



(12) 发明专利申请

(10) 申请公布号 CN 112614149 A

(43) 申请公布日 2021.04.06

(21) 申请号 202011459573.2

(22) 申请日 2020.12.11

(71) 申请人 北京影谱科技股份有限公司
地址 100000 北京市朝阳区朝阳公园南路3
号楼6层601内702室

(72) 发明人 孙腾

(74) 专利代理机构 北京万思博知识产权代理有
限公司 11694
代理人 高镇

(51) Int. Cl.

G06T 7/194 (2017.01)

G06T 7/11 (2017.01)

G06T 5/50 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

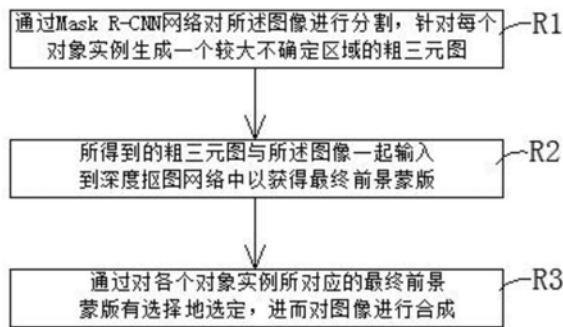
权利要求书1页 说明书5页 附图2页

(54) 发明名称

基于实例分割的语义合成方法

(57) 摘要

本申请公开了一种基于实例分割的语义合成方法,用于对图像进行合成,通过MaskR-CNN网络对所述图像进行分割,针对每个对象实例生成一个较大不确定区域的粗三元图,然后将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版,通过所述最终前景蒙版对图像进行合成。本申请能够实现自动前景对象提取方法,即使在复杂背景下也可以工作,对于复杂背景下的图像可对各个对象实例进行分割和图像抠图过程,允许对多个对象实例进行分类,并从背景中分割和提取,便于后续有选择地进行图像合成,有效提高处理效率,同时降低人力投入。



1. 一种基于实例分割的语义合成方法,用于对图像进行合成,其特征在于,通过Mask R-CNN网络对所述图像进行分割,针对每个对象实例生成一个较大不确定区域的粗三元图,然后将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版,通过所述最终前景蒙版对图像进行合成。

2. 根据权利要求1所述的基于实例分割的语义合成方法,其特征在于,通过Mask R-CNN网络对所述图像进行分割时,利用MASK R-CNN的中间输出为每个检测到的对象实例生成对象边界框和实例掩码,利用所得到的对象边界框和实例掩码对每个实例估计出一个较大不确定区域的粗三元图。

3. 根据权利要求1或2所述的基于实例分割的语义合成方法,其特征在于,将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版的过程,是在前次的蒙版输出和深度抠图网络输入之间插入了反馈连接,从而得到逐次质量提升的蒙版输出,并在多次反馈循环后得到所述最终前景蒙版。

4. 根据权利要求3所述的基于实例分割的语义合成方法,其特征在于,所述反馈循环的次数至少为3次。

5. 根据权利要求3所述的基于实例分割的语义合成方法,其特征在于,将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版的过程包括如下步骤:

步骤S1:

将粗三元图与原始的所述图像一起输入到深度抠图网络中,得到蒙版输出;

步骤S2:

将步骤S1中的蒙版输出的alpha值扩张来创建新的三元图,将新的三元图再连通所述图像反馈到深度抠图网络中再次细化得到新的蒙版输出,所得到的新的蒙版输出的alpha值再次扩张进而循环前述过程以得到进一步细化的蒙版输出,构成闭环的反馈循环;

步骤S3:

在完成至少三次的所述反馈循环后得到质量提高的所述最终前景蒙版。

6. 根据权利要求5所述的基于实例分割的语义合成方法,其特征在于,在所述步骤S2中,所采用的深度抠图网络是对三元图中的未知区域卷积生成前景蒙版块,再将所述前景蒙版块与三元图中的原始块求平均值,从而得到新的蒙版输出。

7. 根据权利要求5所述的基于实例分割的语义合成方法,其特征在于,所述深度抠图网络采用的是VGG-16架构。

8. 一种计算设备,包括存储器、处理器和存储在所述存储器内并能由所述处理器运行的计算机程序,其中,所述处理器执行所述计算机程序时实现如权利要求1-7中任一项所述的方法。

9. 一种计算机可读存储介质,优选为非易失性可读存储介质,其内存储有计算机程序,所述计算机程序在由处理器执行时实现如权利要求1-7中任一项所述的方法。

10. 一种计算机程序产品,包括计算机可读代码,当所述计算机可读代码由计算机设备执行时,导致所述计算机设备执行如权利要求1-7中任一项所述的方法。

基于实例分割的语义合成方法

技术领域

[0001] 本申请涉及图像处理技术领域,特别是涉及一种基于实例分割的语义合成方法。

背景技术

[0002] 图像合成是电影制作和图像编辑中的关键步骤,它将来自不同来源的视觉元素组合成一个图像。当背景为纯蓝色或绿色时,使用色度键控可以在工作室中轻松完成自动图像合成。然而,在具有复杂背景的自然场景中进行图像合成仍然是一项繁琐的工作,需要有经验的人员进行手工分割。从图像源中提取视觉元素的过程通常需要专业人员手工从确定的背景中精确提取要观察的前景对象,或者通过自动色度键控从彩色背景(例如绿屏)中识别前景对象,这无疑会给图像自动处理带来了困难,不但增加了人工成本,还影响了生产效率。

发明内容

[0003] 本申请的目的在于克服上述问题或者至少部分地解决或缓减解决上述问题。

[0004] 根据本申请的一个方面,提供了一种基于实例分割的语义合成方法,用于对图像进行合成,通过Mask R-CNN网络对所述图像进行分割,针对每个对象实例生成一个较大不确定区域的粗三元图,然后将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版,通过所述最终前景蒙版对图像进行合成。

[0005] 可选地,通过Mask R-CNN网络对所述图像进行分割时,利用MASK R-CNN的中间输出为每个检测到的对象实例生成对象边界框和实例掩码,利用所得到的对象边界框和实例掩码对每个实例估计出一个较大不确定区域的粗三元图。

[0006] 可选地,将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版的过程,是在前次的蒙版输出和深度抠图网络输入之间插入了反馈连接,从而得到逐次质量提升的蒙版输出,并在多次反馈循环后得到所述最终前景蒙版。

[0007] 可选地,所述反馈循环的次数至少为3次。

[0008] 可选地,将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版的过程包括如下步骤:

[0009] 步骤S1:

[0010] 将粗三元图与原始的所述图像一起输入到深度抠图网络中,得到蒙版输出;

[0011] 步骤S2:

[0012] 将步骤S1中的蒙版输出的alpha值扩张来创建新的三元图,将新的三元图再连通所述图像反馈到深度抠图网络中再次细化得到新的蒙版输出,所得到新的蒙版输出的alpha值再次扩张进而循环前述过程以得到进一步细化的蒙版输出,构成闭环的反馈循环;

[0013] 步骤S3:

[0014] 在完成至少三次的所述反馈循环后得到质量提高的所述最终前景蒙版。

[0015] 可选地,在所述步骤S2中,所采用的深度抠图网络是对三元图中的未知区域卷积

生成前景蒙版块,再将所述前景蒙版块与三元图中的原始块求平均值,从而得到新的蒙版输出。

[0016] 可选地,所述深度抠图网络采用的是VGG-16架构。

[0017] 特别地,本发明还提供了一种计算设备,包括存储器、处理器和存储在所述存储器内并能由所述处理器运行的计算机程序,其中,所述处理器执行所述计算机程序时实现如上所述的方法。

[0018] 本发明还提供了一种计算机可读存储介质,优选为非易失性可读存储介质,其内存储有计算机程序,所述计算机程序在由处理器执行时实现如上所述的方法。

[0019] 本发明还提供了一种计算机程序产品,包括计算机可读代码,当所述计算机可读代码由计算机设备执行时,导致所述计算机设备执行如上所述的方法。

[0020] 本申请的基于实例分割的语义合成方法,其能够实现自动前景对象提取方法,即使在复杂背景下也可以工作,对于复杂背景下的图像可对各个对象实例进行分割和图像抠图过程,允许对多个对象实例进行分类,并从背景中分割和提取,便于后续有选择地进行图像合成,有效提高处理效率,同时降低人力投入。

[0021] 根据下文结合附图对本申请的具体实施例的详细描述,本领域技术人员将会更加明了本申请的上述以及其他目的、优点和特征。

附图说明

[0022] 下文将参照附图以示例性而非限制性的方式详细描述本申请的一些具体实施例。附图中相同的附图标记标示了相同或类似的部件或部分。本领域技术人员应该理解,这些附图未必是按比例绘制的。附图中:

[0023] 图1是根据本申请一个实施例的基于实例分割的语义合成方法的方法流程图;

[0024] 图2是根据本申请另一个实施例的基于实例分割的语义合成方法的方法流程图;

[0025] 图3是根据本申请另一个实施例的计算设备结构图;

[0026] 图4是根据本申请另一个实施例的计算机可读存储介质结构图。

具体实施方式

[0027] 本实施例可以在复杂的背景下实现自动提取前景对象实例。本实施例结合了对象实例分割和图像匹配过程,允许从复杂背景中对多个前景对象实例进行分类、分割和提取。

[0028] 实例分割过程提供的语义标签提供了一种自动提取不同类型对象的方法。例如,可以从背景中提取场景中的所有人,而将其他类型的前景对象留在后面。或者,我们可以获得多个前景蒙版,每个实例对应获得一个前景蒙版,这些前景蒙版可以在以后的合成操作中根据需要使用。

[0029] 目标分割是计算机视觉中最重要、最复杂的任务之一。基于学习方法的分割算法最近取得了许多进展,这在很大程度上是由于大量公开可用的数据集提供了人为绘制的分割掩码作为真实值,比如说COCO数据集,COCO数据集是微软团队提供的一个可以用来进行图像识别的数据集。主要从复杂的日常场景中截取,COCO数据集有超过200000张图片,80种物体类别。所有的物体实例都用详细的分割mask进行了标注,共标注了超过500,000个物体实例。

[0030] 图像匹配是计算机视觉的另一个基本问题,其涉及为每个对象生成粗二值掩模,通过估计包含图像中每个像素的不透明度值的前景蒙版从静态图像或一组视频帧中提取感兴趣的对象。前景中的像素的alpha值 α 等于1,而背景中的像素的alpha值 α 等于0,但是由于像素的扩展大小以及视频中的运动模糊,前景和背景对象之间的边界上的像素有前景和背景对象的贡献,因此给出0到1之间的alpha值 α 。其中 i ,FG,BG分别表示像素位置、前景图像和背景图像:

$$[0031] \quad I_i = \alpha_i \times FG_i + (1 - \alpha_i) \times BG_i$$

[0032] 唯一已知的值是图像输入 I ,而变量FG, α ,BG是未知的并且需要估计。为了简化估计过程,大多数图像消光算法需要以用户标记输入的形式进行手动干预。

[0033] 对此,本实施例提出了一种基于实例分割的语义合成方法,用于对图像进行合成,如图1所示:

[0034] 步骤R1:

[0035] 通过Mask R-CNN网络对所述图像进行分割,针对每个对象实例生成一个较大不确定区域的粗三元图,

[0036] 步骤R2:

[0037] 将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版;

[0038] 步骤R3:

[0039] 通过对各个对象实例所对应的最终前景蒙版有选择地选定,进而对图像进行合成。

[0040] 在本实施例中,首先将一个输入图像导入MASK R-CNN网络,利用MASK R-CNN的中间输出为每个检测到的对象实例生成对象边界框和实例掩码,利用所得到的对象边界框和实例掩码对每个实例估计出一个较大不确定区域的粗三元图。利用这些中间结果,可以对每个被检测对象估计出一个具有较大不确定区域的粗三元图。此处选用Mask R-CNN网络算法作为整个流程的第一阶段生成实例掩码和边界框,Mask R-CNN网络的推理误差会影响后面步骤的准确性,因此,如使用稳定的Mask R-CNN网络,在边界框和掩模推理中实现更高的精度。在本实施例中该Mask R-CNN网络建立在更深层的ResNeXt-152-32x8d-FPN上,在ImageNet-5k上训练,而不是通常的ImageNet-1k,使用Detectron Pytorch预训练Mask R-CNN模型得到最终被使用的Mask R-CNN网络。三元图估计利用Mask R-CNN网络的中间输出为每个检测到的对象实例生成三元图。

[0041] 为了估计最终的前景蒙版,这个粗三元图和输入图像被输入到深度抠图网络中,但由于实例掩模的粗糙性导致了低质量的三元图,也就是说最初得到的确实为粗三元图。从生成的前景蒙版中估计一个新的三元图,然后再次经过抠图阶段,通常可以改进分割和matting结果。在本实施例中,将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版的过程,是在前次的蒙版输出和深度抠图网络输入之间插入了反馈连接,从而得到逐次质量提升的蒙版输出,并在多次反馈循环后得到所述最终前景蒙版。

[0042] 如上所述,在三元图与深度抠图网络所形成的蒙版输出互相改进的过程中就形成了一个反馈循环,本实施例假设掩模边界附近的区域是最需要进一步估计的区域,即图深度抠图网络应该关注的未知区域。对象掩模扩张的特定区域被定义为三元图的未知区域 α_i

$=0.5$,掩模内侧的区域指定为前景区域 $\alpha_i=1$,而该掩模外侧的区域被指定为背景 $\alpha_i=0$,扩张量由物体大小决定,选择扩张率为宽度和高度平均值的固定百分比。精确的三元图受到抠图网络的青睐,因为它强化了约束。当掩模边界接近物体真实边界时,小的扩张率就能够覆盖需要优化的区域。

[0043] 在另一实施例中,将所得到的粗三元图与所述图像一起输入到深度抠图网络中以获得最终前景蒙版的过程包括如下步骤:

[0044] 步骤S1:

[0045] 将粗三元图与原始的所述图像一起输入到深度抠图网络中,得到蒙版输出;

[0046] 步骤S2:

[0047] 将步骤S1中的蒙版输出的alpha值扩张来创建新的三元图,将新的三元图再连通所述图像反馈到深度抠图网络中再次细化得到新的蒙版输出,得到新的蒙版输出的alpha值再次扩张进而循环前述过程以得到进一步细化的蒙版输出,构成闭环的反馈循环;

[0048] 步骤S3:

[0049] 在完成至少三次的所述反馈循环后得到质量提高的所述最终前景蒙版。

[0050] 可选地,在所述步骤S2中,所采用的深度抠图网络是对三元图中的未知区域卷积生成前景蒙版块,再将所述前景蒙版块与三元图中的原始块求平均值,从而得到新的蒙版输出。

[0051] 具体说来,本实施例中对于深度抠图网络的选择,采用了Lin等人提出的基于学习的深度图像抠图方法,它是一个基于vgg-16的编解码器网络,再加一个全连接的优化阶段。深度抠图网络的训练数据的预处理是 320×320 图像块并将未知区域像素(即 $\alpha_i=0.5$)从训练数据中随机裁剪。下采样方法会导致细节损失,特别是当测试图像的大小与期望大小之比非常高时。为了避免这种情况,我们对输入测试图像实现了一种基于块的区域块预处理方法。这种方法包括将测试图像裁剪成 320×320 的区域块,单个图像的每个区域块都被单独地输入到深度抠图网络中,得到多个蒙版结果。然后,将这些结果混合在一起。使用这种技术的一个限制是,我们必须确保裁剪后的输入图像位于三元图中一个未知像素的中心。如果不是这样,网络的性能可能会很差。此外,当裁剪后的区域块的内容与网络训练阶段使用的训练数据差别非常大时,网络性能会受到影响。为了解决这些问题,首先将高分辨率的测试图像下采样到 640×640 。然后,随机裁剪区域块,让它们的中心沿着三元图的灰色区域移动,直到覆盖整个未知区域。对于每个区域块,都会生成一个前景蒙版区域块。所有这些前景蒙版区域块都被粘贴回它们的原始位置,并与现有区域块的重叠部分进行平均,进而得到最终前景蒙版。通常,四个反馈回路可以得到准确的结果。

[0052] 根据下文结合附图对本申请的具体实施例的详细描述,本领域技术人员将会更加明了本申请的上述以及其他目的、优点和特征。

[0053] 本申请实施例还提供了一种计算设备,参照图3,该计算设备包括存储器1120、处理器1110和存储在所述存储器1120内并能由所述处理器1110运行的计算机程序,该计算机程序存储于存储器1120中的用于程序代码的空间1130,该计算机程序在由处理器1110执行时实现用于执行任一项根据本发明的方法步骤1131。

[0054] 本申请实施例还提供了一种计算机可读存储介质。参照图4,该计算机可读存储介质包括用于程序代码的存储单元,该存储单元设置有用于执行根据本发明的方法步骤的程

序1131',该程序被处理器执行。

[0055] 本申请实施例还提供了一种包含指令的计算机程序产品。当该计算机程序产品在计算机上运行时,使得计算机执行根据本发明的方法步骤。

[0056] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时,可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机加载和执行所述计算机程序指令时,全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、获取其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一个计算机可读存储介质传输,例如,所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如同轴电缆、光纤、数字用户线(DSL))或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质,(例如,软盘、硬盘、磁带)、光介质(例如,DVD)、或者半导体介质(例如固态硬盘 Solid State Disk (SSD))等。

[0057] 专业人员应该还可以进一步意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0058] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分步骤是可以通程序来指令处理器完成,所述的程序可以存储于计算机可读存储介质中,所述存储介质是非短暂性(英文:non-transitory)介质,例如随机存取存储器,只读存储器,快闪存储器,硬盘,固态硬盘,磁带(英文:magnetic tape),软盘(英文:floppy disk),光盘(英文:optical disc)及其任意组合。

[0059] 以上所述,仅为本申请较佳的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应该以权利要求的保护范围为准。

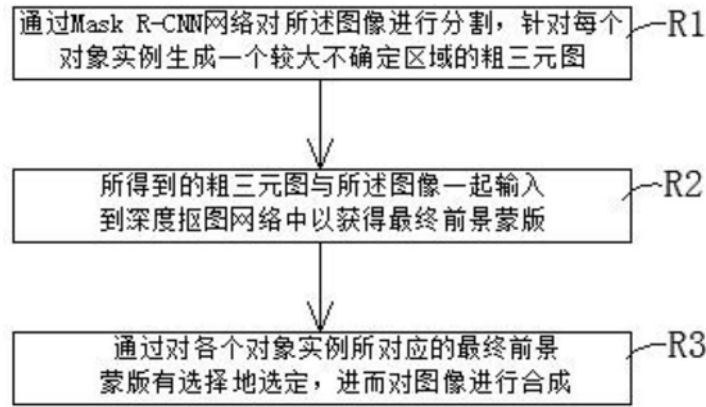


图1

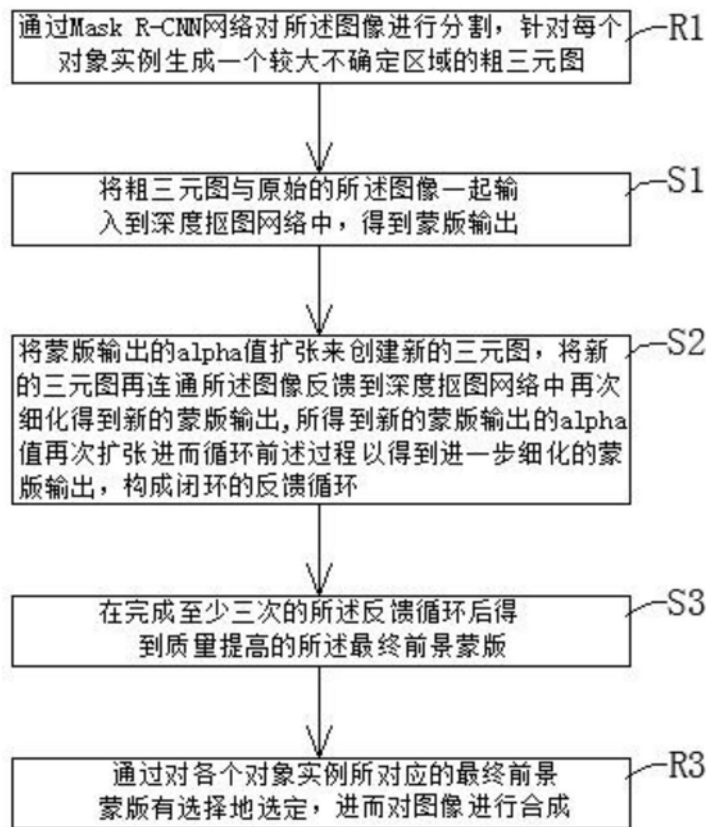


图2

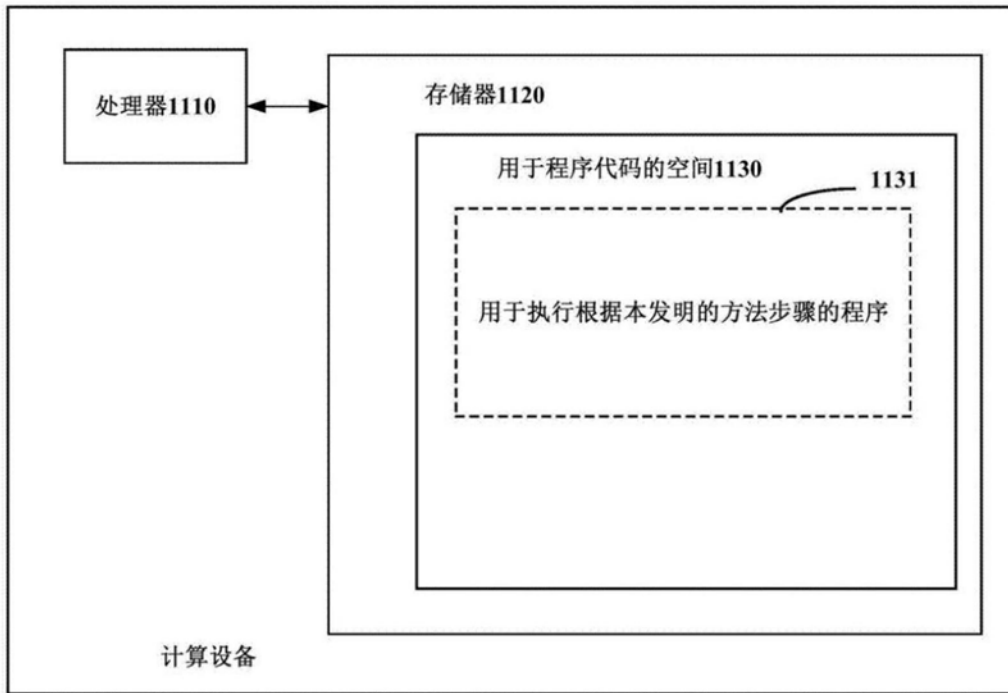


图3

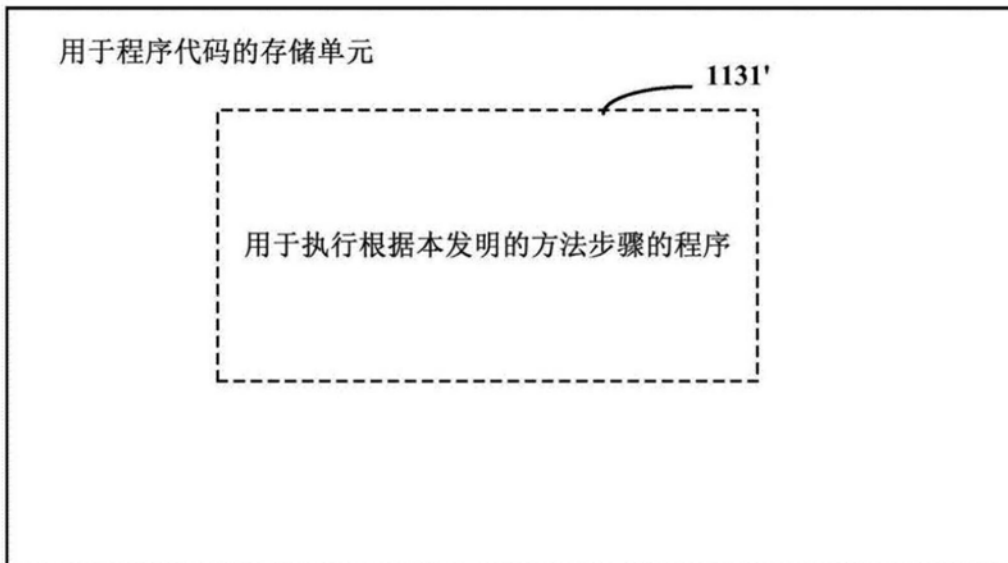


图4