(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119032384 A
(43)申请公布日 2024.11.26

(21)申请号 202280081511.3
(22)申请日 2022.10.07
(30)优先权数据
63/253,496 2021.10.07 US
(85)PCT国际申请进入国家阶段日
2024.06.07
(86)PCT国际申请的申请数据
PCT/US2022/046040 2022.10.07
(87)PCT国际申请的公布数据
WO2023/059876 EN 2023.04.13
(71)申请人 康耐视公司
地址 美国马萨诸塞州
(72)发明人 Z·H·刘 P·汉哈特 R·怀斯 S·A·巴克
(74)专利代理机构 广州嘉权专利商标事务所有限公司 44205
专利代理师 邓柏威
(51)Int.Cl.
G06V 20/62(2022.01)
G06V 10/82(2022.01)
G06T 7/00(2017.01)

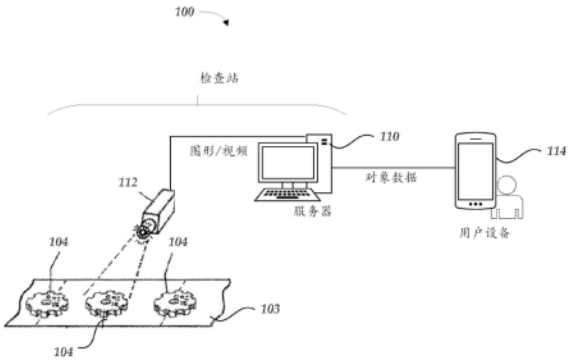
权利要求书4页 说明书25页 附图9页

(54)发明名称

用于检测对象的系统和方法

(57)摘要

本文描述的技术涉及用于检测图像中的对象的计算机化方法和设备。本文描述的技术还涉及使用预训练的机器学习模型和可以在现场训练过程中训练的一个或多个其他机器学习模型来检测一个或多个对象的计算机化方法和装置。预训练的机器学习模型可以是深度机器学习模型。



CN 119032384 A

1.一种计算机化方法,用于检测图像中的一个或多个对象,所述方法包括:

确定所述图像的特征图;

使用第一机器学习模型处理所述图像的所述特征图以生成所述图像的对象中心热图,其中所述对象中心热图包括多个样本,每个样本具有指示所述图像中的对应样本是对象中心的可能性的值;和

基于所述对象中心热图确定所述图像中一个或多个对象的位置。

2.根据权利要求1所述的方法,还包括:

使用第二机器学习模型和所述特征图来处理所述图像中的一个或多个对象的位置,以识别所述一个或多个对象中的对象。

3.根据权利要求2所述的方法,其中识别所述一个或多个对象中的对象包括:

使用所述图像的所述特征图的一部分生成对象特征向量,其中所述特征图的所述一部分是基于所述对象的位置周围的区域;

使用所述第二机器学习模型处理所述对象特征向量以生成类别向量,其中所述类别向量包括多个值,每个值对应于多个已知标签之一;和

使用所述类别向量将所述对象分类到所述多个已知标签中的一个标签。

4.根据权利要求3所述的方法,其中:

所述类别向量的每个值表示与所述多个已知标签中对应的一个标签相关联的被预测分数;和

对所述对象进行分类包括在所述类别向量的所述多个值中选择最大值,其中所选择的值对应于所述标签。

5.根据权利要求3和4中任一项所述的方法,其中所述多个已知标签包括多个文本字符标签。

6.根据权利要求3和4中任一项所述的方法,其中所述多个已知标签还包括背景标签。

7.根据权利要求2-6中任一项所述的方法,还包括使用相应的机器学习方法并使用相应的一组现场训练数据来训练所述第一机器学习模型和所述第二机器学习模型中的每一个。

8.根据权利要求2-6中任一项所述的方法,其中:

所述一个或多个对象包括包含在所述图像中的零件上的一个或多个印刷文本字符;和

所述方法还包括使用从所述一个或多个文本字符中已识别的至少一个文本字符来跟踪所述零件。

9.根据权利要求1-8中任一项所述的方法,其中确定所述一个或多个对象的位置包括:

平滑所述对象中心热图以生成被平滑的对象中心热图;和

选择所述一个或多个对象的位置,其中所述被平滑的对象中心热图中的每个相应位置处的值高于所述位置的邻近区域中的值。

10.根据权利要求9所述的方法,其中平滑所述对象中心热图包括应用具有与对象大小成比例的标准偏差的高斯滤波器。

11.根据权利要求9和10中任一项所述的方法,其中选择所述位置还包括过滤其在所述被平滑的对象中心热图中的值低于阈值的一个或多个位置。

12.根据权利要求1-11中任一项所述的方法,其中:

所述第一机器学习模型包括权重向量;

所述图像的所述特征图包括多个样本,每个样本与相应的特征向量相关联;和

所述对象中心热图中每个样本的值是所述特征图中相应样本的特征向量和权重向量的点积。

13.根据权利要求1-12中任一项所述的方法,其中确定所述图像的所述特征图包括:

使用预训练的神经网络模型处理所述图像以生成所述图像的所述特征图。

14.根据权利要求1-13中任一项所述的方法,还包括使用1D条形码扫描仪或2D条形码扫描仪捕获所述图像。

15.一种非暂时性计算机可读介质,包括指令,当由计算设备上的一个或多个处理器执行时,所述指令可操作来使所述一个或多个处理器执行权利要求1-14中任一项所述的方法。

16.一种系统,包括:

扫描仪,包括图像捕获装置,所述图像捕获装置配置成捕获检查站上的零件的图像;和

处理器,配置为执行编程指令以执行包括权利要求1-14中任一项的操作。

17.一种计算机化方法,用于识别图像中的一个或多个字符,所述方法包括:

使用预训练的机器学习模型处理所述图像以生成所述图像的特征图;

使用第一机器学习模型处理所述图像的所述特征图以生成所述图像的字符中心热图,其中所述字符中心热图包括多个样本,每个样本具有指示所述图像中的对应样本是字符的中心的可能性的值;和

使用第二机器学习模型处理所述图像的所述特征图和所述图像的所述字符中心热图以识别所述图像中的一个或多个字符。

18.根据权利要求17所述的方法,其中识别所述一个或多个字符中的至少一个字符包括:

使用所述图像的所述特征图和所述图像的所述字符中心热图生成字符特征向量;

使用所述第二机器学习模型处理所述字符特征向量以生成类别向量,其中所述类别向量包括多个值,每个值对应于多个已知标签之一;和

使用所述类别向量将所述字符分类到所述多个已知标签中的一个标签。

19.根据权利要求18所述的方法,其中:

所述类别向量的每个值表示与所述多个已知标签中对应的一个标签相关联的被预测分数;和

对所述字符进行分类包括在所述类别向量的多个值中选择最大值,其中所选择的值对应于所述标签。

20.根据权利要求18和19中任一项所述的方法,其中所述多个已知标签还包括背景标签。

21.根据权利要求18-20中任一项所述的方法,还包括基于所述字符中心热图来确定所述一个或多个字符在所述图像中的多个位置,其中生成所述至少一个字符的所述字符特征向量包括使用所述图像的特征图的一部分来生成所述字符特征向量,其中所述特征图的一部分是基于所述字符的所述位置周围的区域。

22.根据权利要求21所述的方法,其中确定所述一个或多个字符的位置包括:

平滑所述字符中心热图以生成被平滑的字符中心热图;和

选择所述一个或多个字符的位置,其中所述被平滑的字符中心热图中的每个相应位置处的值高于所述位置的邻近区域中的值。

23.根据权利要求22所述的方法,其中平滑所述字符中心热图包括应用具有与字符大小成比例的标准偏差的高斯滤波器。

24.根据权利要求22和23中任一项所述的方法,其中选择所述位置还包括过滤其在所述被平滑的字符中心热图中的值低于阈值的一个或多个位置。

25.根据权利要求17-24中任一项所述的方法,其中:

所述第一机器学习模型包括权重向量;

所述图像的所述特征图包括多个样本,每个样本与相应的特征向量相关联;和

所述字符中心热图中每个样本的值是所述特征图中相应样本的特征向量和权重向量的点积。

26.根据权利要求17-25中任一项所述的方法,还包括通过以下方式训练所述第二机器学习模型:

获得多个训练图像和多个训练标签,所述多个训练标签中的每一个与所述多个训练图像中的相应一个中的字符相关联;

分别使用所述多个训练图像中的一个和所述预训练机器学习模型来确定多个训练特征图;和

使用所述多个训练特征图和所述多个训练标签来确定所述第二机器学习模型的权重。

27.根据权利要求26所述的方法,其中确定所述第二机器学习模型的所述权重包括:

对于所述多个训练标签中的每个训练标签:

使用与所述训练标签相关联的相应训练图像的所述特征图的一部分来确定相应的字符特征向量;和

确定目标向量;和

使用所述多个训练标签的字符特征向量和目标向量来确定所述第二机器学习模型的所述权重。

28.根据权利要求27所述的方法,其中确定所述第二机器学习模型的权重包括对所述字符特征向量和所述目标向量使用机器学习方法。

29.根据权利要求27和28中任一项所述的方法,还包括:

对于所述多个训练标签中的每个训练标签:

获得所述训练标签在相应训练图像中的位置;和

基于所述训练标签的所述位置,使用对应的训练图像的特征图的所述部分来确定对应的字符特征向量。

30.根据权利要求29所述的方法,其中所述特征图的所述部分由以相应训练图像中的训练标签的位置为中心的边界框表示。

31.根据权利要求29和30中任一项所述的方法,其中,确定对应的字符特征向量包括连接多个子特征向量,每个子特征向量是基于所述特征图的所述部分中的相应位置而形成的。

32.根据权利要求17-31中任一项所述的方法,还包括通过以下方式训练所述第二机器

学习模型:

确定多个训练特征图的多个基本特征中心热图;和

使用所述多个训练特征图和所述多个真实字符中心热图来确定所述第一机器学习模型的权重。

33.根据权利要求32所述的方法,确定所述第一机器学习模型的权重包括使用机器学习方法确定所述第一机器学习模型的权重。

34.根据权利要求32和33中任一项所述的方法,其中确定所述多个真实字符中心热图包括多个样本,每个样本具有基于到所述训练图像中的训练标签的最近真实位置的距离的值。

35.根据权利要求17至34中任一项所述的方法,还包括:

通过图形用户界面接收所述多个训练图像中的字符与所述多个训练图像中的一个或多个字符中的每一个的字符标签之间的一对一映射;和

基于所接收的所述多个训练图像中的一个或多个字符的一对一映射来生成所述多个训练标签。

36.根据权利要求35所述的方法,还包括:

通过所述图形用户界面,接收对一个或多个已识别的字符的校正;

基于所述校正更新所述多个训练标签;和

还基于已更新的多个训练标签来重新训练所述第一机器学习模型和所述第二机器学习模型。

37.根据权利要求17-36中任一项所述的方法,其中:

所述一个或多个字符包括包含在所述图像中的零件上的一个或多个印刷文本字符;和

所述方法还包括使用从所述一个或多个文本字符中已识别的至少一个文本字符来跟踪所述零件。

38.一种非暂时性计算机可读介质,包括指令,当由计算设备上的一个或多个处理器执行时,所述指令可操作来使所述一个或多个处理器执行权利要求17-37中任一项所述的方法。

39.一种系统,包括:

扫描仪,包括图像捕获装置,所述图像捕获装置配置成捕获检查站上的零件的图像;和

处理器,配置为执行编程指令以执行包括权利要求17-37中任一项的操作。

用于检测对象的系统和方法

[0001] 相关申请

[0002] 本申请根据35U.S.C.119(e)要求2021年10月7日提交的题为“用于检测对象的系统和方法”的美国临时申请第63/253,496号的优先权,其全部内容通过引用结合于本文。

技术领域

[0003] 本技术涉及机器视觉系统和方法,尤其涉及用于检测对象的系统和方法。

背景技术

[0004] 可能需要检测图像中的对象,例如字符。可以使用各种技术来检测图像中的对象。例如,光学字符识别(OCR)技术经常用于许多机器视觉系统中,以便检测与各种制造过程相关的文本,例如印刷和/或粘贴在机器零件上的文本。然而,为给定的OCR应用程序设置参数可能很困难,尤其是对于新用户。例如,对于字符识别,用户可以选择OCR字符串周围的区域,以向机器视觉系统指示用于进行OCR处理的字符在哪里。如果系统不能正确识别字符,用户很难手动解决问题。此外,与这些问题可以提前解决(例如,通过系统集成商)的一些应用程序不同,作为制造过程一部分的参数调整通常需要技术人员或工程师在实际生产现场解决这些问题(例如,训练或修改运行时间参数以提供更好的OCR)。例如,当制造新零件或使用新的印刷或标签时,可能需要这种训练或调整。

发明内容

[0005] 本公开涉及用于检测图像中诸如字符的对象的技术。所描述的技术的一些方面提供了用于识别图像中的一个或多个对象的计算机化方法、非暂时性计算机可读介质和/或系统。该方法包括确定图像的特征图。该方法可以使用预训练的机器学习模型来确定特征图。预训练的机器学习模型可以是深度机器学习模型。该方法还包括使用第一机器学习模型处理图像的特征图,以生成该图像的对象中心热图,其中该对象中心热图包括多个样本,每个样本具有一个值,该值指示该图像中的对应样本是对象中心的可能性。该方法还包括基于该对象中心热图确定图像中一个或多个对象的位置。在一些示例中,该方法还可以包括使用第二机器学习模型和特征图来处理图像中的一个或多个对象的位置,以识别该一个或多个对象中的至少一个对象。该第一和第二机器学习模型可以在现场训练过程中进行训练。

[0006] 在一些实施例中,识别该一个或多个对象中的至少一个对象包括:使用该图像的该特征图的一部分生成对象特征向量,其中该特征图的一部分是基于该对象位置周围的区域;使用该第二机器学习模型来处理该对象特征向量以生成类别向量,其中该类别向量包括多个值,每个值对应于多个已知标签中的一个;以及使用该类别向量将该对象分类到该多个已知标签中的一个标签。

[0007] 在一些实施例中,类别向量的每个值指示与多个已知标签中对应的一个标签相关联的预测分数,并且对对象进行分类包括在该类别向量的多个值中选择最大值,其中所选

择的值对应于该标签。

[0008] 在一些实施例中,该多个已知标签包括多个文本字符标签。

[0009] 在一些实施例中,该多个已知标签还包括背景标签。

[0010] 在一些实施例中,该方法还包括使用相应的机器学习方法并使用相应的一组现场训练数据来训练该第一机器学习模型和该第二机器学习模型中的每一个。

[0011] 在一些实施例中,该一个或多个对象包括包含在该图像中的零件上的一个或多个印刷文本字符;并且该方法还包括使用从该一个或多个文本字符中已识别的至少一个文本字符来跟踪该零件。

[0012] 在一些实施例中,确定该一个或多个对象的位置包括:平滑该对象中心热图以生成被平滑的对象中心热图;以及选择该一个或多个对象的位置,其中该被平滑的对象中心热图中的每个相应位置处的值高于该位置的邻近区域中的值。

[0013] 在一些实施例中,平滑对象中心热图包括应用具有与对象大小成比例的标准偏差的高斯滤波器。

[0014] 在一些实施例中,选择位置还包括过滤其在该被平滑的对象中心热图中的值低于阈值的一个或多个位置。

[0015] 在一些实施例中,该第一机器学习模型包括权重向量;该图像的该特征图包括多个样本,每个样本与相应的特征向量相关联;并且该对象中心热图中每个样本的值是该特征图中相应样本的特征向量和权重向量的点积。

[0016] 在一些实施例中,确定该图像的该特征图包括:使用预训练的神经网络模型处理该图像,以生成该图像的该特征图。

[0017] 在一些实施例中,该方法还包括使用1D条形码扫描仪或2D条形码扫描仪捕获该图像。

[0018] 在一些实施例中,提供了一种非暂时性计算机可读介质,其包括指令,当由计算设备上的一个或多个处理器执行时,所述指令可操作来使得该一个或多个处理器执行上述任何方法。

[0019] 在一些实施例中,提供了一种系统,该系统包括:扫描仪,该扫描仪包括配置成捕获检查站上的零件的图像的图像捕获设备;以及处理器,配置成执行编程指令以执行包括上述任何方法的操作。

[0020] 所描述的技术的一些方面提供了用于识别图像中的一个或多个字符的计算机化方法、非暂时性计算机可读介质和/或系统。在一些实施例中,一种方法包括使用预训练的机器学习模型来处理图像,以生成图像的特征图;使用第一机器学习模型处理该图像的该特征图,以生成该图像的字符中心热图,其中该字符中心热图包括多个样本,每个样本具有指示该图像中的对应样本是字符中心的可能性的值;以及使用第二机器学习模型处理该图像的该特征图和该图像的该字符中心热图,以识别该图像中的一个或多个字符。

[0021] 在一些实施例中,识别一个或多个字符中的至少一个字符包括:使用该图像的该特征图和该图像的该字符中心热图生成字符特征向量;使用该第二机器学习模型处理该字符特征向量,以生成类别向量,其中该类别向量包括多个值,每个值对应于多个已知标签中的一个;以及使用该类别向量将字符分类到该多个已知标签中的一个标签。

[0022] 在一些实施例中,该类别向量的每个值指示与该多个已知标签中对应的一个标签

相关联的预测分数,并且对该字符进行分类包括在该类别向量的多个值中选择最大值,其中所选择的值对应于该标签。

[0023] 在一些实施例中,该多个已知标签还包括背景标签。

[0024] 在一些实施例中,该方法还包括基于字符中心热图来确定一个或多个字符在图像中的位置,其中为至少一个字符生成字符特征向量包括使用图像的特征图的一部分来生成字符特征向量,其中特征图的一部分是基于字符位置周围的区域。

[0025] 在一些实施例中,确定该一个或多个字符的位置包括:平滑该字符中心热图以生成被平滑的字符中心热图;以及选择该一个或多个字符的位置,其中该被平滑的字符中心热图中的每个相应位置处的值高于该位置的邻近区域中的值。

[0026] 在一些实施例中,平滑该字符中心热图包括应用具有与字符大小成比例的标准偏差的高斯滤波器。

[0027] 在一些实施例中,选择位置还包括过滤其在该被平滑的字符中心热图中的值低于阈值的一个或多个位置。

[0028] 在一些实施例中,第一机器学习模型包括权重向量;图像的特征图包括多个样本,每个样本与相应的特征向量相关联;并且字符中心热图中每个样本的值是特征图中相应样本的特征向量和权重向量的点积。

[0029] 在一些实施例中,该方法包括通过以下步骤训练该第二机器学习模型:获得多个训练图像和多个训练标签,该多个训练标签中的每一个与该多个训练图像中的相应一个中的字符相关联;分别使用该多个训练图像之一和该预训练机器学习模型来确定多个训练特征图;以及使用该多个训练特征图和该多个训练标签来确定该第二机器学习模型的权重。

[0030] 在一些实施例中,确定该第二机器学习模型的权重包括:(1)对于该多个训练标签中的每个训练标签:使用与该训练标签相关联的相应训练图像的该特征图的一部分来确定相应的字符特征向量;并确定目标向量;以及(2)使用该多个训练标签的字符特征向量和目标向量来确定该第二机器学习模型的权重。

[0031] 在一些实施例中,确定该第二机器学习模型的权重包括对该字符特征向量和该目标向量使用机器学习方法。

[0032] 在一些实施例中,该方法还包括:对于该多个训练标签中的每个训练标签:(1)获得该训练标签在相应训练图像中的位置;以及(2)基于训该练标签的位置,使用相应训练图像的该特征图的该部分来确定相应的字符特征向量。

[0033] 在一些实施例中,该特征图的该部分由以相应训练图像中训练标签的位置为中心的边界框来表示。

[0034] 在一些实施例中,确定对应的字符特征向量包括连接多个子特征向量,每个子特征向量基于该特征图的该部分中的相应位置形成。

[0035] 在一些实施例中,该方法还包括通过以下步骤训练该第一机器学习模型:确定多个训练特征图的多个真实字符中心热图;以及使用该多个训练特征图和该多个真实字符中心热图来确定该第一机器学习模型的权重。

[0036] 在一些实施例中,确定该第一机器学习模型的权重包括使用机器学习方法来确定该第一机器学习模型的权重。

[0037] 在一些实施例中,确定该多个真实字符中心热图包括多个样本,每个样本具有一

值,该值基于到该训练图像中训练标签的最近真实位置的距离。

[0038] 在一些实施例中,该方法还包括:通过图形用户界面,接收该多个训练图像中的字符和该多个训练图像中的一个或多个字符中的每一个的字符标签之间的一对一映射;以及基于所接收的该多个训练图像中的该一个或多个字符的一对一映射来生成该多个训练标签。

[0039] 在一些实施例中,该方法还包括:通过图形用户界面,接收对一个或多个已识别字符的校正;基于该校正更新该多个训练标签;以及进一步基于已更新的多个训练标签来重新训练该第一机器学习模型和该第二机器学习模型。

[0040] 在一些实施例中,该一个或多个字符包括包含在该图像中的零件上的一个或多个印刷文本字符;并且该方法还包括使用从该一个或多个文本字符中已识别的至少一个文本字符来跟踪该零件。

[0041] 在一些实施例中,提供了包括指令的非暂时性计算机可读介质,当由计算设备上的一个或多个处理器执行时,所述指令可操作来使得该一个或多个处理器执行上述任何方法。

[0042] 在一些实施例中,提供了一种系统,该系统包括:扫描仪,该扫描仪包括配置成捕获检查站上的零件的图像的图像捕获设备;以及处理器,配置为执行编程指令以执行上述任何方法。

[0043] 应该理解的是,前述概念和下面更详细描述的附加概念的所有组合(假设这些概念不是相互矛盾的)被认为是本文公开的发明主题的一部分。具体而言,出现在本公开末尾的所要求保护的主题的所有组合都被视为本文公开的发明主题的一部分。

附图说明

[0044] 结合附图参考本文的描述,本公开的附加实施例及其特征和优点将变得更加明显。附图中的组件不一定按比例绘制。此外,在附图中,贯穿不同的视图,相同的附图标记表示相应的部分。

[0045] 图1是根据本文描述的技术的一些实施例的系统的示意图,该系统用于检测对象。

[0046] 图2是根据本文描述的技术的一些实施例的系统的另一示意图,该系统用于检测对象。

[0047] 图3是根据本文描述的技术的一些实施例的说明性方法的流程图,该方法用于检测对象。

[0048] 图4是根据本文描述的技术的一些实施例的说明性方法的流程图,该方法用于检测字符。

[0049] 图5A是根据本文描述的技术的一些实施例的说明性方法的流程图,该方法用于训练将在检测字符中使用的机器学习模型。

[0050] 图5B是根据本文描述的技术的一些实施例的说明性方法的流程图,该方法用于训练将在检测字符中使用的另一机器学习模型。

[0051] 图6是根据本文描述的技术的一些实施例的示例零件,该零件包括印刷在其上的可被检测的字符。

[0052] 图7示出了根据本文描述的技术的一些实施例的示例图形用户界面,该图形用户

界面用于在训练过程中纠正标签。

[0053] 图8示出了根据本文描述的技术的一些实施例的计算机系统的说明性实施方式，该计算机系统可用于执行本文公开的技术和实施例的任何方面。

具体实施方式

[0054] 为了促进对本公开的原理的理解，现在将参考附图中示出的实施例，并且将使用特定语言来描述这些实施例。然而，应当理解，这并不意味着对本发明范围的限制。

[0055] 对于各种机器视觉应用程序，可能希望检测图像中的对象，例如字符。例如，作为制造或检查过程的一部分，可能希望检测印刷和/或粘贴在机器零件上的文本。各种技术可用于检测机器视觉应用程序中的对象。例如，制造过程中的OCR应用可以包括识别具有描述零件号的一个或多个文本字符的打印标签。在该应用中，图像包括具有打印标签的部分，该部分可以被捕获。OCR技术可用于确定打印标签中一个或多个字符的位置并识别该字符。发明人已经认识到并意识到，将传统的对象检测技术应用于上述制造过程特别具有挑战性，并且可以对这种传统技术进行改进。例如，在传统的对象检测系统中，由于捕获图像的各种条件，在制造过程中被捕获的图像可能是低质量的(例如，与扫描的文档相比)。这些各种条件的示例可以包括例如场景中要捕获的零件的运动、相机和零件之间的距离、场景中零件的未知或不可控姿态和/或场景中部件的照明条件。例如，这些条件可能会影响图像的对比度。例如，对于字符，相机和零件之间的距离可能会影响图像中的字符大小。作为另一示例，零件上打印标签位置的变化可能影响被捕获的图像中字符的位置。制造过程中的这些不同条件和其他条件给应用这些传统技术带来了困难。

[0056] 一些对象检测系统被设计成通过训练某些系统参数来适配特定的应用或运行该应用的环境。从目标应用领域收集用于训练系统参数的训练数据。例如，训练数据可以在模拟部署系统的条件的实验室环境中收集或生成。然而，这种方法有几个缺点。首先，对象检测系统进行训练和开发的环境(例如，在实验室环境中)可能与对象检测系统在现场部署的环境非常不同。其次，为特定应用程序或环境收集训练数据可能成本高昂。

[0057] 一些对象检测系统，例如OCR系统，被设计成适用于某些应用程序，以通过允许用户调整某些参数来提供改进的对象检测性能。然而，这些系统具有缺点，因为调整参数通常需要系统集成商、技术人员甚至具有关于对象检测系统如何工作的知识的人。例如，在一个典型的系统中，可能有20-30个参数需要调整。调整这些参数可能超出了大多数用户的技能水平。

[0058] 一些对象检测系统使用机器学习技术。例如，深度机器学习技术已经用于OCR系统中，以从图像中检测字符。然而，发明人已经认识到并意识到，使用机器学习技术的用于对象识别的传统系统通常在调整部署期间使用的机器学习模型方面只能提供很少或没有灵活性。例如，机器学习(例如，深度机器学习)技术继承了前文讨论的相同缺点，因为机器学习技术通常需要庞大的训练数据集，例如数百、数千或数万个训练图像和相关联的真实数据(例如，提供关于图像中包含的对象的信息)。收集这样庞大的训练数据集可能是困难的和/或昂贵的。此外，当使用深度机器学习技术时，考虑到深度机器学习模型中的大量图像和大量权重，训练过程甚至可能极其耗时。例如，在卷积神经网络(CNN)中，可能有数万个权重。因此，以允许在该领域中调整机器学习模型的方式部署这些基于机器学习的技术通常

是不可行的。一旦进入现场,可能没有足够的训练数据和/或计算资源来在合理的时间段内训练和/或重新训练这些系统(例如,训练或调整过程足够快速从而不会导致糟糕的用户体验)。

[0059] 此外,发明人已经认识到并意识到传统的基于机器学习的系统通常使用黑盒方法来端到端地处理图像。例如,传统系统可以提供接收图像作为输入并生成对象检测结果作为输出,例如OCR结果,但是不提供对机器学习模型和/或用于生成对象检测结果的处理流水线的其他部分的任何可见性。附加地或可选地,当场景中存在多个对象时,这种系统不提供对单个对象的识别。例如,OCR系统不提供识别字符串的单个字符或生成处理流水线的中间结果或数据,例如提供指示正在进行OCR的图像中的字符位置的数据(例如,该数据进而可用于配置该OCR过程)。相反,这些系统通常仅提供对文本块的识别,例如图像中的一行或多行文本。

[0060] 因此,发明人已经开发了本文所述的技术,以提供可以在现场容易且快速地训练的对象检测系统。在一些实施例中,可以提供包括一个或多个机器学习模型的系统,该机器学习模型可以在该领域中进行训练。与在黑盒方法中使用复杂的机器学习模型相反,使用多个机器学习模型可以允许系统使用更小和更简单的机器学习模型,每个模型专注于提取图像的某些特征。

[0061] 对象检测可以包括用于检测图像、视频和/或任何其他合适的媒体资产中的一个或多个对象的操作。对象检测的示例可以包括确定一个或多个对象和/或识别图像、视频和/或任何其他合适的媒体资产中的对象。例如,在图像可以包括在装配线上行进的一个或多个机器零件的制造过程中,对象检测可以包括定位图像中的一个或多个零件。对象检测还可以包括识别一个或多个零件以及对应于每个已识别零件的相应已知类别。

[0062] 可以理解,图像中的一个或多个对象可以包括各种类型的项目。要从图像中检测的对象的其他示例可以包括文本(例如印刷标签中的字符)、图形、符号、图标、条形码、机器、人脸、动物、装置、标志、地标或任何其他合适的文本/图形表示或其组合。图像中包含的对象可以包括任何颜色或多种颜色的组合。文本字符可以是英语或任何其他合适的语言。因此,本文描述的技术可以应用于检测各种类型的项目,包括图像中的字符。字符可以包括任何符号,无论是文本符号还是图形符号。字符的示例可以包括ASCII集中的任何字符、任何合适语言的字母、或非文本符号,例如数学符号或其他符号。

[0063] 因此,所描述的技术包括处理图像以检测图像中的一个或多个对象的系统、计算机化方法和非暂时性指令。在一些实施例中,该系统可以获得包含一个或多个对象的图像,并使用预训练的机器学习模型来处理该图像以确定该图像的特征图。在一些示例中,该预训练的机器学习模型可以是深度机器学习模型,例如具有一个或多个隐藏层的卷积神经网络(CNN)或任何其他合适的模型。该预训练的机器学习模型可以配置成对该图像的一些特征进行编码,用于后续处理以输出特征图。该特征图可以包括多个样本,每个样本包括一个具有多个值的向量(例如,特征向量),这些值在每个相应位置对关于该图像的语义信息进行编码。该特征图的大小可能与该图像的大小不同。例如,该特征图可能比该图像小,其大小为该图像大小的1/4或1/8。

[0064] 在一些实施例中,该系统还可以使用第一机器学习模型进一步处理图像的特征图,以生成该图像的对象中心热图。可以在现场训练过程中训练该第一机器学习模型。对象

中心热图可以是描绘图像中的一个或多个像素是对象中心的可能性的图像表示。例如,该对象中心热图可以包括多个样本,每个样本具有指示该图像中对应样本是对象中心的可能性的值。因此,在该对象中心热图中,该图像中对象中心附近的样本可能比该图像中其他区域中的样本具有更高的值。在一些示例中,该第一机器学习模型可以包括一个权重向量,该权重向量包括在一个向量中的多个权重。对象中心热图的大小可以与该特征图的大小相同。在一些实施例中,该对象中心热图中每个样本的值可以是该特征图中相应样本的一个特征向量和该权重向量的点积。在一些示例中,该权重向量可以具有K个维度值,即该权重向量可以具有K个值,其中K是该图像中一个或多个对象的已知标签的数量。

[0065] 在一些实施例中,该系统还可以基于该对象中心热图的对应样本的值来确定一个或多个对象的位置。例如,可以通过滤除对象中心热图中对应样本的值低于一个阈值的位置,来选择该图像中一个由多个位置组成的列表作为对象中心。在一些实施例中,该系统可以对该对象中心热图应用过滤器以生成被平滑的对象中心热图,并使用该被平滑的对象中心热图来选择该一个或多个对象的位置。在一些示例中,滤波器可以是高斯滤波器,例如各向异性高斯滤波器。高斯滤波器的大小可以基于图像中对象的大小来确定。例如,为了检测文本字符标签,高斯滤波器的大小(例如,宽度和高度)可以基于字符大小来确定,该字符大小可以被训练并被存储在系统中。

[0066] 在一些实施例中,该系统可以通过使用第二机器学习模型和该特征图来处理该图像中的该一个或多个对象的位置,以识别该图像中的该一个或多个对象。该第二机器学习模型也可以在现场训练过程中进行训练。在从一个或多个对象的位置中的每个位置识别对象时,系统可以生成一个对象特征向量。对象特征向量可以是描述对象的某些特征的对象的表示,例如使用该预训练的机器学习模型从该图像中提取的特征。例如,可以使用该图像的该特征图的一部分来生成该对象特征向量,其中该特征图的一部分是基于相应对象的位置周围的区域。该第二机器学习模型可以包括一个权重矩阵,该权重矩阵包括一个矩阵中的多个权重。该权重矩阵中的权重可以被训练。在一些示例中,该系统可以使用该第二机器学习模型来处理该对象特征向量以生成一个类别向量,并且通过使用该类别向量将每个对象分类到该多个已知标签中的标签来识别该一个或多个对象。类别向量可以是一个或多个对象(例如,字符、机器零件或其他对象)的已知类别的表示。该类别向量可以包括多个值,每个值对应于要检测的对象的多个已知标签中的一个。

[0067] 在一些方面,上述技术可以应用于字符检测(例如,OCR)的问题。例如,该第一机器学习模型可以配置成提取该图像的某些特征,这些特征允许检测该图像中的字符位置。第二机器学习模型可以配置成提取该图像的某些其他特征,这些特征允许该系统识别该图像中的字符。该第一和第二机器学习模型中的每一个都可以包括比复杂机器学习模型少得多的权重,因此可以更快地进行训练。此外,训练这些较小的机器学习模型可能不需要庞大的训练集,这使得在现场训练这些模型更加可行。

[0068] 在一些实施例中,该系统可以使用预训练的机器学习模型来提取图该像的特征图,用于其他处理步骤。该特征图可以是一个图像的表示,并且包含该图像中的某些特征。例如,该特征图可以包括多个样本,每个样本包括一个具有多个值的向量,这些值在该图像中相应位置处对语义信息进行编码。在一些示例中,该预训练的机器学习模型可以是深度机器学习模型,例如CNN或其他神经网络模型。可以使用大量离线训练数据来训练该预训练

的机器学习模型。一旦经过训练,该预训练的机器学习模型将无需调整,从而使其在现场应用中部署变得可行。使用该预训练机器学习模型来生成该图像的特征图利用了深度机器模型在其鲁棒性和抗噪性方面的优势。

[0069] 在一些实施例中,该系统可以组合各种类型的机器学习模型。例如,发明人已经意识到,可能需要(1)使用预训练的机器学习模型来提取该图像中的某些语义特征;以及(2)使用其他机器学习模型来提取该图像的其他特征,其中该其他机器学习模型在现场部署时可以被快速训练。例如,该预训练的机器学习模型可以是深度机器学习模型,例如CNN或其他神经网络模型。该其他机器学习模型可以包括上述第一机器学习模型和第二机器学习模型。在一些实施例中,该系统可以使用该第一机器学习模型和从该预训练的机器学习模型获得的该特征图来检测这些字符的位置。附加地,该系统可以使用该第二机器学习模型和从该预训练的机器学习模型获得的特征图来识别该图像中的这些字符。深度机器学习模型(例如,给定训练这种模型所需的时间量和/或计算能力,该深度机器学习模型可以预训练)和在部署时可以训练的其他类型的模型的这种组合因此可以在系统性能和鲁棒性方面获得深度机器学习模型的益处,同时提供使该系统适应特定现场应用的灵活性。这可以克服如上文所述的传统系统中的一些或全部缺点。

[0070] 在一些方面,上述实施例的一些变型可以包括系统、计算机化方法和非暂时性指令,其可以被提供来处理包含一个或多个字符的图像并检测该一个或多个字符。检测一个或多个字符可以包括用于检测一个或多个字符的位置的操作。此外,检测图像中的字符还可以包括将一个或多个字符识别为相应的已知标签(类别)。在一些实施例中,该系统可以获得包含一个或多个字符的图像,并使用预训练的机器学习模型来处理该图像以生成该图像的特征图。在一些示例中,该预训练的机器学习模型可以是深度机器学习模型。该系统可以通过使用第一机器学习模型处理该图像的该特征图来生成字符中心热图。类似于对象中心热图,字符中心热图可以是描绘该图像中的一个或多个像素是字符中心的可能性的图像表示。例如,该字符中心热图可以包括多个样本,每个样本具有指示该图像中对应样本是字符中心的可能性的值。在一些示例中,当该图像包含文本字符时,该字符中心热图中每个样本的值可以指示该图像中对应样本是文本字符中心的可能性。因此,该字符中心热图可以表示该图像中一个字符的每个对应位置的“字符中心度”分数。例如,该字符中心热图中样本的值在该图像中靠近字符中心的位置可能较高,而在远离字符中心的位置或在该图像没有字符的区域中可能较低。这可能有助于定位该图像中的多个独立字符。

[0071] 在一些示例中,该系统可以基于该字符中心热图的对应样本的值来确定一个或多个字符的位置。例如,通过过滤掉字符中心热图中的位置,该位置对应样本的值低于一个阈值,可以选择该图像中的一个由多个位置组成的列表作为多个字符中心。在一些实施例中,该系统可以应用过滤器来平滑字符中心热图,并使用该被平滑的字符中心热图来选择该一个或多个字符的位置。在一些示例中,滤波器可以是高斯滤波器,例如各向异性高斯滤波器。高斯滤波器的大小可以基于字符的大小来确定,例如可以训练和/或存储的预存字符大小。由多个所获得的位置组成的结果列表可以包括该图像中多个字符的多个预测的中心。

[0072] 在一些实施例中,该系统可以通过使用第二机器学习模型和该特征图处理该图像中多个字符的位置来识别该图像中的一个或多个字符。在识别字符时,在一个或多个字符的每个位置,该系统可以生成一个字符特征向量。类似于对象特征向量,字符特征向量可以

是描绘该字符的某些特征的字符表示,例如在该图像的该特征图(使用预训练的机器学习模型获得)中表示的对应于该字符的多个特征。例如,可以使用该图像的该特征图的一部分来生成该字符特征向量,其中该特征图的一部分是基于相应字符位置周围的区域。该第二机器学习模型可以包括一个权重矩阵,该权重矩阵包括一个矩阵中的多个权重。该权重矩阵中的权重可以在现场训练。在一些示例中,该系统可以使用该第二机器学习模型来处理该字符特征向量以生成一个类别向量,并且通过使用该类别向量将每个字符分类到该多个已知标签中的一个标签来识别该一个或多个字符。类别向量可以是一个或多个字符的已知类别的表示。例如,该类别向量可以包括多个值,每个值对应于多个已知标签中的一个。在一个非限制性示例中,对于工业零件,该多个已知标签可以包括在工业零件的印刷标签中使用的整组字符或符号。

[0073] 在一些实施例中,上述系统可以包括在一个或多个应用程序中应用检测对象(例如,字符)的结果。例如,该图像中的该一个或多个对象可以包括该图像中包含的一个机器零件上的一个或多个印刷文本字符。因此,该系统可以配置成识别该零件上的该一个或多个文本字符,并使用至少一个已识别的文本字符跟踪该零件。在一些实施例中,该系统可以在条形码扫描仪中实施,该系统的图像捕获设备可以配置成捕获包含零件的图像,零件上具有打印的标签。该系统可以配置成通过检测条形码的位置并解码已定位的条形码来处理被捕获的图像。

[0074] 如上所述,各种实施例使用第一机器学习模型和第二机器学习模型。在一些实施例中,本文描述了在现场训练过程中训练该第一和/或第二机器学习模型的各种技术,包括系统、计算机化方法和非暂时性指令。在一些实施例中,训练过程可以包括使用初始训练数据集的初始训练过程,以及使用更新的训练数据集的重新训练处理。在重新训练过程中,更新的训练数据集可以包括初始训练数据集的一部分,和/或可以由用户例如通过用户界面获取的新训练数据。

[0075] 现场训练过程可以包括训练过程,其中被训练的机器学习模型可以使其自身在特定客户现场或生产线处适应与特定现场应用相关联的特定条件。特定条件的示例包括所使用的特定字体或打印过程、特定照明条件或特定背景、要处理的特定类型的图像,例如包含装配线上的机器零件的图像。现场训练还可以包括与用于训练预训练机器学习模型的训练数据不同的训练数据,有重叠或无重叠。现场训练过程可以包括例如用户添加定制符号(标签)并提供包含用于训练机器学习模型的新符号的训练图像。因此,现场训练可以使系统能够达到比专门预配置的系统更好的性能。

[0076] 作为示例,将详细描述可用于在检测对象时训练第一和第二机器学习模型的训练系统。在一些实施例中,训练系统可以训练第一机器学习模型用于检测对象,其中第一机器学习模型可以包括权重向量。该系统可以获得多个训练图像和对应于训练图像中的对象的训练标签。该训练标签可以包括一个由多个真实对象位置组成的列表,每个真实对象位置与该多个训练图像中对应的一个训练图像中的对象相关联,以指示该对象出现在该图像中的何处。

[0077] 在一些实施例中,该训练系统可以分别使用该多个训练图像中的一个和该预训练的机器学习模型来确定多个训练特征图,其方式类似于在用于检测字符/对象的系统中从先前描述的图像中确定特征图。例如,该预训练的机器可以是深度机器学习模型。该用于训

练机器学习模型的系统还可以包括确定真实对象中心热图,以及使用该训练特征图和该真实对象中心热图来确定第一机器学习模型的权重。在一些实施例中,真实对象中心热图中的每个样本的值可以基于到训练图像中最近的真实对象位置的距离来生成。

[0078] 在一些示例中,对于该训练图像中的每个对象,该系统还可以使用相应训练图像的训练特征图的一部分来确定一个相应的对象特征向量。对象特征向量可以是对象的表示,其描述了对象的某些特征,并且可以以与先前描述的用于检测字符/对象类似的方式获得。

[0079] 在一些示例中,该系统可以从该训练图像中的真实对象位置获得对象的位置;并且基于该对象的位置使用相应训练图像的特征图的该部分来确定该对象特征向量。确定特征图的一部分的操作可以类似于先前在用于检测字符/对象的系统中描述的操作。在确定对象特征向量时,用于训练机器学习模型的系统可以使用对准方法来确定多个子特征向量,每个子特征向量基于该特征图的该部分中的相应位置,并且连接该多个子特征向量以形成该对象特征向量。

[0080] 因此,对于所有训练图像,系统可能已经根据上述技术确定了多个对象特征向量和多个真实对象中心热图值。在确定第一机器学习模型中的权重时,系统可以使用合适的机器学习方法,使得第一机器学习模型(例如,权重向量)和每个对象特征向量之间的点积尽可能接近相应的真实对象中心热图值。用于训练权重向量的方法可以包括线性回归、SVM或其他合适的机器学习方法。

[0081] 在一些示例中,可以提供训练系统来训练该第二机器学习模型以用于检测对象(例如,字符),其中该第二机器学习模型可以包括一个权重矩阵。在一些示例中,该系统可以获得多个训练图像和训练标签。该多个训练标签中的每一个可以与多该个训练图像中的相应一个中的字符相关联。例如,该训练标签可以包括一个由多个真实字符位置组成的列表,每个位置与该多个训练图像中的相应一个中的一个字符相关联,以指示该字符出现在该图像中的何处。该系统还可以包括真实字符类别列表,该真实字符类别列表包含一个由多个类别组成的列表,这些类别表示在每个真实字符位置处的字符身份。在文本字符的情况下,字符类别可以对应于文本字符,例如“A”到“Z”或可以在现场应用中使用的其他标签。在一些示例中,系统可以通过使用标签编码方法来获得字符类别列表。例如,该系统可以收集出现在训练标签中的所有字符类别,并丢弃重复的字符类别,以生成一个由不同字符类别组成的列表。

[0082] 此外,该系统可以以与先前在训练第一机器学习模型中描述的类似方式,使用预训练的机器学习模型来确定训练特征图。例如,预训练的机器可以是深度机器学习模型。该系统还可以使用多个训练特征图和多个训练标签来确定第二机器学习模型的权重。在确定第二机器学习模型的权重时,对于训练图像中的每个对象,系统可以以与先前在训练第一机器学习模型中描述的方式类似的方式,使用对象所在的对应训练图像的特征图的一部分来确定对应的对象特征向量。

[0083] 附加地,系统还可以为训练图像中的每个对象确定目标向量。目标向量可以是描述对象与一个或多个已知类别的关联的对象(例如,字符)的表示。在一些示例中,可以使用编码方法来确定目标向量。例如,可以使用独热编码方法。因此,在确定训练图像中的每个真实对象位置的对象特征向量和目标向量之后,系统可以确定第二机器学习模型(例如,权

重矩阵),使得当该第二机器学习模型(例如,权重矩阵)与每个对象特征向量相乘时,该乘积尽可能接近地再现相应的目标向量。类似于第一机器学习的训练,可以使用机器学习方法,例如线性回归、SVM或其他合适的机器学习方法来训练第二机器学习模型中的权重矩阵。

[0084] 尽管描述了使用对象训练机器学习模型的技术,但是这种训练技术也可以应用于针对对象的多个类型(例如,检测字符)训练机器学习模型。虽然本文描述的一些示例是结合OCR技术来描述的,但是应当理解,这些示例仅仅是为了说明的目的,而不是为了限制,因为这些技术不限于OCR应用程序。

[0085] 本文描述的技术可以在提高对象/字符检测的性能方面提供优于传统系统的优势。通过使用预训练的深度神经网络作为特征提取器,系统可以使用更高层次的语义信息进行目标定位和分类。与不使用深度神经网络的系统相比,本文描述的系统更加鲁棒,并且更少受到图像噪声、对象外观变化、印刷质量变化、照明和背景差异等的影响。

[0086] 本文描述的技术还可以通过附加地使用一个或多个可以在现场训练过程中训练的其他机器学习模型,在进一步提高对象/字符检测的性能方面提供优于传统系统的优势。这些机器学习模型可以在现场训练过程中进行训练,并适应现场应用中可能出现的特定条件。因此,该系统可以使用户能够用定制提供的训练图像和/或更新标签来适配该系统。通过只训练系统的一小部分,可以使用简单的机器学习方法,例如线性回归或其他合适的方法,来有效地训练其他机器学习模型,而不像训练深度机器学习模型那样使用昂贵的计算资源或大量训练图像。

[0087] 尽管已经描述了各种实施例,但是对于本领域普通技术人员来说,显然更多的实施例和实施方式是可能的。因此,本文描述的实施例是示例,而不是唯一可能的实施例和实施方式。此外,上述优点不一定是仅有的优点,也不一定期望每个实施例都能达到所有描述的优点。

[0088] 图1是根据本文描述的技术的一些实施例的系统的示意图,该系统用于检测对象。

[0089] 在一些实施例中,系统100可包括图像捕获装置112,例如照相机,其配置成捕获零件104或检查平台103上的任何组件的视频或图像。被捕获的视频可以包括有多个帧图像组成的序列,并且每个图像可以包含检查平台103上的一个或多个零件104。系统100还可以包括服务器110,服务器110至少具有处理器、存储器、存储介质和/或其他组件。该至少一个处理器可以配置成执行存储在存储器中的编程指令,以分析图像并对每个图像执行对象检测,以检测图像中的一个或多个对象。图6是根据本文描述的技术的一些实施例的一个示例零件,该零件包括印刷在其上的可以被检测字符。在图6中,一个或多个对象602可以包括印刷在零件604上的文本字符标签。可以理解,图像中的对象可以不限于文本字符标签。

[0090] 回到图1,服务器110可以配置成在检查过程中实时处理一个或多个零件的被捕获的图像。在非限制性配置中,零件104可以放置在检查平台103上。例如,检查平台103可以是装配线的传送带或其上放置多个待检查零件的转盘。可以控制检查平台103来移动这些零件。每当检查平台103将一个零件(例如,零件104)移动到用于检查的位置时,检查平台103上的该零件被图像捕获装置112捕获,使得系统100可以处理该被捕获的图像以检测该零件上的一个或多个对象(例如,字符)。当检测完成时,检查平台103移动,使得下一个零件移动到用于检查的位置,并且捕获包含刚刚被移动到该位置的零件的新图像,并且处理该新图

像以检测该图像中的一个或多个对象。对每个要检查的零件重复该检查过程。服务器110可以配置成在检查过程中处理每个零件的捕获图像。虽然图1和图6中的示例仅示出了每个被捕获的图像一个零件104,但是应当理解,一个被捕获的图像中可以包含多个零件,可以根据需要对其进行处理以检测多个零件中的每个零件中的对象。

[0091] 系统100可以配置成检测图像中的一个或多个对象。在一些实施例中,检测一个或多个对象可以包括使用一个或多个机器学习模型来确定该一个或多个对象在图像中的位置。例如,对于零件的输入图像(例如,由图像捕获设备112捕获的图像),系统100可以配置成使用预训练的机器学习模型来处理该输入图像以提取该图像的特征图。预训练机器模型可以包括先前已训练和已存储的深度机器学习模型。例如,该特征图可以包括多个样本,每个样本是一个向量,该向量在该图像中相应位置处对语义信息进行编码。

[0092] 在一些实施例中,系统100可以使用机器学习模型(例如,第一机器学习模型)从该特征图生成对象中心热图。对象中心热图可以包括多个样本,每个样本具有指示图像中对应像素是对象中心的可能性的值。因此,在对象中心热图中,接近对象中心的样本可以具有与远离该对象中心的另一样本不同的值,其中该对象中心热图中的样本之间的值的差异可以用于将该图像中接近该对象中心的样本与远离该对象中心的样本区分开。因此,服务器110可以配置成基于该对象中心热图来确定该图像中该对象的中心位置。

[0093] 在一些实施例中,检测该一个或多个对象还可以包括通过使用另一机器学习模型(例如,第二机器学习模型)和从该预训练的机器学习模型获得的该特征图来处理该图像中的该一个或多个对象的位置来识别该图像中的该一个或多个对象。在一些示例中,对于该图像中的一个或多个已定位的对象中的每一个,系统100可以确定该图像的特征图中与该对象的位置周围的区域相对应的部分。系统100可以使用该特征图的该部分来将该对象分类到多个已知标签中的一个标签。标签可以包括该图像中一个对象的表示。例如,标签可以是文本字符的ASCII表示或者可以表示多个类别之一的任何合适的值。多个已知标签可以根据应用而变化。例如,为了检测零件上的标签,多个已知标签可以限于文本字符。在这种情况下,用于一个零件的已知标签可以包括数字和字母,并且该已知标签可以被转换成字符类别列表。因此,可以训练第二机器学习模型,使得该第二机器学习模型可以配置成输出图像中的一个或多个对象的该多个已知标签的一个类别。在诸如(图5A中)500的现场训练过程中进一步描述了生成类别列表。

[0094] 在一些实施例中,上述系统100中的第一机器学习模型和第二机器学习模型可以不同于预训练机器学习模型,因为预训练机器学习模型是深度机器学习模式,而该第一和第二机器学习模型不是深度机器学习模型。例如,第一和第二机器学习模型可以各自包括一个相应向量或矩阵,该向量或矩阵具有可以快速训练的多个权重。这对于在训练深度机器学习模型(例如,预训练机器学习模型)可能由于缺少计算资源和缺少用户处的训练数据而不可行的领域中部署系统100可能特别有用。相反,可以在现场训练第一和第二机器学习模型以提取特定于应用的某些特征。例如,在识别某些零件中的标签时,用户可以获得一组训练数据集来训练该第一和第二机器学习模型,其中该训练数据集包括训练图像,该训练图像包含其上具有标签的零件。参考图2-图4和图5A-图5B,进一步描述了可以在系统100中实施的用于使用一个或多个机器学习模型检测图像中的对象并训练该机器学习模型的方法的细节。

[0095] 进一步参考图1,一旦检测到图像中的一个或多个对象,对象数据可以被传输到用户设备114。在一些实施例中,该对象数据可以包括包含在该图像中的一个零件上的一个或多个对象的位置。在其他实施例中,该对象数据可以包括来自该图像中的该一个或多个对象的已识别的标签,例如零件上的文本字符标签。用户设备114可以配置成使用从服务器110获得的对象数据来执行与图像中的零件相关联的各种操作。

[0096] 在非限制性示例中,图像中的一个或多个对象可以包括包含在该图像中的零件上的一个或多个印刷文本字符。因此,系统100可以配置成识别该零件上的该一个或多个文本字符,并使用至少一个已识别的文本字符跟踪该零件。在一些实施例中,系统100可以在条形码扫描仪中实施,其中图像捕获设备(例如,112)可以配置成捕获图像,该图像包含其上打印有标签的零件。系统100可以配置成通过检测条形码的位置来处理被捕获的图像,并对已定位的条形码进行解码。系统100中组件的各种布置可以配置成激活特定的应用。例如,用户设备114可以配置成条形码(例如,1D或2D)扫描仪。用户设备114可以包括图像捕获设备112(例如,手机相机)以捕获其上具有条形码的零件的图像。用户设备可以配置成条形码扫描仪,以检测捕获的图像中的条形码(例如,1D或2D条形码)的位置,并基于检测到的条形码位置对图像中的条形码进行解码。各种其他应用程序也可以在系统100中实施。

[0097] 在一些实施例中,用户设备114可以设有图形用户界面来显示该对象数据。例如,该对象数据可以包括零件上的已识别的标签,并且用户设备114可以显示该已识别标签以及包含该零件的图像,使得用户可以确定该标签是否被正确识别。此外,用户设备114可以配置成如果标签被错误识别,则允许用户进行校正。图7示出了根据本文描述的技术的一些实施例的示例图形用户界面,该图形用户界面用于在训练/再训练过程中纠正标签。如图7所示,图形用户界面700可以在用户设备114(图1)中实施,可以包括部分702,部分702在第一区域702-1中显示图像中的每个对象(例如,字符)并在第二区域702-2中显示相应的已识别标签。如图所示,第一区域702-1中的字符和第二区域702-2中的已识别标签对齐,使得用户可以确定相应的标签是否被正确识别。在一些实施例中,用户界面700可以允许用户对未正确识别的标签进行校正。例如,在区域702-2中,已识别的标签可以显示在允许用户修改标签的可编辑文本框中。

[0098] 在一些实施例中,从服务器110(图1)传输到用户设备(例如,图1的114)的对象数据可以包括每个字符的位置以及已识别的标签。用户设备可以使用接收到的关于每个字符的位置信息来从被捕获的图像中分割这些字符,并将它们显示在用户界面700中,例如图7的显示区域702-1。此外,用户界面700还可以包括区域704,该区域被配置成显示包含字符的图像的一部分。

[0099] 回到图1,在一些实施例中,用户设备114可以配置成将任何被用户修改的信息发送到服务器110,例如用户通过图形用户界面700(图7)输入的被修改的标签。服务器110可以使用该被修改的信息来训练系统100中使用的一个或多个机器学习模型。在本公开中参考图4和图5A-图5B进一步描述了可以在系统100中实施的用于检测图像中的对象的训练方法的细节。

[0100] 在图1中,尽管服务器110和用户设备114被示为使用机器学习模型来执行对象检测并训练该机器学习模型,但是应当理解,单个计算设备(例如,在服务器或用户设备处)可以配置成执行该检测/训练操作。在其他变型中,本文描述的实施例也可以在分布式系统中

的多个计算设备中实施。

[0101] 图2是根据本文描述的技术的一些实施例的系统200的示意图,该系统使用机器学习模型检测对象并训练该机器学习模型。系统200可以包括可以在系统100中实施的一个或多个方框(图1)。在一些实施例中,系统200可以包括配置成使用机器学习模型从图像中执行对象检测的一个或多个方框。系统200还可以包括一个或多个方框,其配置成执行对检测操作中使用的一个或多个机器学习模型的训练。如图2所示,系统200可以配置成在检测模式和现场训练模式下操作。例如,在检测模式中,系统200可以使用预训练的机器学习模型220和其他机器学习模型(例如222、226)来执行一个或多个检测操作。附加地,在现场训练模式中,系统200可以配置成在一个或多个现场训练过程中训练其他机器学习模型(例如,222、226)。

[0102] 预训练的机器学习模型可以是在机器学习模型部署和在现场使用(例如,在客户应用中)之前被训练的机器学习模型。该预训练机器学习模型可以在不同于机器学习模型在现场应用中部署的地点进行训练。在一些实施例中,一旦部署,预训练的机器学习模型可能不会被修改。在一些实施例中,一些类型的机器学习模型(例如,222、226)可以在现场训练过程中进行训练。这些机器学习模型可以在与训练预训练机器学习模型的地点不同的地点、在训练预训练机器学习模型之后的时间和/或在与训练预训练机器学习模型的阶段不同的阶段进行训练。在非限制性示例中,预训练的机器学习模型(例如220)可以在部署系统200之前的开发/设计阶段进行训练。当部署系统200时,可以在部署阶段训练其他机器学习模型(例如222、226)。例如,部署阶段可以包括如前所述的现场训练模式,其中机器学习模型可以在现场训练过程中进行训练。

[0103] 预训练机器学习模型220可以是深度机器学习模型,并且在不同于部署系统200的地点或位置进行训练。在部署系统200时,预训练的机器学习模型220可以与由用户习惯训练的其他机器学习模型(例如222、226)结合使用。例如,用户可以收集应用特定的训练数据并使用训练系统212来训练其他机器学习模型(例如222、226)。现在进一步详细描述系统200。

[0104] 在本文描述的技术的一些实施例中,系统200可以包括特征提取器202,特征提取器202配置成处理图像并使用预训练的机器学习模型220生成该图像的特征图。作为特征提取器的输入的图像可以包括一个在系统100中被捕获的图像(例如,通过图1中的图像捕获设备112)。该预训练机器学习模型可以是深度机器学习模型,其被训练来处理输入图像并输出图像的特征图。在一些示例中,该预训练的机器学习模型可以通过包括多个图像的训练数据集来训练,该多个图像包含一个或多个对象,例如包括文本字符的标签。该其他机器学习模型(例如,222、226)可以使用不同的训练数据集来训练,例如包括从特定领域中的应用程序捕获的图像的定制数据集,例如图6中所示的图像。

[0105] 在一些实施例中,作为该预训练机器学习模型的输出,该特征图可以包括多个样本,每个样本包括一个具有多个值的向量(特征向量),这些值在该图像中的每个相应位置处对语义信息进行编码。例如,该特征图可以具有形状为h×w×D的张量,其中h是该特征图的高度,w是该特征图的宽度,D是该特征图的通道数。预训练的机器学习模型220可以配置成对图像的一些特征进行编码以供后续处理。在一些示例中,该预训练的机器学习网络可以包括特征金字塔网络(feature pyramid network)和上采样和级联网络(upsampling-

and-concatenation network)。特征金字塔网络可以是神经网络,其将图像作为输入,并输出一个或多个例如以不同比例独立生成的特征图。在非限制性示例中,特征金字塔网络可以生成不同大小的多个特征图,例如,h1×w1×D1、h2×w2×D2和h3×w3×D3。特征金字塔网络的各种实施方式是可用的。例如,可以使用在(Linetal.,"Feature Pyramid Networks for Object Detection,"December,2016,arxiv.org/abs/1612.03144)中描述的特征金字塔网络,并且其全部内容通过引用结合于此。

[0106] 在一些实施例中,该预训练机器学习网络中的上采样和级联网络可以配置成获取每个特征图,并将其上采样(例如,使用双线性插值)到共同的大小h×w。这样,多个特征图可以分别被转换成大小为:h×w×D1、h×w×D2和h×w×D3的特征图。这些特征图可以被连接以生成大小为h×w×D的输出特征图。在一些示例中,输出特征图中的通道数量D可以具有值D=D1+D2+D3。在非限制性示例中,特征图可以具有与图像不同的大小。例如,h和W可以分别取H/8和W/8的值,或者其他合适的值,其中H和W是图像的大小。图像的宽度和高度可能不相等,特征图的宽度和高度也不相等。例如,h可以具有值16,而w可以具有值24。对于h和w,任何其他合适的值都是可能的在非限制性示例中,通道数量D可以具有128的值,或者其他合适的值。当该预训练的机器学习网络是深度CNN时,通过选择CNN中的一些层并使用它们的激活作为输出,CNN可以配置成包括特征金字塔网络。这种预训练的机器学习网络可以部署在各种领域的应用中,而不需要重新训练。

[0107] 进一步参考图2,系统200可以包括对象中心热图生成器204,其配置成通过使用第一机器学习模型(例如,图2的222)处理图像的特征图来生成对象中心热图。对象中心热图可以包括多个样本,每个样本具有指示图像中的对应样本是对象中心的可能性的值。在非限制性示例中,该第一机器学习模型(例如,图2中的222)可以包括一个权重向量。该权重向量可以具有K的维度值,即,该权重向量可以具有K个值,其中K是该图像中的该一个或多个对象的已知标签的数量。在一些示例中,该权重向量中的每个值可以是浮点数。应当理解,该权重向量中的值也可以是整数。在一些实施例中,对象中心热图中每个样本的值可以是特征图中相应样本的特征向量和权重向量的点积。

[0108] 继续参考图2,系统200还可以包括配置成确定该图像中一个或多个对象的位置的对象定位器206。在一些实施例中,在该对象中心热图(由对象中心热图生成器204提供)中,图像中对象中心附近的样本可能比该图像中其他区域中的样本具有更高的值。这样,对象定位器206可以基于该对象中心热图的对应样本的值来确定一个或多个对象在图像中的位置。例如,可以通过滤除该对象中心热图中对应样本的值低于一个阈值的位置,来选择图像中的一个由多个位置组成的列表作为对象中心。

[0109] 在一些实施例中,对象定位器206可以对该对象中心热图执行平滑操作以生成被平滑的对象中心热图,并使用该被平滑的对象中心热图来选择该一个或多个对象的位置。在非限制性示例中,高斯滤波器,例如各向异性高斯滤波器,可以用于平滑操作。在一些实施例中,高斯滤波器的大小可以基于图像中对象的大小来确定。例如,为了检测文本字符标签,高斯滤波器的大小(例如,宽度和高度)可以基于对象大小224来确定,对象大小224可以被训练并被存储在系统200中。

[0110] 在一些实施例中,系统200可包括对象识别器208,其配置成通过使用第二机器学习模型(例如,图2的226)和该特征图处理该图像中的该一个或多个对象的位置(由对象定

位器206确定)来识别该图像中的该一个或多个对象,其中该第二机器学习模型226可以包括一个权重矩阵。在识别对象时,对象识别器208可以在由一个或多个对象的位置组成的列表中的每个位置处,使用该图像的该特征图的一部分来生成一个对象特征向量,其中该特征图的一部分是基于相应对象的位置的周围区域。对象识别器208可以使用第二机器学习模型226来处理该对象特征向量以生成一个类别向量,其中该类别向量包括多个值,每个值对应于多个已知标签中的一个。因此,对象识别器208可以使用该类别向量将一个对象分类到多个已知标签中的一个标签。

[0111] 进一步参考图2,系统200可以包括训练系统212,其配置成训练在检测过程中使用的各种机器学习模型,例如第一机器学习模型222和第二机器学习模型226。第一和第二机器学习模型222、226每个都可以是非深度机器学习模型,并且可以在现场训练过程中在训练系统中被有效地训练。系统200还可以包括图形用户界面210,其可以配置成提供用户界面工具以允许用户对来自对象中心热图生成器204或对象识别器208的检测结果进行修改。在用户界面210中接收到的用户修改可以被提供给训练系统212。参考图5A和图5B详细描述了用于训练该第一和第二机器学习模型的方法。

[0112] 图3是示出根据本文描述的技术的一些实施例的示例性计算机化方法300的流程图,该方法用于检测对象。例如,方法300可以在系统100中实施(图1)或200(图2)。在本文描述的一些技术中,方法300可以从动作302开始以获得图像。图像可以包括在系统中被捕获的图像(例如,通过图1中系统100的图像捕获设备112)。方法300可以在动作304使用预训练的机器学习模型来确定图像的特征图。动作304可以在特征提取器例如202(图2)中实施。在非限制性示例中,该特征提取器可以使用预训练的机器学习模型,例如深度CNN,其中输入层具有大小(例如,高度、权重、频道)。例如,作为该预训练机器学习模型输入的输入图像可以具有H×W×C的大小,其中H是图像的高度,W是图像的宽度,C是图像的通道数。应当理解,通道的数量可以是任何合适的数量。例如,在RGB彩色图像中C的值可以是3,而在灰度图像中C的值可以是1。

[0113] 在一些实施例中,作为该预训练机器学习模型输出的特征图可以具有形状为h×w×D的张量,其中h是该特征图的高度,w是该特征图的宽度,D是该特征图的通道数。换句话说,该特征图的h×w网格中的每个样本可以与相应的向量(特征向量)相关联。先前在图2中参考预训练机器学习模型220描述了预训练机器学习模型的示例。如上所述,该特征图可以包括多个样本,每个样本包括一个向量,该向量在图像中每个相应位置对语义信息进行编码。

[0114] 进一步参考图3,方法300可以包括在动作306生成对象中心热图。在一些示例中,动作306可以在对象中心热图生成器(例如,图2的204)中实施。动作306可以使用第一机器学习模型(例如,图2的222)来处理该图像的该特征图以生成该图像的对象中心热图,其中可以在现场训练过程中训练该第一机器学习模型。该对象中心热图可以包括多个样本,每个样本具有指示该图像中的对应样本是对象中心的可能性的值。该对象中心热图的大小可以与该特征图的大小相同,例如h×w。对于该对象中心热图中的样本,该图像中的对应样本可以是网格H×W中的像素。在上述示例中,其中h=H/8,w=W/8,对于该对象中心热图中的样本(x,y),图该像中的对应样本是位置(x*8,y*8)处的像素。在一些示例中,当该图像包含文本字符时,该对象中心热图中每个样本的值可以指示该图像中对应样本是字符中心的可

能性。

[0115] 在非限制性示例中,该第一机器学习模型(例如,图2中的222)可以包括一个权重向量。在上述示例中,该特征图可以具有h×w个样本,这些样本排列在与该图像中的像素阵列相对应的阵列中,其中每个样本由D维向量表示。该对象中心热图还可以包括以与该特征图中样本的排列方式相似的方式排列的多个样本,使得该对象中心热图中的一个样本可以一对一地对应于该特征图中的一个样本,该样本可以对应于该图像中的一个像素。在一些实施例中,该对象中心热图中每个样本的值可以是该特征图中相应样本的特征向量和权重向量的点积。

[0116] 在一些实施例中,该权重向量可以具有K个维度值,即该权重向量可以具有K个值,其中K是该图像中该一个或多个对象的已知标签的数量。可以训练该权重向量,使得该对象中心热图中的值可以指示该图像中的对应样本是对象中心的可能性。在一些示例中,该权重向量中的每个值可以是浮点数。应当理解,该权重向量中的值也可以是整数。参考图5A进一步描述用于训练该权重向量的训练过程。

[0117] 继续参考图3,方法300还可以包括在动作308确定一个或多个对象的位置。在一些实施例中,动作308可以在系统200中的对象定位器(例如,图2中的206)中实施。如前所述,在该对象中心热图中,该图像中对象中心附近的样本可能比该图像中其他区域中的样本具有更高的值。这样,动作308可以基于对象中心热图的对应样本的值来确定一个或多个对象在图像中的位置。例如,可以通过滤除该对象中心热图中对应样本的值低于阈值的位置,来选择该图像中的一个由多个位置组成的列表作为对象中心。

[0118] 在一些实施例中,动作308可以对该对象中心热图执行平滑操作以生成被平滑的对象中心热图,并使用该被平滑的对象中心热图来选择该一个或多个对象的位置。在一些示例中,动作308可以使用诸如高斯滤波器之类的平滑滤波器来执行平滑操作。高斯滤波器可以是各向异性的。高斯滤波器的大小可以基于图像中对象的大小来确定。例如,为了检测文本字符标签,高斯滤波器的大小(例如,宽度和高度)可以基于字符大小来确定。在一些示例中,可以训练和存储对象大小(例如,在图2的224处)供后续使用。如上所述,在动作308中获得的由多个位置组成的结果列表可以包括图像中对象的预测中心。在一些示例中,如果样本在该对象中心热图中的对应位置不在图像的网格上(例如,H×W),则可以使用插值来确定网格上的位置。

[0119] 在一些实施例中,方法300可以可选地包括在动作310识别图像中的该一个或多个对象。在一些实施例中,动作310可在对象识别器(例如,图2的208)中实施。动作310可以使用第二机器学习模型(例如,图2的226)和该特征图来处理该图像中的该一个或多个对象的位置以识别该一个或多个对象。该第二机器学习模型也可以在现场训练过程中进行训练。在识别对象时,动作310可以在由一个或多个对象的位置组成的列表中的每个位置处,使用该图像的该特征图的一部分来生成一个对象特征向量,其中该特征图的一部分是基于相应对象的位置周围的区域。动作310可以使用该第二机器学习模型来处理该对象特征向量以生成一个类别向量,其中该类别向量包括多个值,每个值对应于多个已知标签之一。因此,动作310可以使用该类别向量将一个对象分类到多个已知标签中的一个标签。

[0120] 在一些实施例中,动作310可以使用对齐方法从该特征图的一部分中提取一个特征张量。例如,动作310可以定义由多个位置(一个或多个对象的预测位置)组成的列表的每

个位置周围的区域,这些位置在动作308中获得。位置周围区域的大小可以等于预存储的对象大小(例如,图2中的对象大小224)。在一些实施例中,每个对象的区域可以在边界框中表示。该系统可以使用对准方法来使用该边界框从该图像的该特征图的一部分中提取一个特征张量。例如,该系统可以通过在该边界框内的多个位置中的每个位置处对该特征图进行采样来提取多个特征张量。在(Heetal.,"MaskR-CNN",arxiv.org/abs/1703.06870)中描述了可以使用的对准方法的一个示例,例如RoIAlign方法,其全部内容通过引用结合于本文。在非限制性示例中,上述对准方法(例如RoIAlign)可以从该特征图的一部分中提取一个特征张量。例如,被提取的特征张量可以是一个大小为3×3×D的张量,其中D是特征图中通道的数量。其他大小,例如4×4、5×5、8×8或任何其他合适的大小也是可能的。

[0121] 因此,动作310可以通过将被提取的特征张量转换成一个一维向量来确定每个对象的对象特征向量。在上述示例中,在该特征张量的大小为3x3xD的情况下,动作310可以通过将从对准方法(例如,上述RoIAlign方法)获得的多个独立向量连接成一个一维向量来确定该对象特征向量,该一维向量具有在一维上排列的该特征图的该部分的所有值。在这种情况下,该对象特征向量具有fxD值,其中f=9,D是该特征图中的通道数。

[0122] 在一些实施例中,用于识别图像中的字符的该多个已知标签可以包括可能出现在一个标签中的多个字符。例如,对于英文标签,对于英文大写字母"A"到"z",该已知标签的数量可以是26个。该第二机器学习模型(例如,图2的226)可以包括一个权重矩阵。例如,该权重矩阵的大小可以是Kx(f*D),其中D是该特征图的通道数,K是已知标签的数量(例如,在英文大写字母的情况下是26),f是取决于特征向量池的具体方法的因子。在上述例子中,f=3x3=9。

[0123] 在一些实施例中,系统可以通过将形状为K×(f*D)的权重矩阵乘以对象特征向量(大小为f*D)来确定每个对象的一个分类得分向量,从而得到大小为K的分类得分向量,其中K是已知标签的数量。因此,该类别分数向量中的每个值对应于多个已知类别(例如已知标签)之一的一个预测分数。在一些示例中,该系统可以在该类别向量中的多个值中选择最大值,其中所选择的值对应于要识别的标签。在上述示例中,用于对象检测的该已知标签可以包括多个文本字符。在一些实施例中,该已知标签还可以包括非文本类别,其可以被称为背景标签。在一些实施例中,可以该训练权重矩阵,使得该对象类别向量中的结果值可以指示该对象是对应的已知标签的可能性。用于训练该权重矩阵的训练过程将参考图5A进一步描述。

[0124] 可以对从动作308确定的该由多个位置组成的列表中的每个位置重复动作310,其中该列表中的每个位置可以对应于一个潜在的对象(或上述示例中的字符)。一旦动作310完成,该系统可以在从动作308获得的相应位置处输出一个由多个被预测字符组成的列表,其中该预测字符可以是包括一个背景标签(例如,非文本)的已知标签中的一个。在这种情况下,如果一个被预测的字符是一个背景标签,这意味着被识别的对象是非文本的或未知的标签。

[0125] 如上文参考方法300所述,该第一机器学习模型可以包括一个权重向量,该第二机器学习模型可以包括一个权重矩阵。这些机器学习模型可以通过相应的机器学习方法来训练,例如线性回归方法、支持向量机(SVM)、梯度引导决策树或其他合适的方法。因为该第一和第二机器学习模型不是深度机器学习模型,所以训练可以是快速的,使得它们适合于部

署系统的现场训练。参考图5A和图5B详细描述了用于训练该第一和第二机器学习模型的训练过程。

[0126] 图4是根据本文描述的技术的一些实施例的说明性过程400的流程图,该过程使用机器学习模型检测字符并训练该机器学习模型。在一些示例中,方法400可以在(图1的)系统100,或(图2的)系统200中实施。方法400可以类似于(图3的)方法300,除了方法400用于识别图像中的一个或多个字符。在本文描述的各种实施例中,方法400可以包括在动作402获得一个图像。类似于动作302,该图像可以包括来自系统100(图1)的被捕获的图像,例如来自图像捕获设备112。方法400可以继续在动作404使用预训练的机器学习模型来处理该图像,以生成该图像的特征图。在一些示例中,动作404可以在特征提取器(例如,图2的202)中实施。类似于动作304,该预训练的机器学习模型可以是深度机器学习模型(例如,图2中的220)。深度机器学习模型的示例可以包括具有多层的CNN。参考图3描述的预训练机器学习模型的配置也可以用于方法400中。从动作404获得的特征图可以包括多个样本,每个样本包括一个具有多个值的向量,这些值编码在应位置处的对关于该图像语义信息进行编码。

[0127] 进一步参考图4,方法400可包括在动作406生成字符中心热图。在一些示例中,动作406可以在对象中心热图生成器(图2的204)中实施。动作406可以使用第一机器学习模型(例如,图2的222)处理图像的该特征分部以生成该图像的字符中心热图。该字符中心热图可以包括多个样本,每个样本具有指示该图像中的对应样本是字符中心的可能性的值。在一些示例中,当该图像包含文本字符时,该字符中心热图中每个样本的值可以指示该图像中对应样本是文本字符中心的可能性。如在方法300中,字符中心热图的h×w网格中的样本可以对应于该图像的H×W网格中的像素。

[0128] 如上所述,字符中心热图可以表示该图像中字符的每个对应位置的“字符中心度”分数。例如,该字符中心热图中样本的值在图像中靠近字符中心的位置可能较高,而在远离字符中心的位置或在没有字符的图像区域中可能较低。这有助于定位图像中的多个独立字符,这将进一步解释。继续参考图4,方法400还可以包括在动作408确定一个或多个字符的位置。在一些实施例中,动作408可以在系统200中的对象定位器(例如,图2中的206)中实施。如前所述,在该字符中心热图中,该图像中字符中心附近的样本可能比该图像中其他区域中的样本具有更高的值。这样,动作408可以基于该字符中心热图的对应样本的值来确定该一个或多个字符在图像中的位置。例如,通过过滤掉字符中心热图中的位置,该位置对应样本的值低于一个阈值,可以选择该图像中的一个由多个位置组成的列表作为多个字符中心。

[0129] 在一些实施例中,动作408可以包括以与动作308类似的方式对该字符中心热图执行平滑操作。例如,动作408可以通过向字该符中心热图应用平滑过滤器来生成被平滑的字符中心热图,并使用该被平滑的字符中心热图来选择该一个或多个字符的位置。平滑滤波器可以是高斯滤波器,例如各向异性高斯滤波器。高斯滤波器的大小可以基于字符的大小来确定,例如预存的字符大小(例如图2的224),其可以被训练。在动作408中由多个所获得的位置组成的结果列表可以包括图像中字符的预测中心。字符大小可以是任何合适的大小。例如,在原始输入图像(从动作402获得)中,字符大小可以是60像素乘40像素。举例来说,如果该特征图大小与该图像大小相关(例如,h=H/8且w=W/8),那么该字符大小将以特

征向量样本为单位为60/8乘40/8。然后,该高斯滤波器的大小被确定为该字符大小的固定倍数。

[0130] 在一些实施例中,方法400可以包括在动作410识别该图像中的该一个或多个字符。在一些实施例中,动作410可以在对象识别器(图2的208)中实施,并且可以以与动作310中描述的类似方式实施。例如,动作410可以使用第二机器学习模型(例如,图2的226)和该特征图来处理图像中的该一个或多个字符的位置,以从存储在一个字符类别列表(例如,图2中的对象类别列表228)中的已知标签中识别该一个或多个字符。类似于方法300中的第二机器学习模型,该第二机器学习模型也可以在现场训练过程中进行训练。

[0131] 类似于动作310,动作410的输出可以包括一个由在相应位置处的被预测字符组成的列表,这些位置从动作408获得,其中该被预测字符可以是已知标签之一。在一些实施例中,已知标签还可以包括一个背景标签(例如,非文本)。类似于方法300,方法400中使用的该第一机器学习模型可以包括一个权重向量,方法400中使用的该第二机器学习模型可以包括一个权重矩阵。可以理解,分别在动作406、408中使用的该第一和第二机器学习模型不是深度机器学习模型,因此,可以在现场训练过程中有效地训练。例如,这些机器学习模型可以通过相应的机器学习方法来训练,例如线性回归方法、支持向量机(SVM)、梯度引导决策树方法或其他合适的方法。参考图5A和图5B详细描述了用于训练该第一和第二机器学习模型的训练过程。

[0132] 进一步参考图4,方法400还可以包括在412通过用户界面接收用户修改。用户界面可以在GUI(例如,图2的210)或系统100中的任何设备(例如,图1的服务器110或用户设备114)中实施。在一些示例中,系统可以在用户界面中显示来自动作410的识别结果,诸如图7中所示的示例(参见部分702)。该系统可以通过用户界面接收用户修改。用户修改可以包括一个更新的(例如,校正的)标签或一个新标签。因此,可以创建包括该图像中的字符和已知标签之一之间的新的一对一映射的新的训练数据。在一些示例中,对于一个图像中的一个字符,用户可以提供一个已知标签中不存在的新标签。在这种情况下,可以创建包括该图像中的一个字符和该新标签之间的新的一对一映射的新的训练数据。该新标签也可以存储在字符类别列表中,例如图2中的对象类别列表228。在一些示例中,可以基于新的训练数据来更新该字符大小。例如,该字符大小可以是该训练图像中所有字符的字符大小的平均值。当生成新的训练图像时,该字符大小被更新并被存储在字符大小数据库中,例如图2中的224。该新的训练数据可以随后用于训练用于该新标签的该第一和第二机器学习模型(例如,图2中的222、226)。在一些实施例中,方法400可以在动作416使用该新的训练标签来重新训练该第一机器学习模型,或者在动作418重新训练该第二机器学习模型。图5A和图5B进一步描述了用于训练该第一和第二机器学习模型的训练过程。

[0133] 本文进一步描述了在本公开中描述的各种实施例中用于训练机器学习模型的训练方法,例如系统100(图1)、系统200(图2)、训练系统(图2的212)、重新训练第一机器学习模型(图4的416)或第二机器学习模型(图4的418)的方法。这些方法可以在现场训练过程中实施。

[0134] 在一些示例中,在现场训练过程中,用户可以提供或补充包括多个训练图像和对应的多个训练标签的训练数据,其中该多个训练图像可以各自包含已知的对象。该多个训练标签可以包括每个训练图像的真实数据,其中该真实数据可以包括例如该图像中出现的

多个对象的多个标签、该图像中出现的每个对象的(多个)位置以及这些对象的大小。在一些示例中,这些信息可以由用户通过用户界面提供,例如图7所示的用户界面700。

[0135] 还通过字符标签的示例描述了在现场训练过程中提供训练数据和训练机器学习模型的方法。在非限制性示例中,该训练系统可以包括用户界面(例如,图7中的700),该用户界面配置成在显示区域(例如,704)中显示该多个训练图像中的每一个并接收用户对该图像中出现的多个字符中的每一个的位置选择(例如,注释)。该用户界面还可以接收用户输入,该用户输入可以包括一个字符位置和标签对应关系,以指示该图像中的每个字符应该被识别的内容。在一些示例中,现场训练数据可以在批处理中生成。在其他示例中,当用户正在执行识别操作(例如,图3的300或图4的400)时,可以递增地更新现场训练数据,在该过程中,用户可以输入对该图像中未正确识别的字符的校正。

[0136] 图5A是根据本文描述的技术的一些实施例的说明性过程500的流程图,该过程用于训练机器学习模型。在一些实施例中,方法500可以训练第一机器学习模型(例如,图2中的222)以用于检测字符,并且可以在系统100(图1)、系统200(图2)例如训练系统(例如,图2的212)中实施,或者重新训练第一机器学习模型(例如,图4的416)。方法500可以包括在动作502获得多个训练图像,以及在动作504获得对应于该训练图像中的字符的训练标签。举例来说,在用于检测图像中的字符的系统中,训练标签可以包括一个由真实字符位置组成的列表,每个位置与该多个训练图像中的相应一个中的字符相关联,以指示该字符出现在该图像中的何处。

[0137] 方法500还可以包括在动作506,分别使用该多个训练图像之一和预训练的机器学习模型来确定多个训练特征图。例如,可以通过使用对应的训练图像和该预训练的机器学习模型(例如,图2中的220)来确定训练特征图。该操作可以以与在动作404(图4)、304(图3)中描述的检测过程或在特征提取器202中实施的方法类似的方式来执行。

[0138] 方法500还可以包括在动作508确定真实字符中心热图,以及在动作510使用该训练特征图和该真实字符中心热图来确定该第一机器学习模型的权重。该真实特征中心热图中样本的各个值可以被称为真实特征中心值。可以基于到该训练标签中最近的真实字符位置的距离来生成该真实字符中心热图。例如,该真实字符中心热图包括多个样本,这些样本在接近训练标签中真实字符位置的中心的位置处具有高值,而在远离训练标签中真实字符位置的位置处具有低值。在一个示例实施方式中,该真实字符中心热图值可以计算为一个高斯函数,该高斯函数应用于到该训练标签中最近的真实字符位置的距离。在一些示例中,高斯函数的标准偏差或宽度可以被计算为真实字符大小的固定倍数。

[0139] 在一些实施例中,当该训练标签中的多个真实字符彼此相对接近时,该真实字符中心热图值可改为计算为两个高斯函数之间的差,这两个高斯函数应用于到这两个最近真实字符位置的距离。在一些示例中,该地面真实字符中心热图可以具有与该训练特征图相同的大小。可以使用各种使用高斯函数生成字符中心值的方法。例如,可以使用的方法在(Baeketal.,“Character Region Awareness for Text Detection,”3Apr 2019,arxiv.org/abs/1904.01941)中进行了描述,该方法通过引用整体结合于此。

[0140] 在一些实施例中,对于该训练图像中的每个字符,动作510还可以使用相应训练图像的特征图的一部分来确定一个相应的字符特征向量。例如,方法500可以从与该字符所在的相应训练图像相关联的该训练标签中的真实字符位置获得一个字符的一个位置;并且基

于该字符的位置使用相应训练图像的特征图的该部分来确定该字符特征向量。确定特征图的一部分的操作可以类似于检测过程中描述的方法,诸如在(图2的)识别器208、动作310(图3)或动作410(图4)中实施的方法。例如,可以基于真实字符位置、字符大小(例如,存储在图2的224中)以及根据从动作506获得的训练图像确定的特征图来确定字符特征向量。

[0141] 用于确定对应于图像中每个字符的特征图部分的方法以与检测过程中描述的方法类似的方式实施,例如(分别为图3、图4的)动作310、410。在一些示例中,该特征图的该部分可以由以相应训练图像中训练标签的位置为中心的边界框来表示。每个已定位字符的相应字符特征向量可以通过连接多个子特征向量来确定,每个子特征向量都是基于特征图的该部分中的相应位置而形成的。

[0142] 根据上述实施例,对于所有训练图像,该系统已经确定了总共N个字符特征向量和N个真实字符中心热图值(字符中心度值),其中N是训练特征图中网格单元的总数(例如,所有训练图像的h×w之和,其中h、w是每个训练特征图的大小)。在一些实施例中,该第一机器学习模型(例如,图2中的222)可以包括一个权重向量。训练的目标是找到一个权重向量wt,使得其与每个真实字符特征向量v的点积尽可能接近相应的真实字符中心热图值c,使得wt·v≈c。在一些实施例中,找到该权重向量中的权重可以使用任何合适的机器学习方法,例如线性回归、SVM或其他合适的机器学习方法。

[0143] 图5B是根据本文描述的技术的一些实施例的说明性过程的流程图,该过程用于训练另一机器学习模型。在一些实施例中,训练方法520可以训练第二机器学习模型(例如,图2中的226),并且可以在系统100(图1)、系统200(图2)例如训练系统(例如,图2的212)中实实,或者用于重新训练第二机器学习模型(例如,图4的418)。

[0144] 方法520可以包括以与(图5A的)方法500的动作502、504类似的方式,在动作522获得多个训练图像,并在动作524获得多个训练标签。该多个训练标签中的每一个可以与多该个训练图像中的相应一个中的字符相关联。例如,该训练标签可以包括一个由多个真实字符位置组成的列表,每个位置与该多个训练图像中的相应一个中的一个字符相关联,以指示该字符出现在该图像中的何处。该训练标签还可以包括真实字符类别列表,该真实字符类别列表包含一个由多个字符类别组成的列表,该字符类别表示在每个真实字符位置处的字符身份。在以上针对文本字符的示例中,字符类可以对应于文本字符,例如“A”到“Z”或其他标签,这些标签可以专用于部署系统的领域。在一些示例中,可以假设所有字符的字符大小大致相同。在这种情况下,存储在数据库(例如,图2的222)中的字符大小可以包括宽度和高度的两个值。应当理解,字符大小的其他合适表示也是可能的。

[0145] 在一些实施例中,系统可以通过使用标签编码方法来获得字符分类列表。例如,该系统可以收集出现在训练标签中的所有字符类别,并丢弃重复的字符类别,以生成一个由不同字符类别组成的列表。该系统可以按顺序对这些不同的字符类别进行分类。在一些示例中,该系统可以以任意顺序对不同的字符分类进行排序,以获得该字符分类列表。在一些其他示例中,该系统可以以其他合适的顺序对不同的字符类别进行排序,例如英文字母的字母顺序。一旦生成了字符类别列表,就可以将其存储在该系统中(例如,图2中的228)。

[0146] 方法520还可以包括在动作526确定训练特征图,该动作可以以与图5A的动作506类似的方式执行。例如,可以使用具有多个训练图像作为输入的预训练机器学习模型来确定该练特征图,其中该预训练机器(例如,图2中的220)可以是深度机器学习模型,例如具有

多个隐藏层的CNN。该操作可以以与在动作404(图4)、304(图3)中描述的检测过程或在特征提取器202中实施的方法类似的方式来执行。

[0147] 方法520还可以在动作528使用该多个训练特征图和该多个训练标签来确定该第二机器学习模型的权重。在一些示例中,如上文在本公开中描述的检测过程中所解释的,该第二机器学习模型可以是一个权重矩阵,其中该矩阵的大小是K×(f*D),其中D是该特征图的通道数量,K是该已知标签的数量(例如,在英文大写字母的情况下是26),f是取决于特征向量池的特定方法的因子。因此,该权重矩阵的每一行对应于已知标签的多个已知类别之一。

[0148] 在一些示例中,在确定第二机器学习模型的权重时,方法520可以针对该训练图像中的每个字符,以与方法500中确定字符特征向量类似的方式,使用该字符所在的对应训练图像的该特征图的一部分来确定一个对应的字符特征向量(例如,动作510)。方法520还可以确定多个字符中每个字符的一个目标向量。在一些实施例中,可以使用编码方法来确定该目标向量。例如,训练标签中的每个标记字符可以与长度为K的独热目标向量相关联,其中K是不同字符类别的数量。在一些实施例中,可以使用独热编码方法。例如,如果字符类别列表是["A"、"B"、"C"、"D"],并且特定标记的字符属于类别"C",则独热目标向量是向量(0,0,1,0)。可以使用其他编码方法。

[0149] 因此,在确定该多个训练图像中每个真实字符位置的字符特征向量和目标向量之后,训练过程520可以确定该第二机器学习模型(例如,权重矩阵),使得当该权重矩阵与每个字符特征向量相乘时,该乘积重新计算的结果尽可能接近相应的目标向量。类似于方法500中权重向量的训练,诸如线性回归、SVM或其他合适的机器学习方法的机器学习方法可用于训练第二机器学习模型中的权重矩阵。

[0150] 图8示出了可用于执行本文公开的技术和实施例的任何方面的计算机系统800的说明性实施。例如,计算机系统800可以安装在图1的系统100中,例如通过服务器110。计算机系统800可配置成执行如参考图1-7所述的各种方法和动作。计算机系统800可以包括一个或多个处理器810和一个或多个非暂时性计算机可读存储介质(例如,存储器820和一个或多个非易失性存储介质830)以及显示器840。处理器810可以以任何合适的方式控制向存储器820和非易失性存储设备830写入数据和从存储器820和非易失性存储设备830读取数据,因为本文描述的本发明的各方面在这方面不受限制。为了执行本文描述的功能和/或技术,处理器810可以执行存储在一个或多个计算机可读存储介质(例如,存储器820、存储介质等)中的一个或多个指令。),其可以用作存储由处理器810执行的指令的非暂时性计算机可读存储介质。

[0151] 结合本文描述的技术,用于例如检测图像/视频中的对象的代码可以存储在计算机系统800的一个或多个计算机可读存储介质上。处理器810可以执行任何这样的代码来提供如本文所述的用于检测对象的任何技术。本文描述的任何其他软件、程序或指令也可以由计算机系统800存储和执行。应当理解,计算机代码可以应用于本文描述的方法和技术的任何方面。例如,可以应用计算机代码来与操作系统交互,以通过传统的操作系统过程来检测对象。

[0152] 本文概述的各种方法或过程可被编码为可在一个或多个处理器上执行的软件,所述处理器采用各种操作系统或平台中的任何一种。此外,这种软件可以使用多种合适的编

程语言和/或编程或脚本工具中的任何一种来编写,并且还可以被编译为在虚拟机或合适的框架上执行的可执行机器语言代码或中间代码。

[0153] 在这方面,各种发明概念可以体现为至少一种非暂时性计算机可读存储介质(例如,计算机存储器、一个或多个软盘、光盘、磁带、闪存、现场可编程门阵列或其他半导体器件中的电路配置等),其编码有一个或多个程序,当在一个或多个计算机或其他处理器上执行时,这些程序实现本发明的各种实施例。一个或多个非暂时性计算机可读介质可以是可传输的,使得存储在其上的一个或多个程序可以加载到任何计算机资源上,以实施如上所述的本发明的各个方面。

[0154] 术语“程序”、“软件”和/或“应用程序”在本文中以一般意义使用,指可用于对计算机或其他处理器编程以实施上述实施例的各个方面的任何类型的计算机代码或计算机可执行指令集。此外,应当理解,根据一个方面,当被执行时执行本发明的方法的一个或多个计算机程序不需要驻留在单个计算机或处理器上,而是可以以模块化的方式分布在不同的计算机或处理器中以实施本发明的各个方面。

[0155] 计算机可执行指令可以是由一个或多个计算机或其他设备执行的多种形式,例如程序模块。通常,程序模块包括例程、程序、对象、组件、数据结构等。执行特定任务或实施特定抽象数据类型的。通常,程序模块的功能可以在各种实施例中按需组合或分布。

[0156] 此外,数据结构可以以任何合适的形式存储在非暂时性计算机可读存储介质中。数据结构可以具有通过数据结构中的位置相关的字段。这种关系同样可以通过为字段分配存储空间来达到,该存储空间具有在非暂时性计算机可读介质中传达字段之间关系的位置。然而,可以使用任何合适的机制来建立数据结构的字段中的信息之间的关系,包括通过使用指针、标签或在数据元素之间建立关系的其他机制。

[0157] 各种发明概念可以体现为一种或多种方法,已经提供了这些方法的示例。作为方法的一部分执行的动作可以以任何合适的方式排序。因此,可以构建其中以不同于所示顺序执行动作的实施例,这可以包括同时执行一些动作,即使在说明性实施例中示出为顺序动作。

[0158] 说明书和权利要求中使用的不定冠词“一”和“一个”,除非有明确的相反指示,否则应理解为表示“至少一个”如本文说明书和权利要求中所使用的,根据一个由一个或多个元素组成的列表的短语“至少一个”应被理解为表示从该由多个元素组成的列表中的任何一个或多个元素中选择的至少一个元素,但不一定包括该由多个元素组成的列表中具体列出的每个元素中的至少一个,并且不排除该由多个元素组成的列表中多个元素的任何组合。这允许元素可选地存在于短语“至少一个”所指的该由多个元素组成的列表中的特定元素之外,无论是否与该特定元素相关。

[0159] 本文在说明书和权利要求中使用的短语“和/或”应该被理解为表示如此结合的元素中的“一个或两个”,即在某些情况下结合存在而在其他情况下分离存在的元素。用“和/或”列出的多个元件应该以相同的方式解释,即“一个或多个”元素如此连接。除了由“和/或”从句具体标识的元素之外,还可以任选地存在其他元素,无论这些元素是否与具体标识的元素相关。因此,作为非限制性示例,在一个实施例中,当与诸如“包括”的开放式语言结合使用时,对“A和/或B”的引用可以仅指A(可选地包括B以外的元素);在另一个实施例中,仅包括B(可选地包括除A之外的元素);在又一实施例中,应用于A和B两者(可选地包括其他

元素);等等。

[0160] 如说明书和权利要求中所使用的,“或”应该被理解为具有与上面定义的“和/或”相同的含义。例如,当分隔列表中的项目时,“或”或“和/或”应被解释为包含性的,即包括至少一个但也包括一个以上的多个或一系列元素,并且可选地包括额外的未列出的项目。只有明确表示相反的术语,例如“仅一个”或“正好一个”,或者当在权利要求中使用时,“包括”将指包括多个或一系列元素中的正好一个元素。一般而言,本文使用的术语“或”仅应解释为前面带有排他性术语时表示排他性选择(即“一个或另一个但不是两个”),例如“任一个”、“一个”、“仅一个”或“恰好一个”当在权利要求中使用时,“基本上由...组成”应具有专利法领域中使用的普通含义。

[0161] 使用序数术语,如“第一”、“第二”、“第三”等。在权利要求中,修改权利要求元素本身并不意味着一个权利要求元素相对于另一个权利要求元素的任何优先级、优先级或顺序,或者方法的动作被执行的时间顺序。这些术语仅用作标签,以将具有特定名称的一个权利要求元素与具有相同名称的另一个元素区分开来(但使用了序数术语)。

[0162] 本文使用的措辞和术语是为了描述的目的,不应被视为限制。“包括”、“包含”、“具有”、“包含”、“涉及”及其变体的使用意味着包含其后列出的项目和附加项目。

[0163] 已经详细描述了本发明的几个实施例,本领域技术人员将容易想到各种修改和改进。这种修改和改进旨在落入本发明的方案和范围内。因此,前面的描述仅仅是示例性的,而不是限制性的。

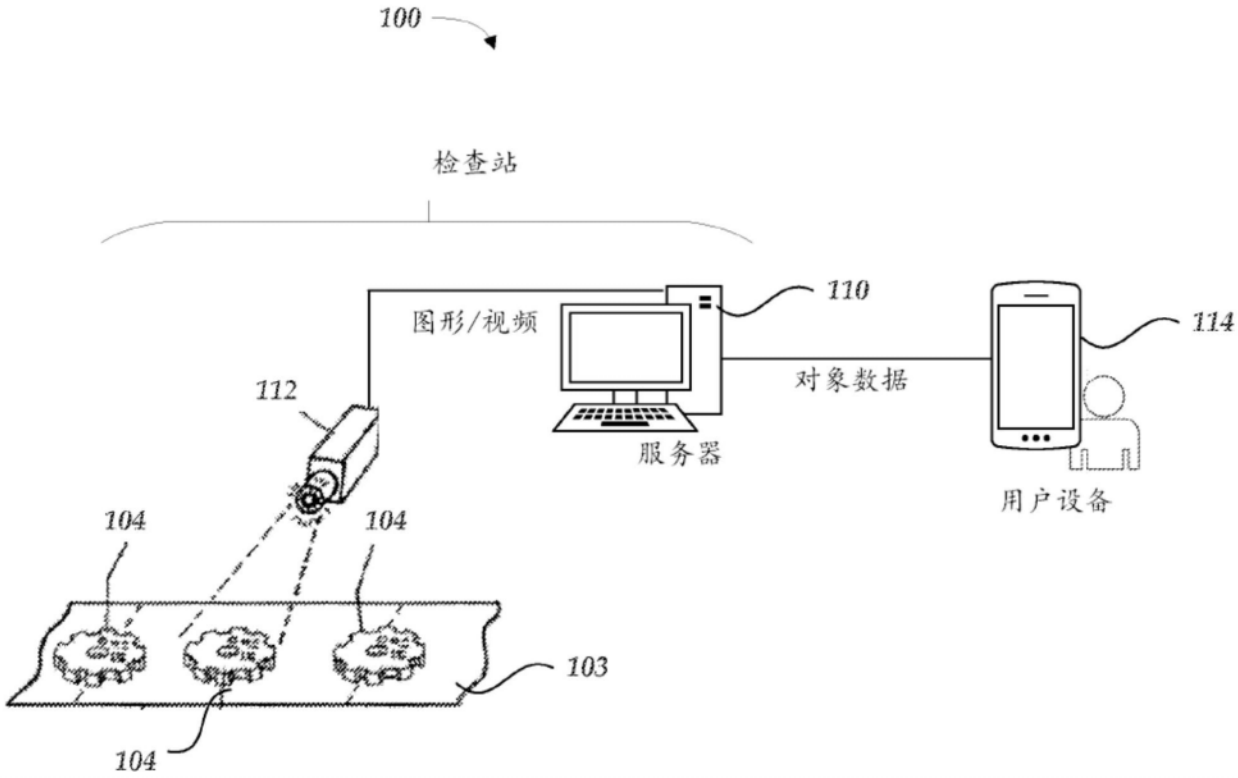


图1

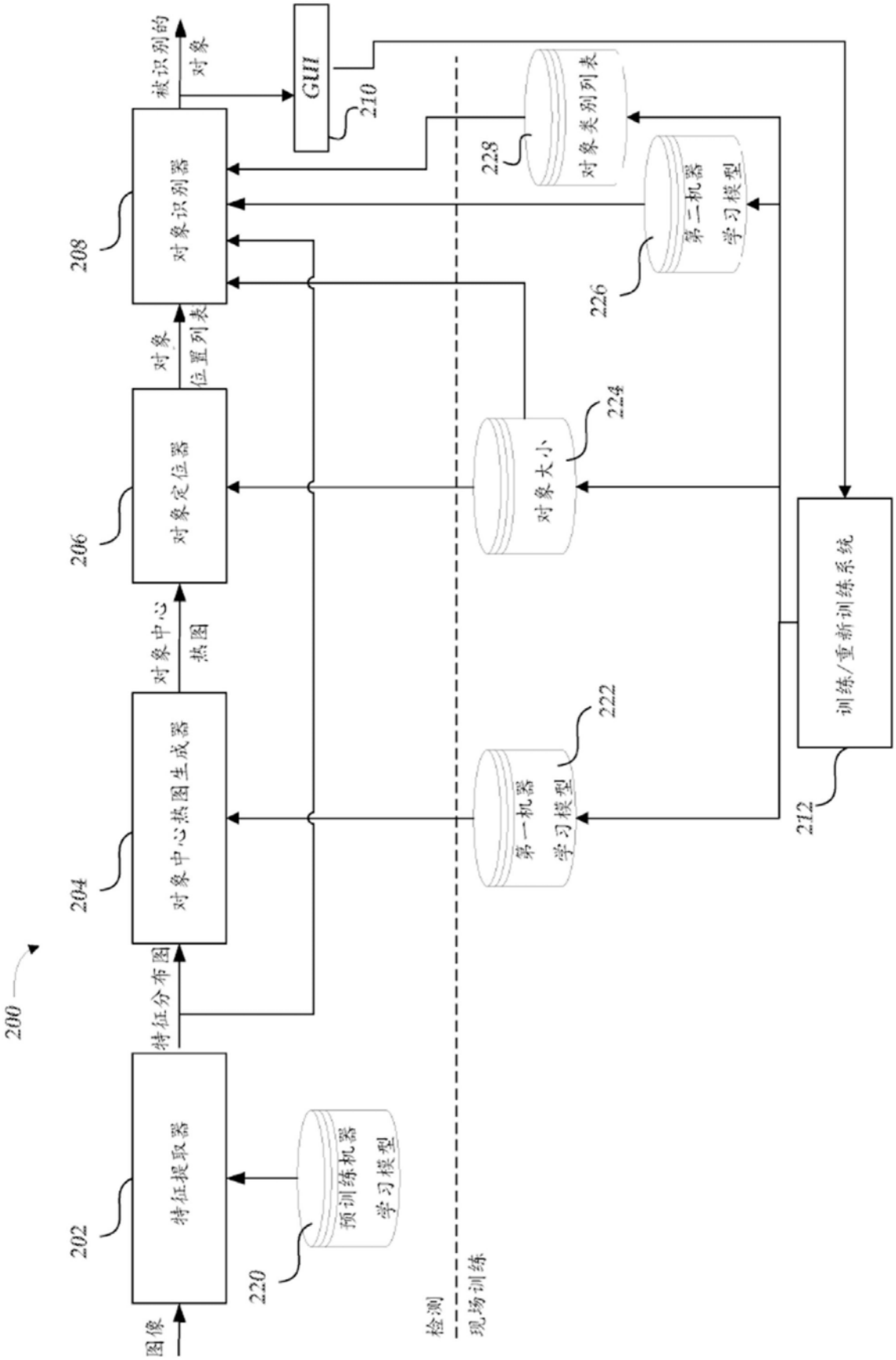

200
图像
特征提取器
202
特征分布图
对象中心热图生成器
204
对象中心热图
对象定位器
206
对象位置列表
对象识别器
208
被识别的对象
GUI
210
检测
现场训练
预训练机器学习模型
220
第一机器学习模型
222
对象大小
224
第二机器学习模型
226
对象类别列表
228
训练/重新训练系统
212

图2

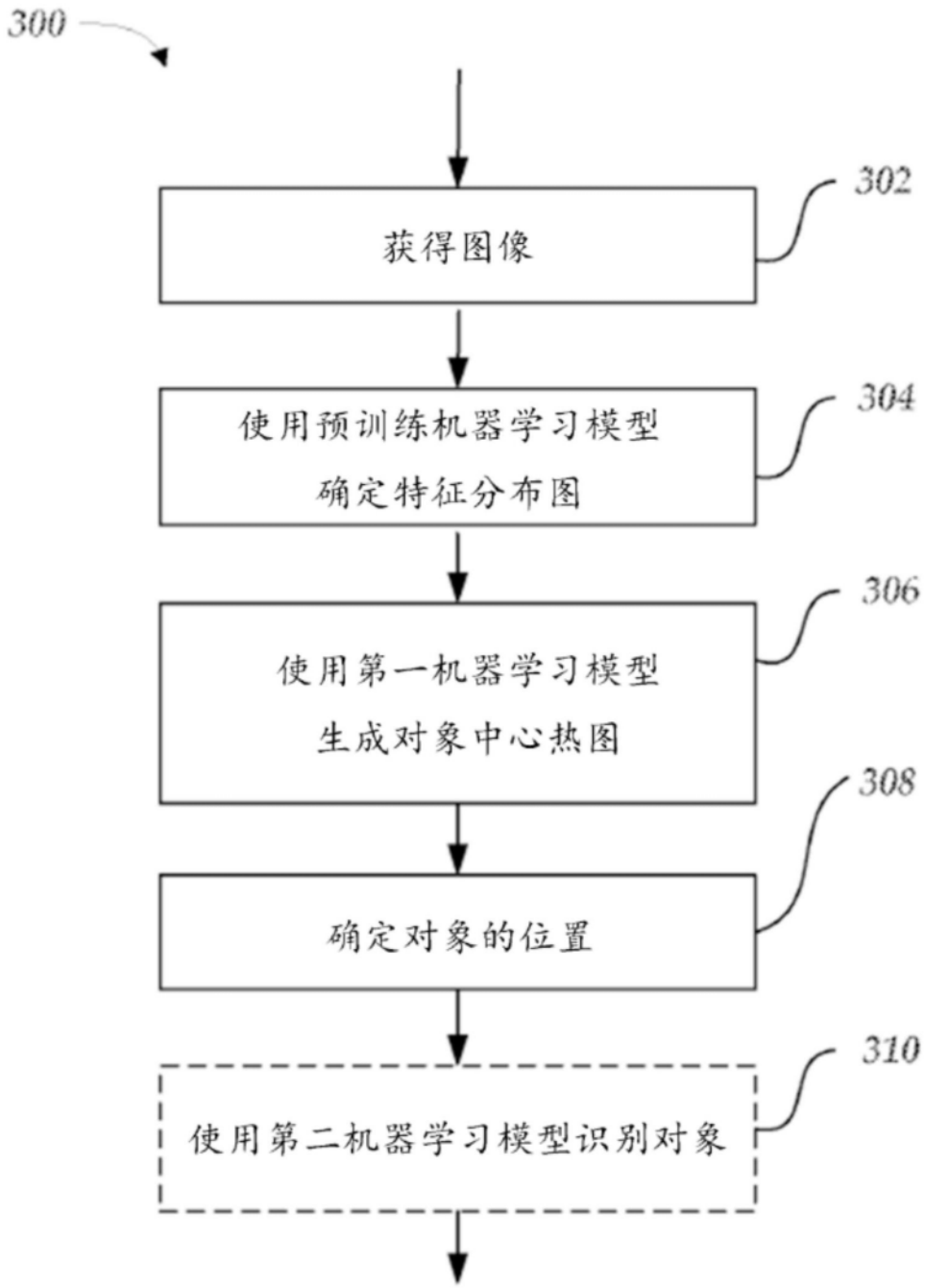

300
获得图像
302
使用预训练机器学习模型
确定特征分布图
304
使用第一机器学习模型
生成对象中心热图
306
确定对象的位置
308
使用第二机器学习模型识别对象
310

图3

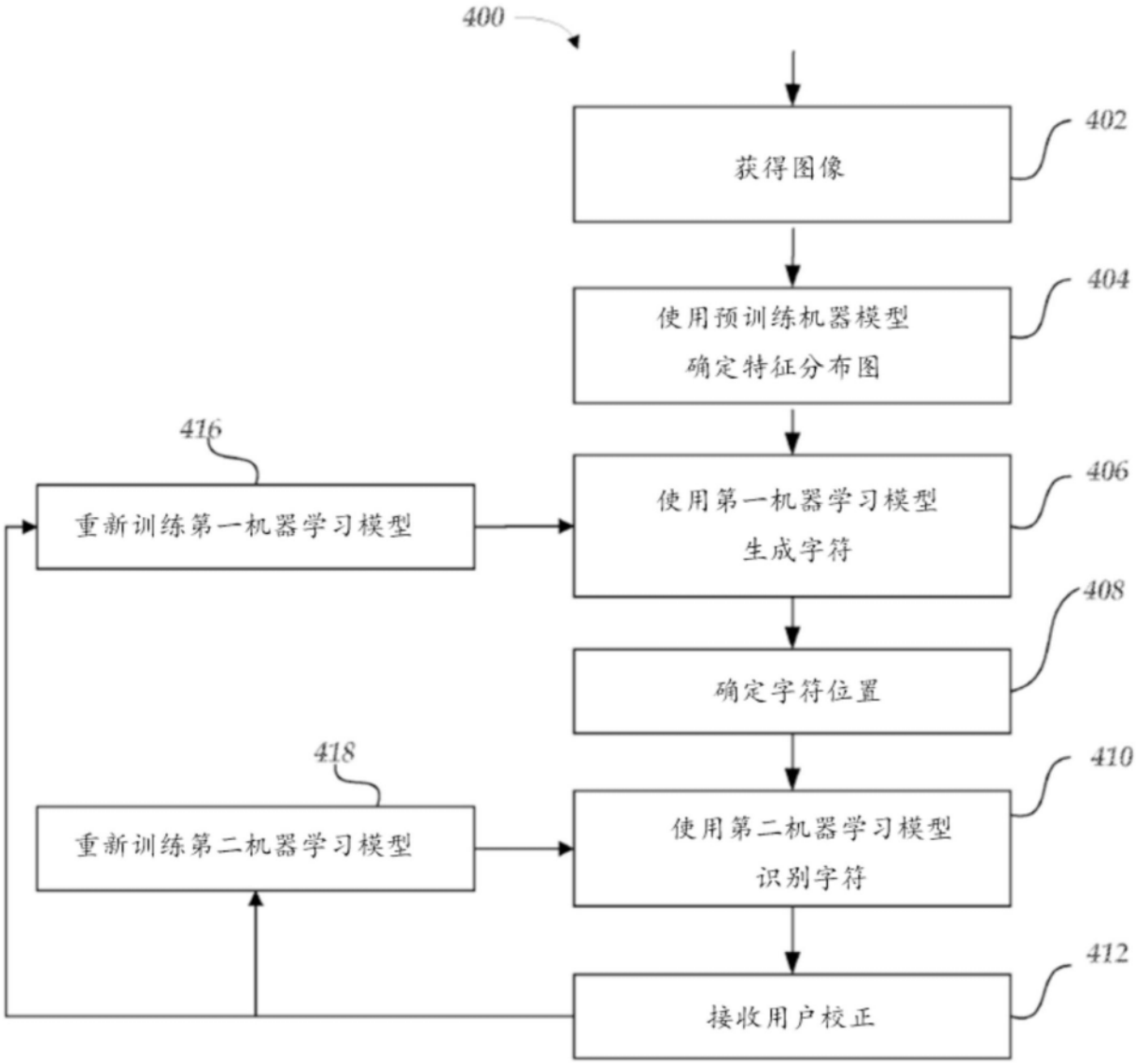

400
获得图像
402
使用预训练机器模型
确定特征分布图
404
416
重新训练第一机器学习模型
使用第一机器学习模型
生成字符
406
408
确定字符位置
418
重新训练第二机器学习模型
使用第二机器学习模型
识别字符
410
412
接收用户校正

图4

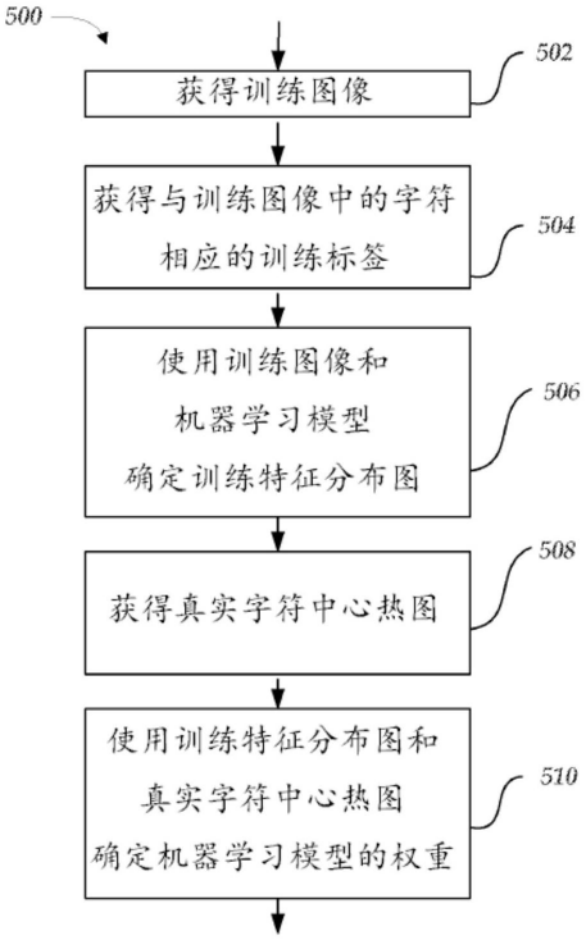

500
获得训练图像
502
获得与训练图像中的字符相应的训练标签
504
使用训练图像和机器学习模型确定训练特征分布图
506
获得真实字符中心热图
508
使用训练特征分布图和真实字符中心热图确定机器学习模型的权重
510

图5A

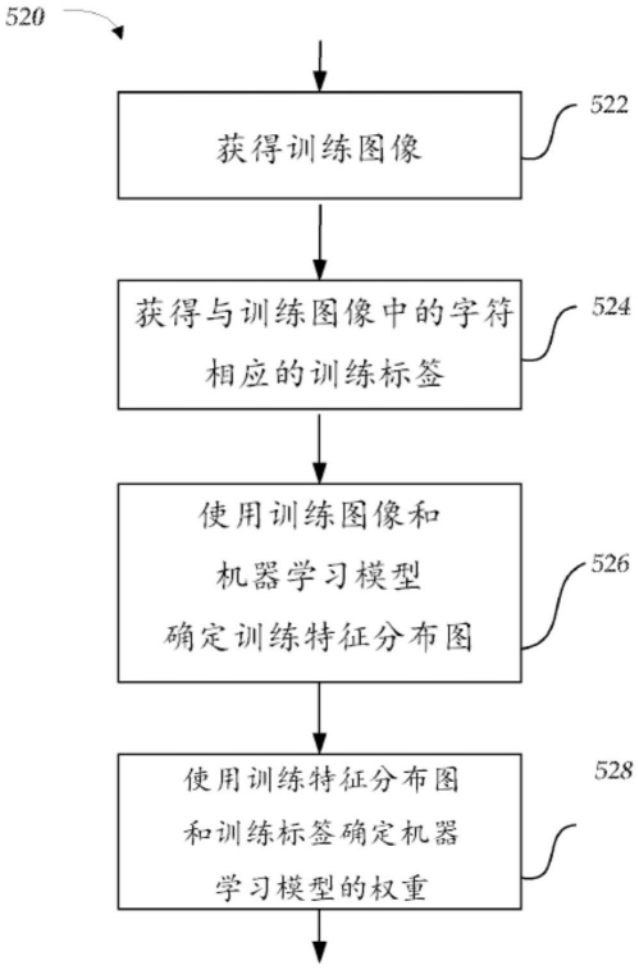

520
获得训练图像
522
获得与训练图像中的字符相应的训练标签
524
使用训练图像和机器学习模型确定训练特征分布图
526
使用训练特征分布图和训练标签确定机器学习模型的权重
528

图5B

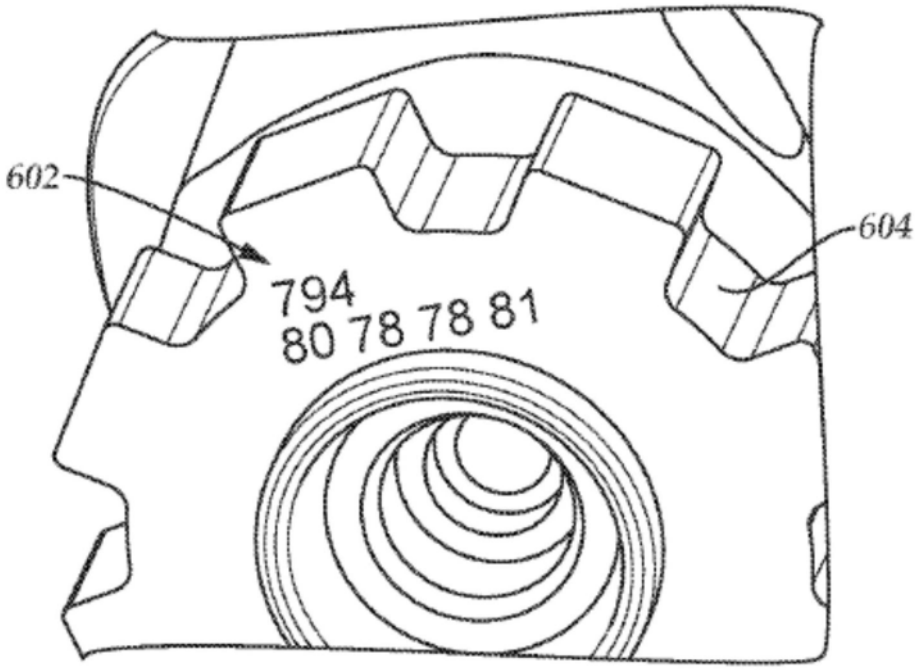


图6

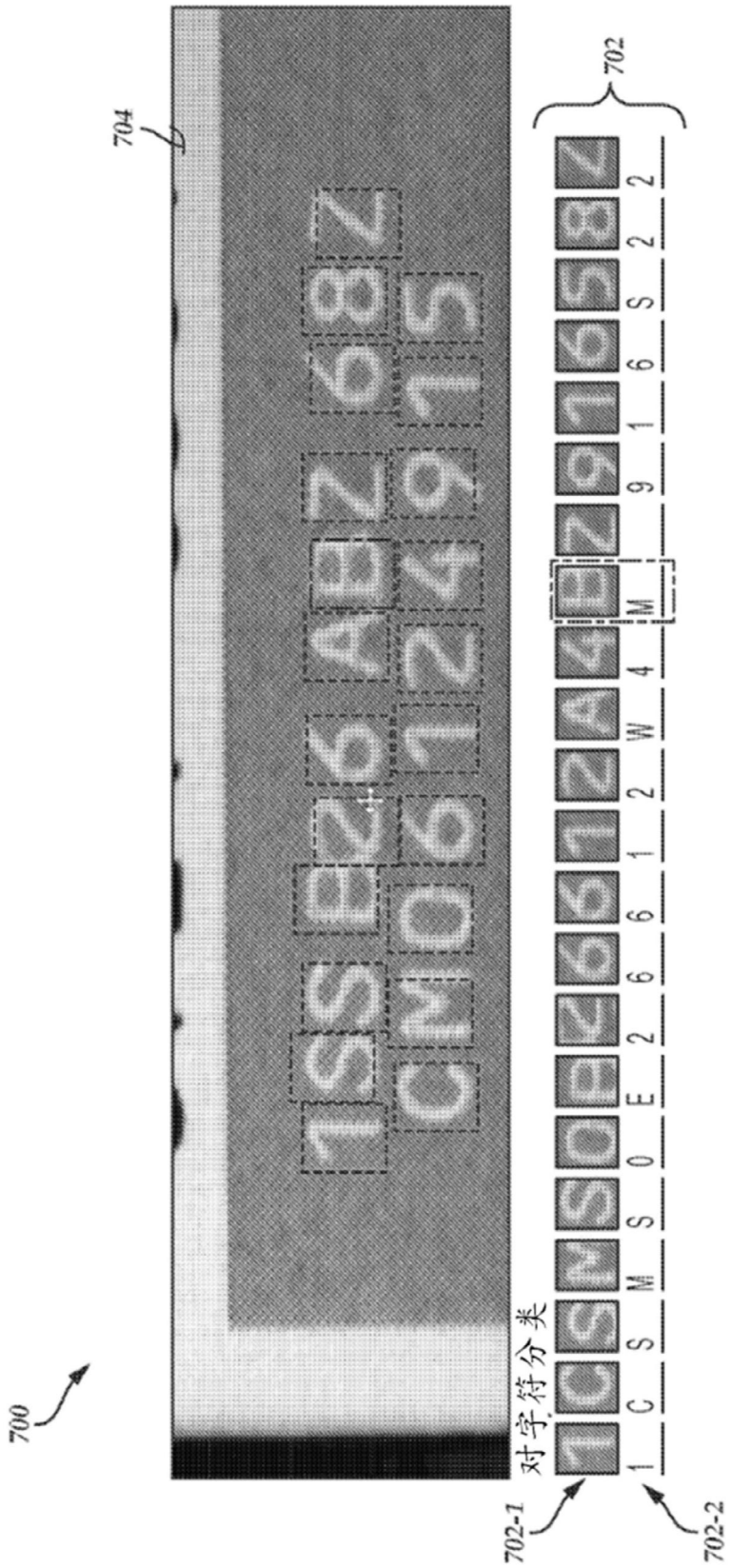


图7

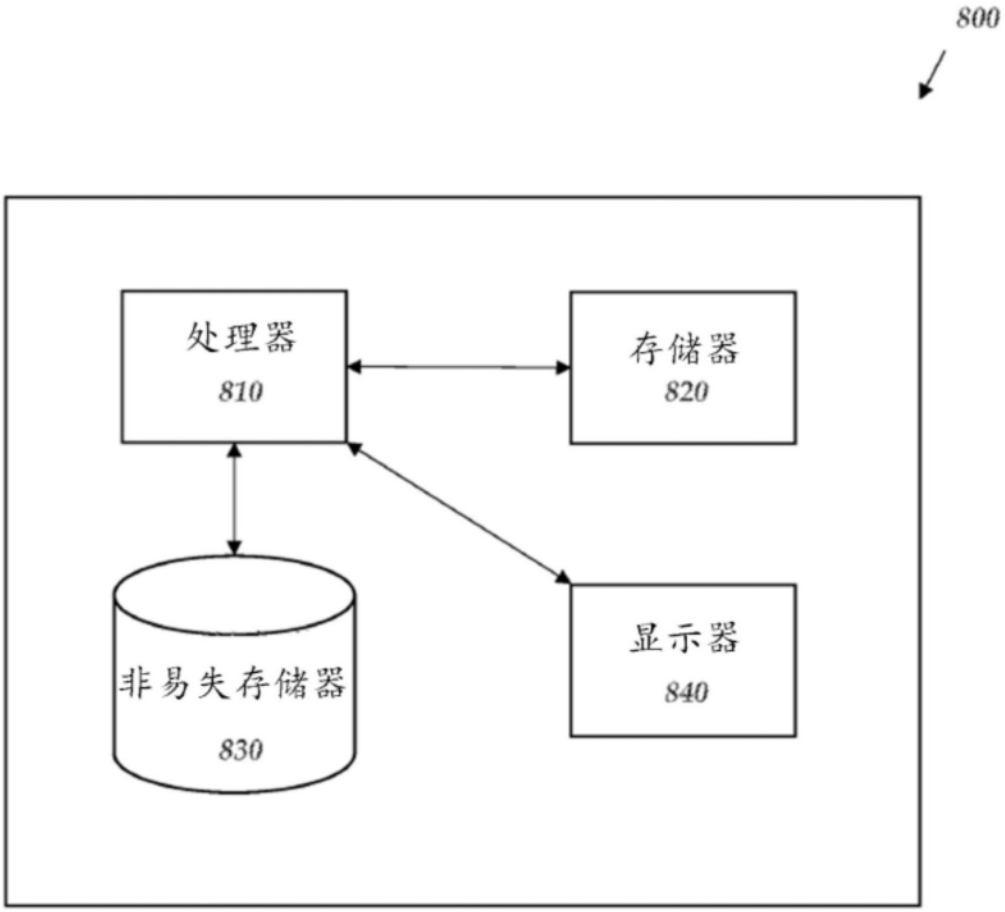


图8