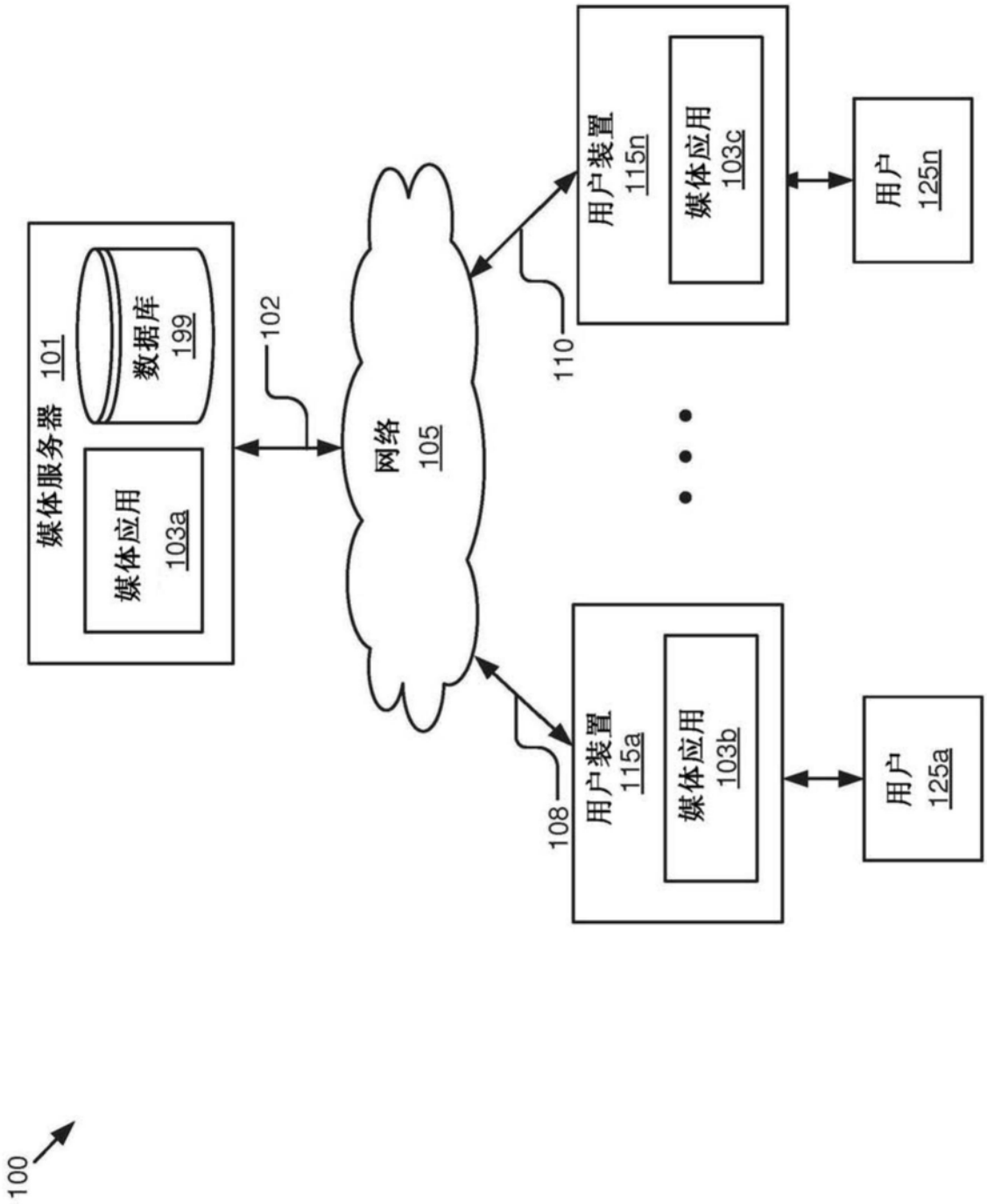

媒体服务器 101
数据库 199
媒体应用 103a
102
网络 105
用户装置 115n
媒体应用 103c
用户 125n
110
用户装置 115a
媒体应用 103b
用户 125a
108
100

图1

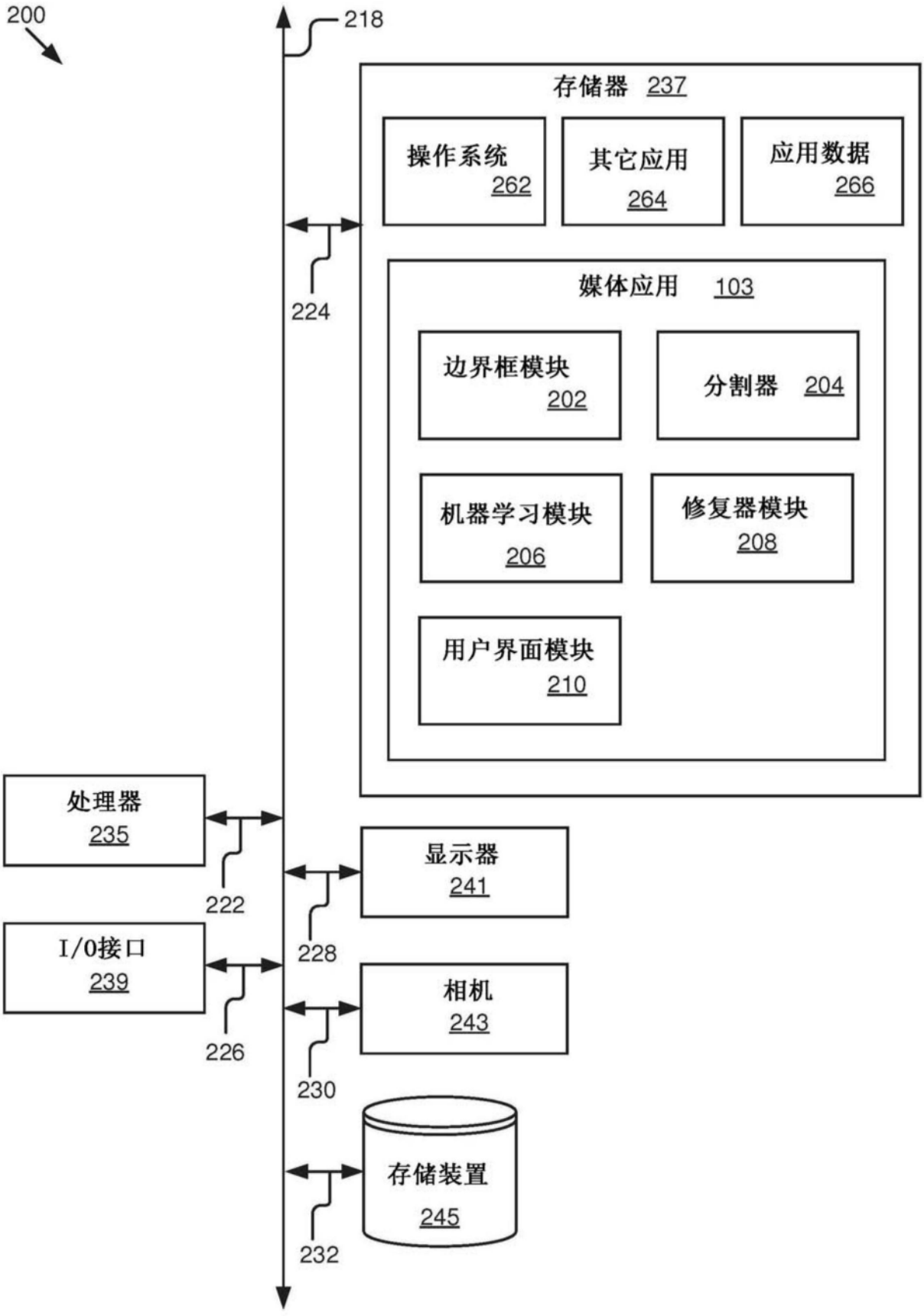

200
218
存储器 237
操作系统 262
其它应用 264
应用数据 266
224
媒体应用 103
边界框模块 202
分割器 204
机器学习模块 206
修复器模块 208
用户界面模块 210
处理器 235
222
显示器 241
228
I/O接口 239
226
相机 243
230
存储装置 245
232

图2

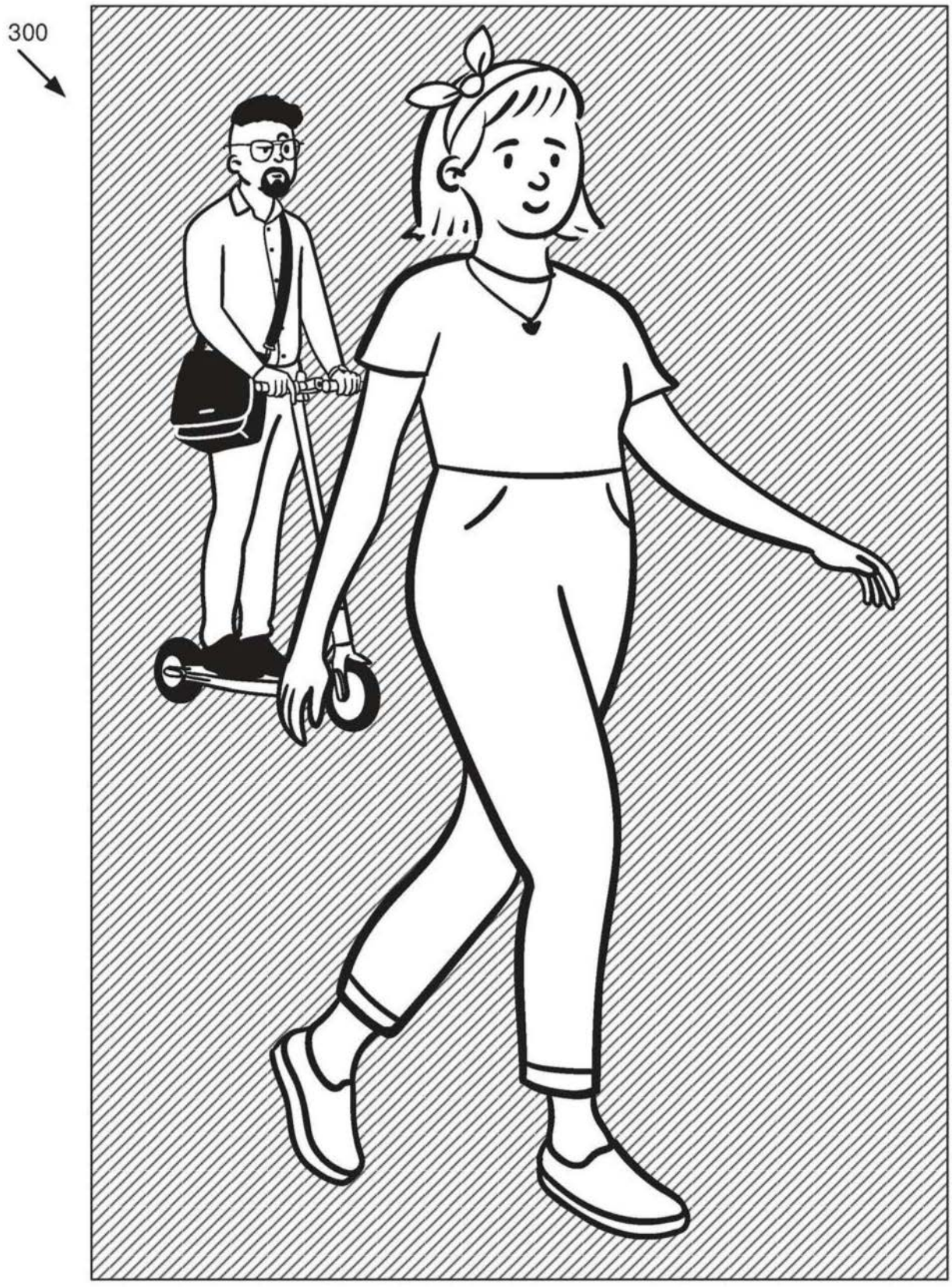


图3A

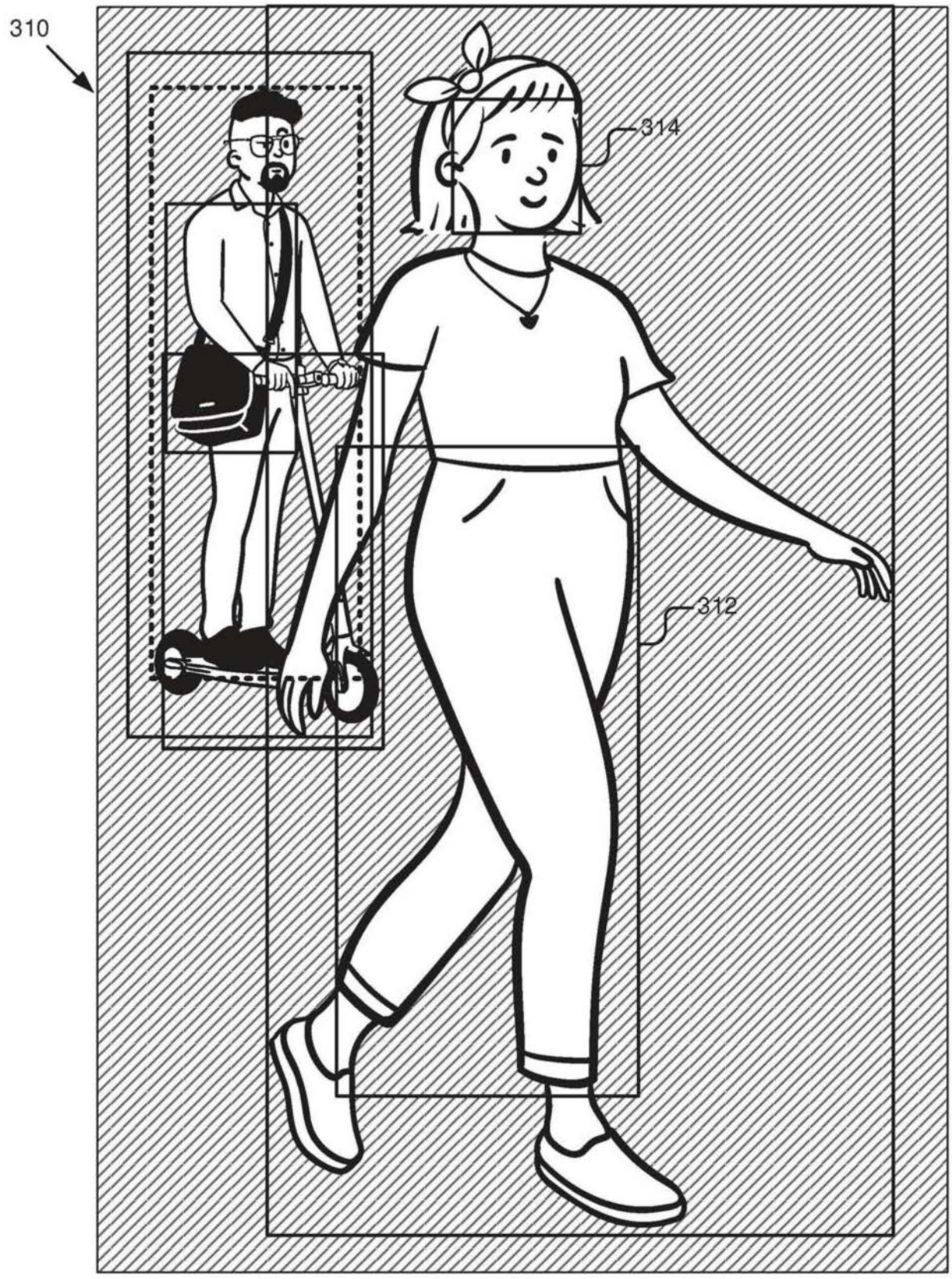


图3B

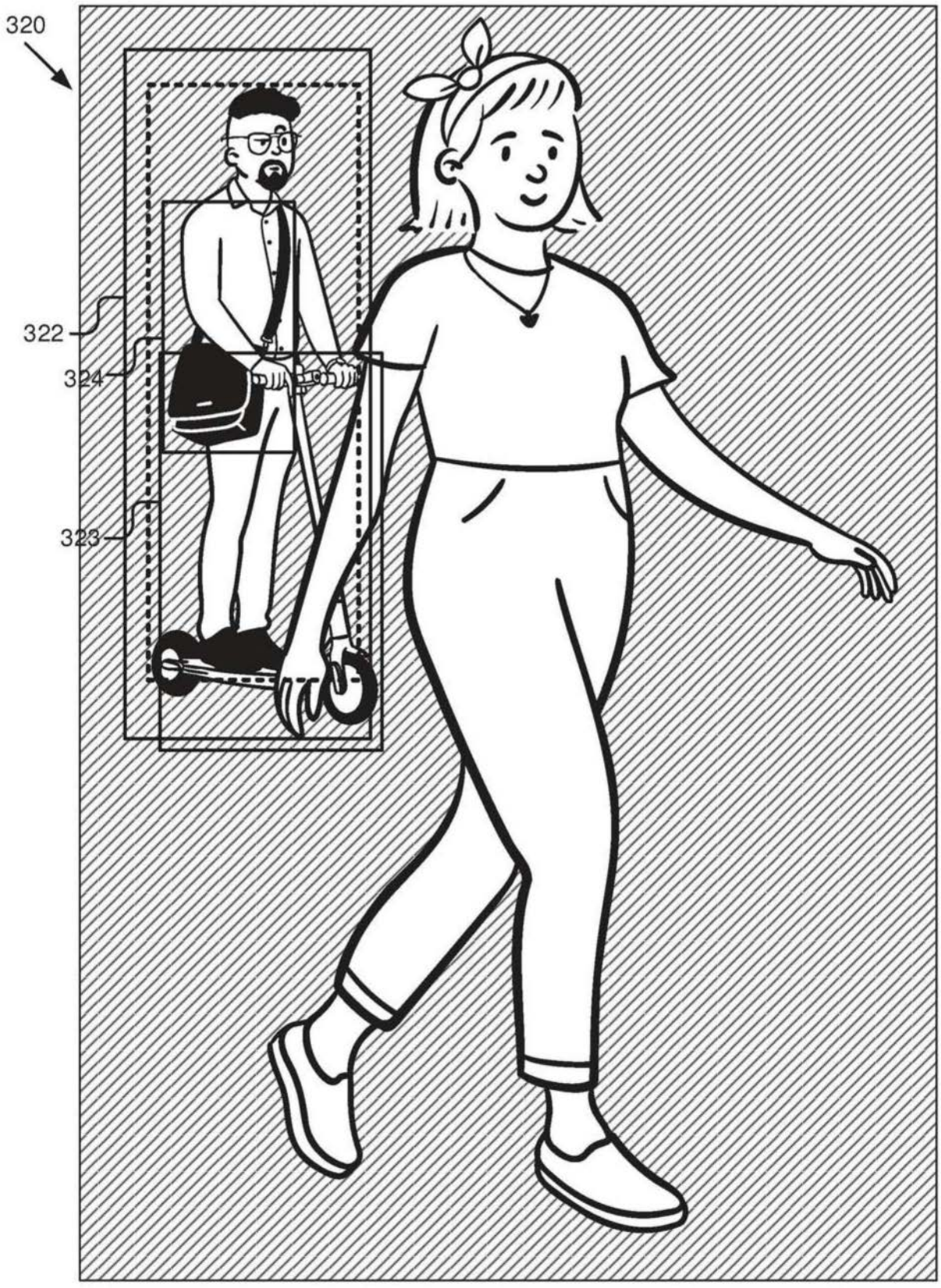


图3C

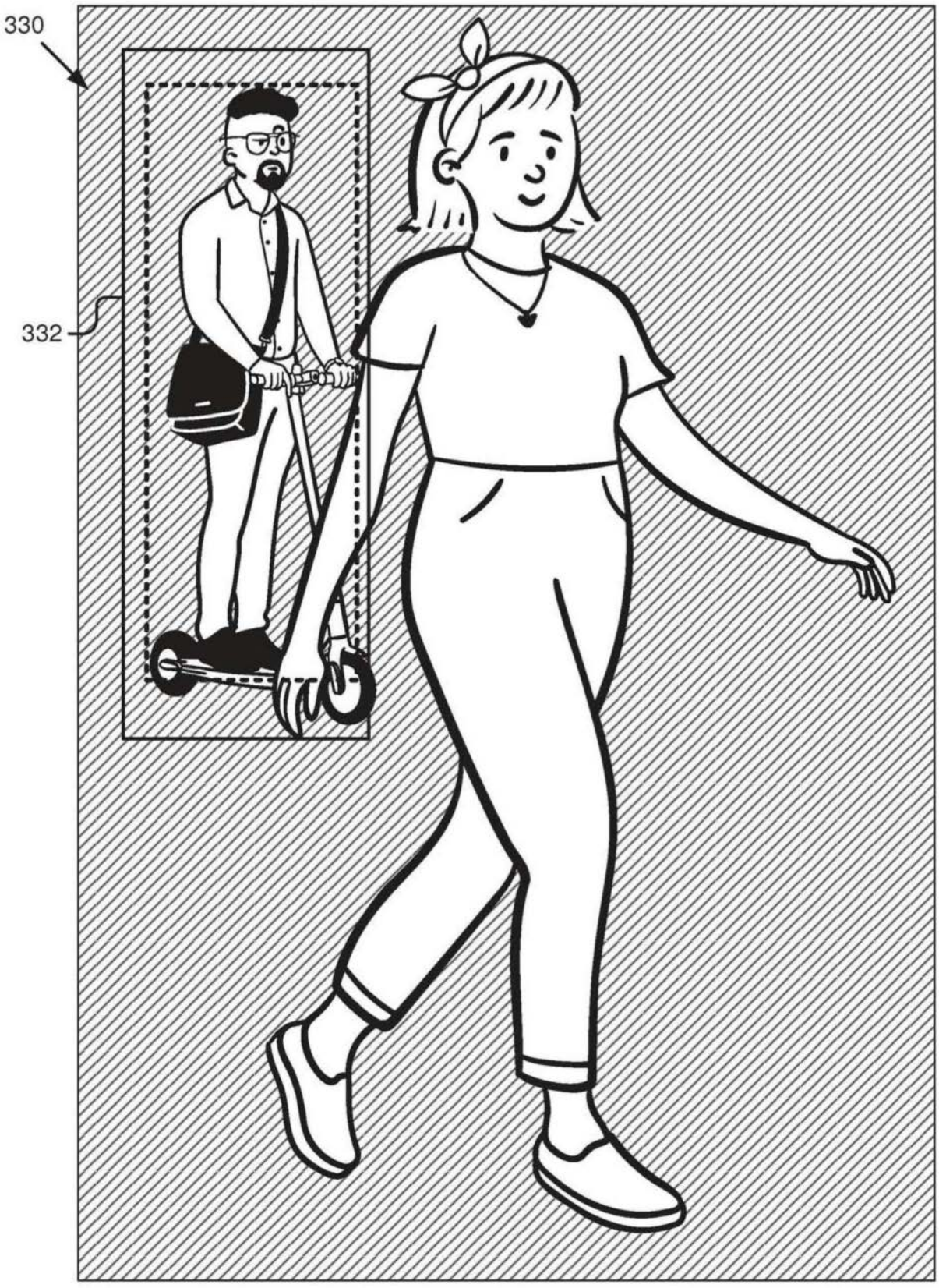


图3D

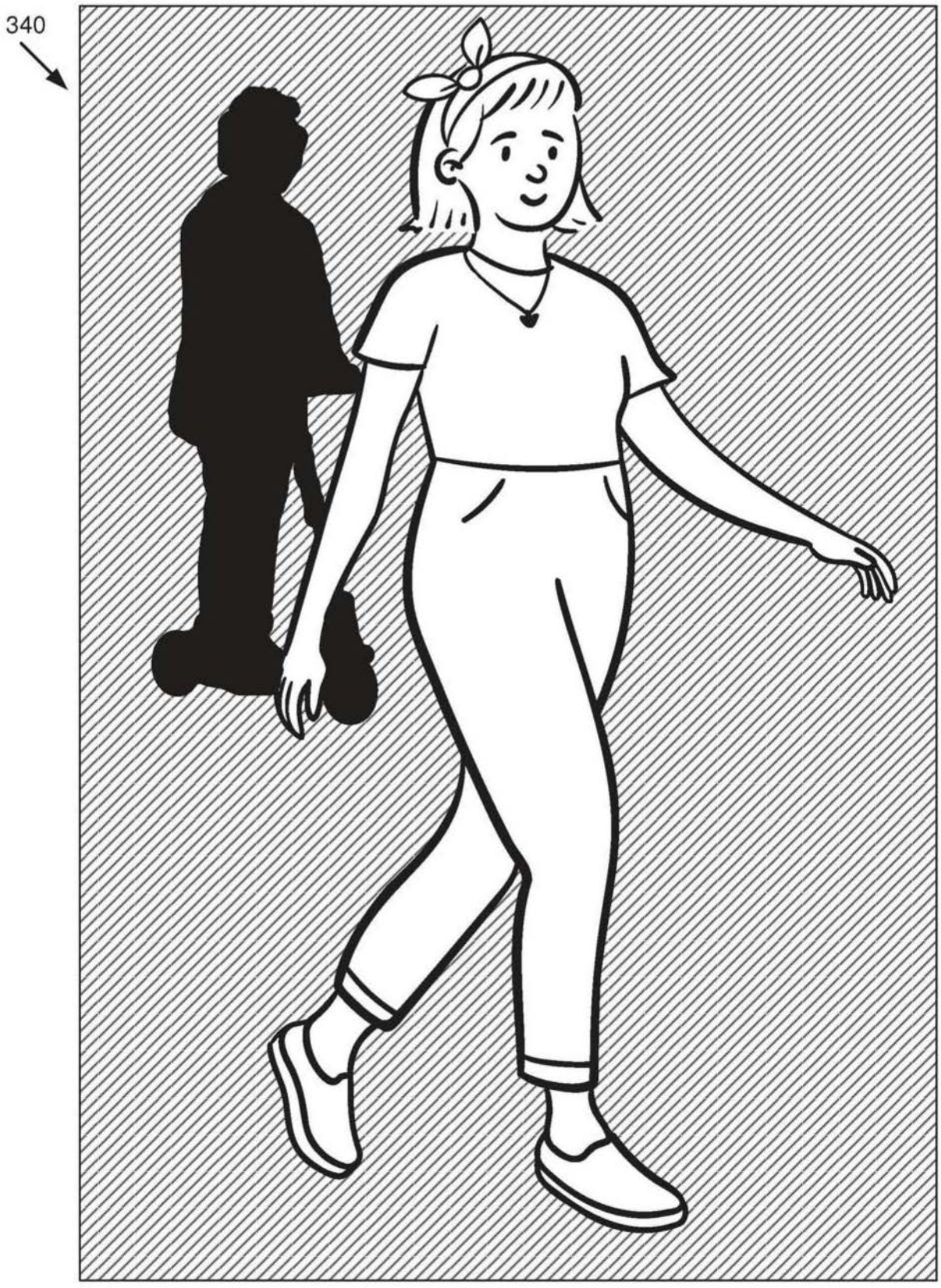


图3E

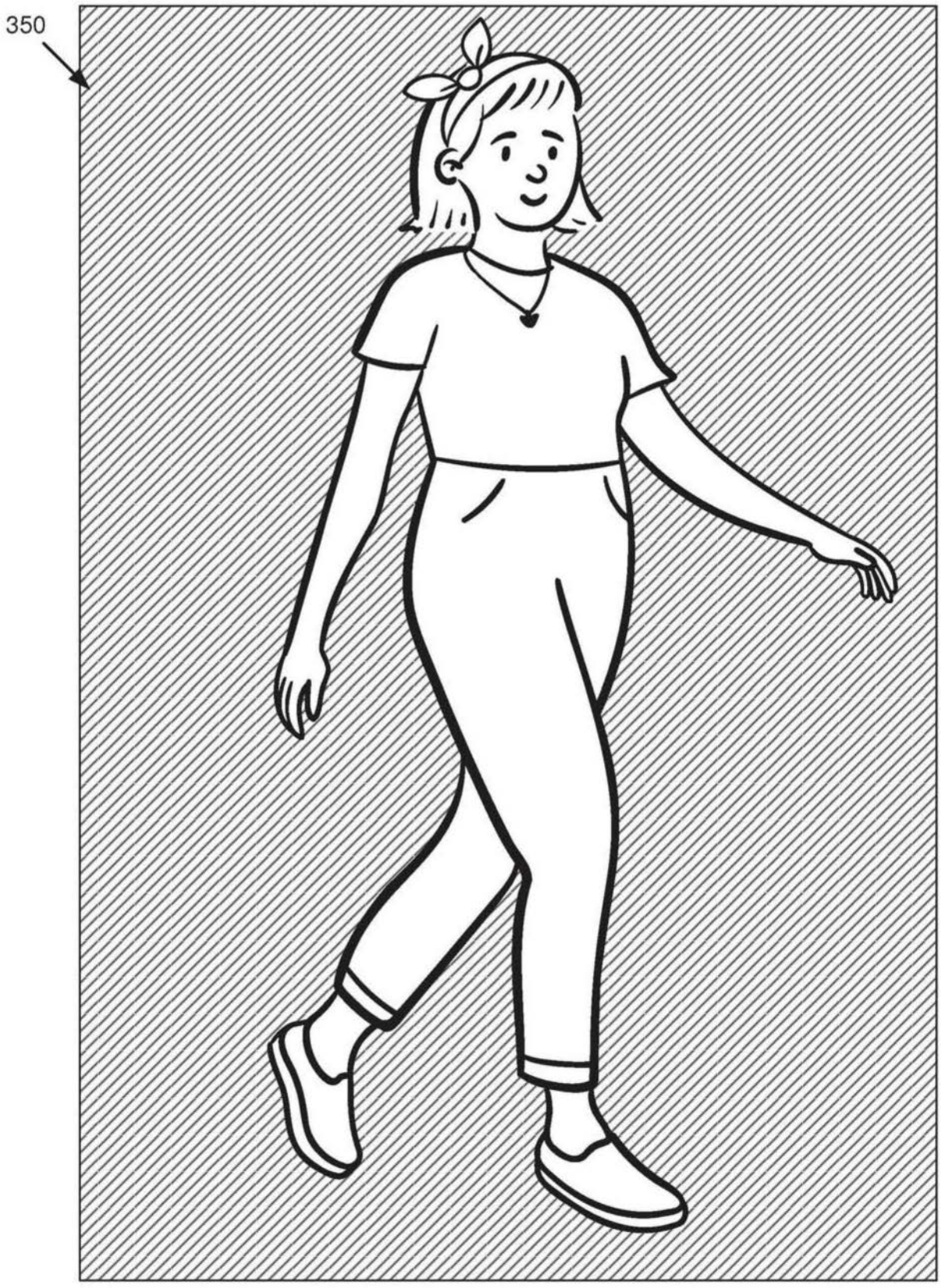


图3F

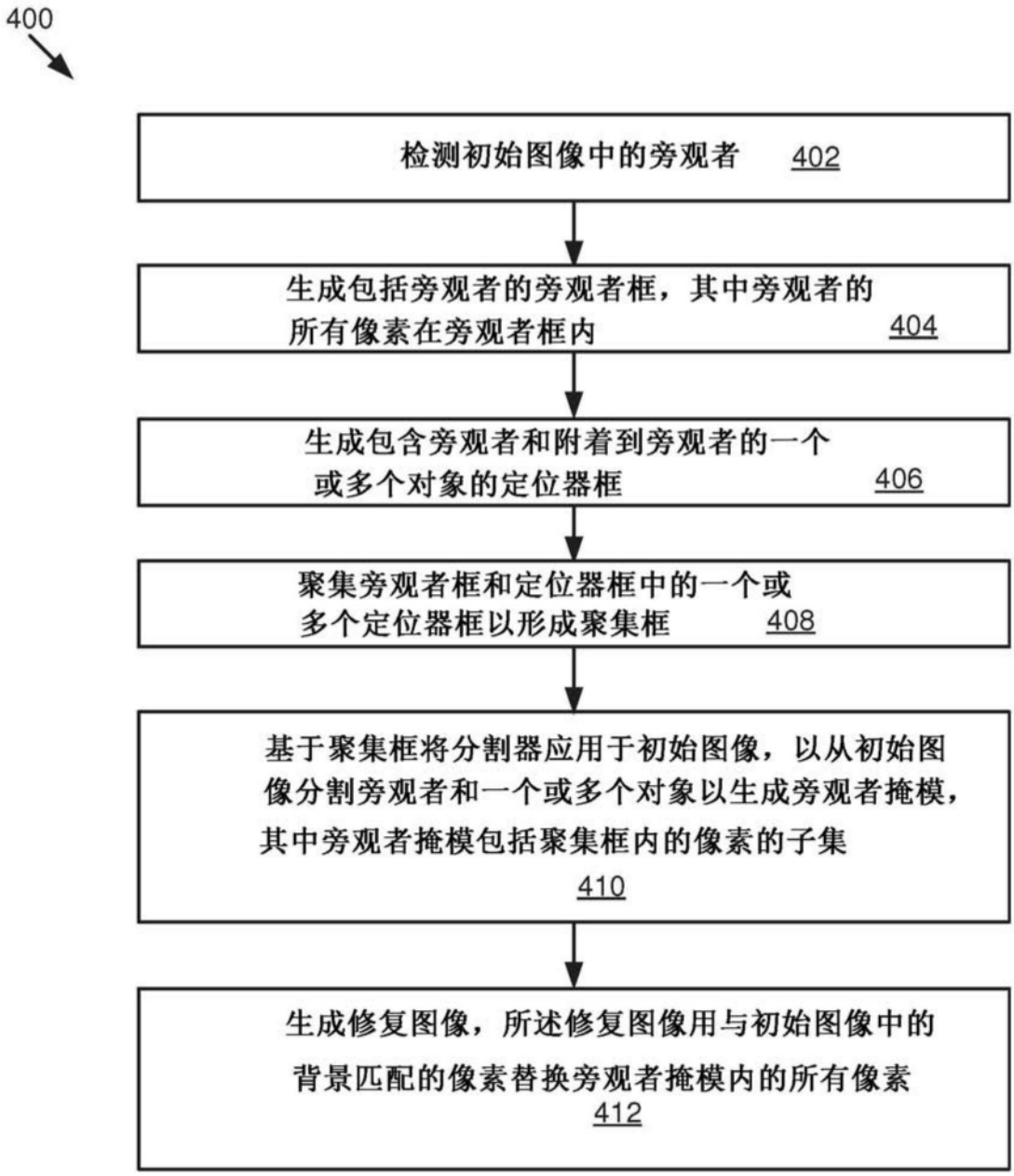

400
检测初始图像中的旁观者 402
生成包括旁观者的旁观者框，其中旁观者的所有像素在旁观者框内 404
生成包含旁观者和附着到旁观者的一个或多个对象的定位器框 406
聚集旁观者框和定位器框中的一个或多个定位器框以形成聚集框 408
基于聚集框将分割器应用于初始图像，以从初始图像分割旁观者和一个或多个对象以生成旁观者掩模，其中旁观者掩模包括聚集框内的像素的子集 410
生成修复图像，所述修复图像用与初始图像中的背景匹配的像素替换旁观者掩模内的所有像素 412

图4

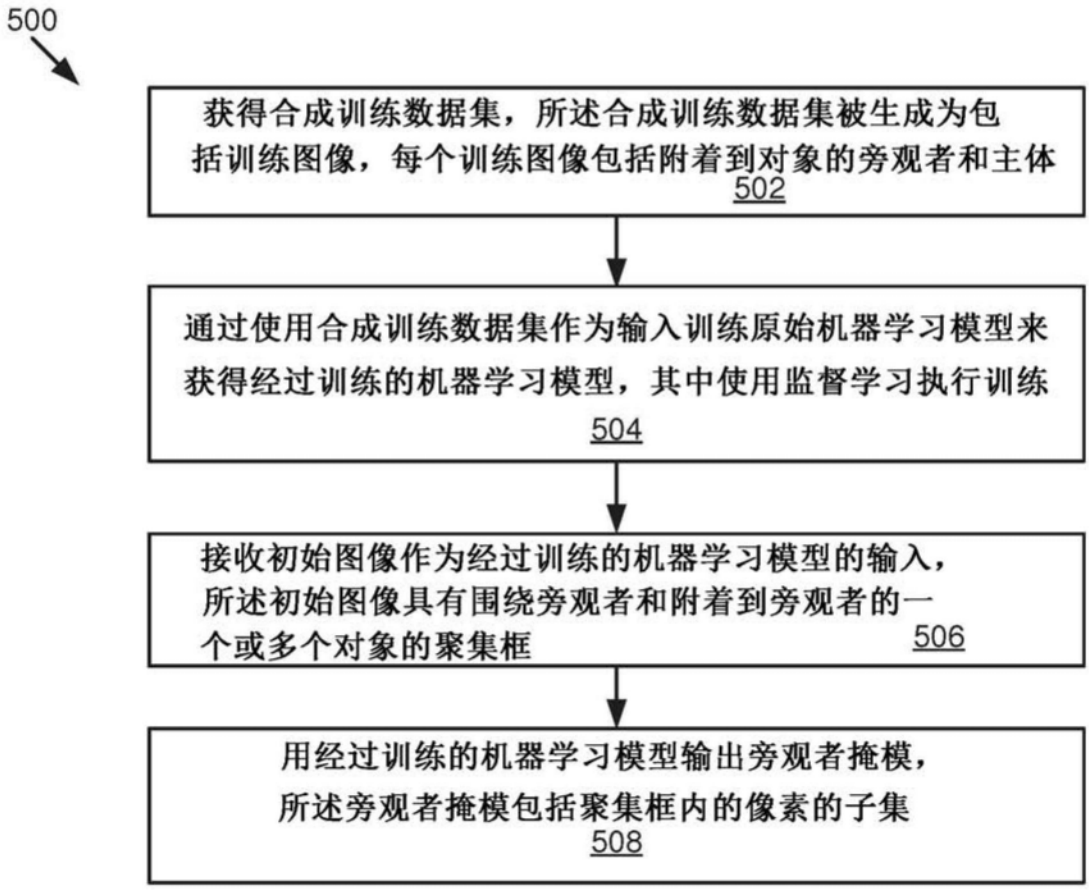

500
获得合成训练数据集，所述合成训练数据集被生成为包括训练图像，每个训练图像包括附着到对象的旁观者和主体
502
通过使用合成训练数据集作为输入训练原始机器学习模型来获得经过训练的机器学习模型，其中使用监督学习执行训练
504
接收初始图像作为经过训练的机器学习模型的输入，所述初始图像具有围绕旁观者和附着到旁观者的一个或多个对象的聚集框
506
用经过训练的机器学习模型输出旁观者掩模，所述旁观者掩模包括聚集框内的像素的子集
508

图5

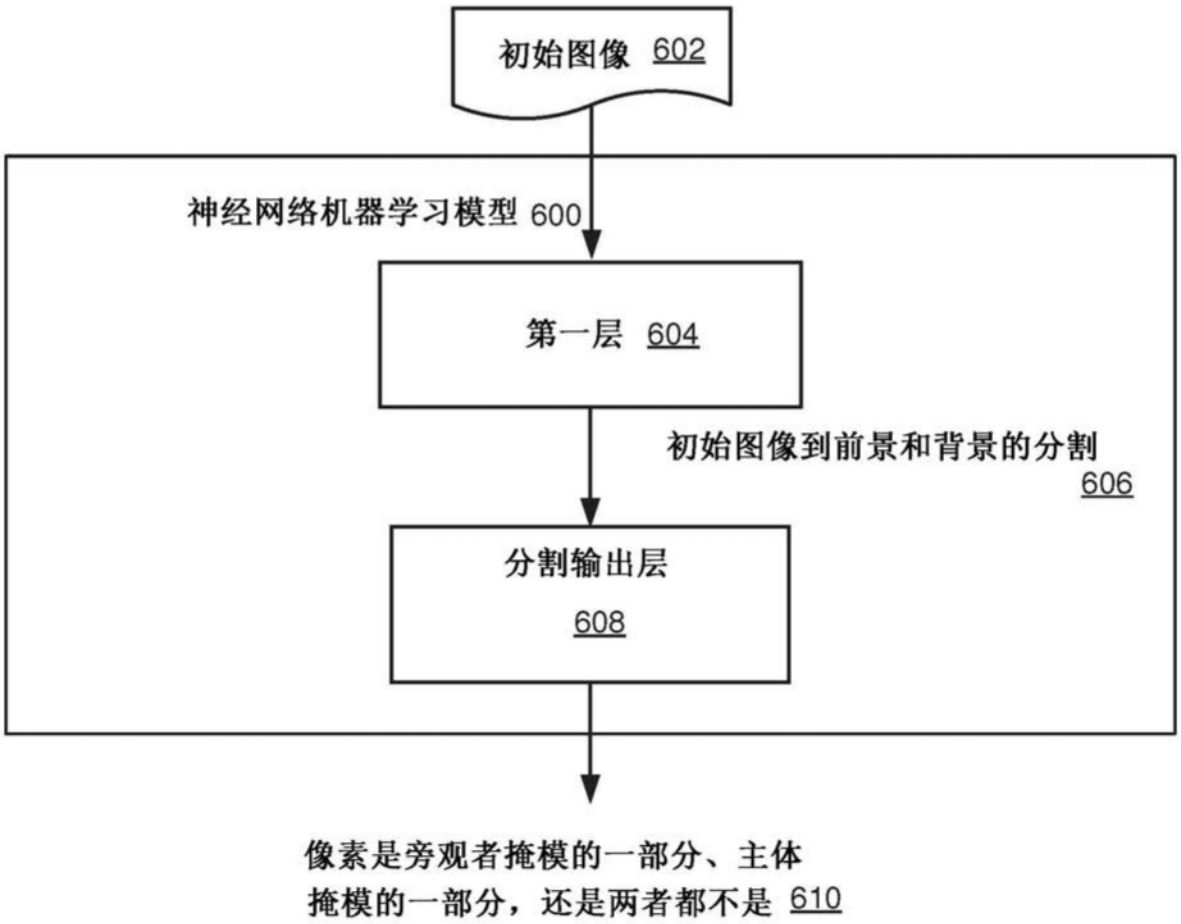

初始图像 602
神经网络机器学习模型 600
第一层 604
初始图像到前景和背景的分割
606
分割输出层
608
像素是旁观者掩模的一部分、主体
掩模的一部分，还是两者都不是 610

图6