



(12)发明专利

(10)授权公告号 CN 108830277 B

(45)授权公告日 2020.04.21

(21)申请号 201810362207.1

G06K 9/72(2006.01)

(22)申请日 2018.04.20

G06N 3/04(2006.01)

G06N 3/08(2006.01)

(65)同一申请的已公布的文献号

申请公布号 CN 108830277 A

(56)对比文件

(43)申请公布日 2018.11.16

CN 107273870 A,2017.10.20,

CN 107609525 A,2018.01.19,

(73)专利权人 平安科技(深圳)有限公司

US 9881207 B1,2018.01.30,

地址 518000 广东省深圳市福田区八卦岭

US 2017147905 A1,2017.05.25,

工业区平安大厦六楼

胡秀华等.一种利用物体性检测的目标跟踪

(72)发明人 王健宗 王晨羽 马进 肖京

算法.《西安电子科技大学学报(自然科学版)》

(74)专利代理机构 深圳市明日今典知识产权代

.2017,第44卷(第4期),第97-104页.

理事务所(普通合伙) 44343

审查员 李现鹏

代理人 王杰辉

(51)Int.Cl.

G06K 9/34(2006.01)

G06K 9/62(2006.01)

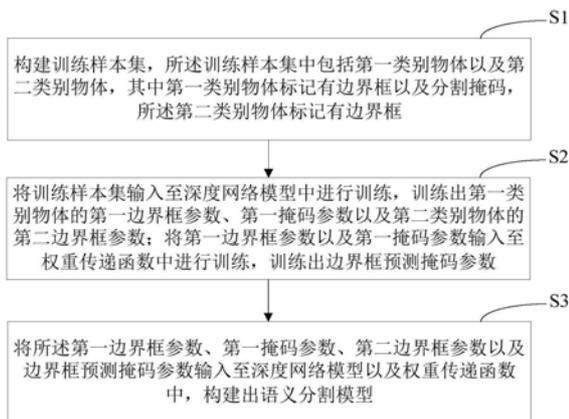
权利要求书2页 说明书12页 附图4页

(54)发明名称

语义分割模型的训练方法、装置、计算机设备和存储介质

(57)摘要

本申请提供了一种语义分割模型的训练方法、装置、计算机设备和存储介质,其中方法包括:构建训练样本集,所述训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,所述第二类别物体标记有边界框;将所述训练样本集输入至深度神经网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数;根据上述训练出的参数构建出语义分割模型。本申请大幅度降低训练样本集的标记成本,且降低后续训练样本和训练计算时的压力,提升训练效率。



1. 一种语义分割模型的训练方法,其特征在于,包括以下步骤:

构建训练样本集,所述训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,所述第二类别物体标记有边界框;

将所述训练样本集输入至深度网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数;

将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至所述深度网络模型以及所述权重传递函数中,构建出语义分割模型。

2. 根据权利要求1所述的语义分割模型的训练方法,其特征在于,所述将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型的步骤之后,包括:

将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果。

3. 根据权利要求2所述的语义分割模型的训练方法,其特征在于,所述将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果的步骤,包括:

将所述待分割图像输入至所述语义分割模型中,通过所述第一边界框参数预测出待分割图像中第一类别物体的边界框,以及通过所述第二边界框参数预测出待分割图像中第二类别物体的边界框;

根据所述第一类别物体的边界框、第二类别物体的边界框以及所述边界框预测掩码参数,分别预测出所述待分割图像中第一类别物体以及第二类别物体的掩码参数;

根据所述待分割图像中第一类别物体以及第二类别物体的掩码参数对所述待分割图像中的第一类别物体以及第二类别物体进行图像语义分割。

4. 根据权利要求1-3中任一项所述的语义分割模型的训练方法,其特征在于,所述深度网络模型为Mask-RCNN网络模型。

5. 根据权利要求4所述的语义分割模型的训练方法,其特征在于,所述权重传递函数的表达式为:

$$w_{\text{seg}}^c = \mathcal{T}(w_{\text{det}}^c; \theta)$$

$$w_{\text{det}}^c = [w_{\text{cls}}^c, w_{\text{box}}^c]$$

其中, \mathcal{T} 为传递函数, ω_{cls} 为类的权重, ω_{box} 为边界框的权重, ω_{det} 为合并向量, θ 为类别不可知的学习参数, ω_{seg} 为边界框预测掩码参数。

6. 根据权利要求5所述的语义分割模型的训练方法,其特征在于,所述权重传递函数为两层全连接神经网络,其中两层全连接层的神经元个数分别是5120以及256,其使用的激活函数为LeakyReLU。

7. 根据权利要求1所述的语义分割模型的训练方法,其特征在于,所述第二类别物体的数量多于所述第一类别物体的数量。

8. 一种语义分割模型的训练装置,其特征在于,包括:

第一构建单元,用于构建训练样本集,所述训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,所述第二类别物体标记有边界

框；

训练单元,用于将所述训练样本集输入至深度网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数;

第二构建单元,用于将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至所述深度网络模型以及所述权重传递函数中,构建出语义分割模型。

9. 一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述方法的步骤。

10. 一种计算机存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至7中任一项所述的方法的步骤。

语义分割模型的训练方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及计算机技术领域,特别涉及一种语义分割模型的训练方法、装置、计算机设备和存储介质。

背景技术

[0002] 图像语义分割的意思就是机器自动分割图像并识别出图像中的内容,比如给出一个人骑摩托车的照片,从照片中将摩托车和人分别分割出来。对图像进行语义分割时,需要得到分割掩码(segmentation mask)来对图像进行分割;目前无法根据边界框去推导出分割掩码,因此,训练时,若要实现对图像进行分割,则需要所有实例上标记有分割掩码。

[0003] 现有的语义分割方法要求所有训练实例都必须标记有分割掩码,即对每一个像素进行标注,使得注释新类别的成本十分昂贵,而且将实例分割模型限制在约100个详细注释的类别(即最多在100个类别都标记有分割掩码)。这种语义分割方法成本太高,到100种分类时,在制作训练样本和训练算力两个方面压力都很大,目前最大的实例分割公开数据集只有80个类别。因此,成本太高,不适合推广使用。

发明内容

[0004] 本申请的主要目的为提供一种语义分割模型的训练方法、装置、计算机设备和存储介质,克服训练语义分割模型时成本高的缺陷。

[0005] 为实现上述目的,本申请提供了一种语义分割模型的训练方法,包括以下步骤:

[0006] 构建训练样本集,所述训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,所述第二类别物体标记有边界框;

[0007] 将所述训练样本集输入至深度网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数;

[0008] 将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至所述深度网络模型以及所述权重传递函数中,构建出语义分割模型。

[0009] 进一步地,所述将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型的步骤之后,包括:

[0010] 将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果。

[0011] 进一步地,所述将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果的步骤,包括:

[0012] 将所述待分割图像输入至所述语义分割模型中,通过所述第一边界框参数预测出待分割图像中第一类别物体的边界框,以及通过所述第二边界框参数预测出待分割图像中第二类别物体的边界框;

[0013] 根据所述第一类别物体的边界框、第二类别物体的边界框以及所述边界框预测掩

码参数,分别预测出所述待分割图像中第一类别物体以及第二类别物体的掩码参数;

[0014] 根据所述待分割图像中第一类别物体以及第二类别物体的掩码参数对所述待分割图像中的第一类别物体以及第二类别物体进行图像语义分割。

[0015] 进一步地,所述深度网络模型为Mask-RCNN网络模型。

[0016] 进一步地,所述权重传递函数的表达式为:

$$[0017] \quad w_{\text{seg}}^c = \mathcal{T}(w_{\text{det}}^c; \theta)$$

$$[0018] \quad w_{\text{det}}^c = [w_{\text{cls}}^c, w_{\text{box}}^c]$$

[0019] 其中, \mathcal{T} 为传递函数, ω_{cls} 为类的权重, ω_{box} 为边界框的权重, ω_{det} 为合并向量, θ 为类别不可知的学习参数, ω_{seg} 为边界框预测掩码参数。

[0020] 进一步地,所述权重传递函数为两层全连接神经网络,其中两层全连接层的神经元个数分别是5120以及256,其使用的激活函数为LeakyReLU。

[0021] 进一步地,所述第二类别物体的数量多于所述第一类别物体的数量。

[0022] 本申请还提供了一种语义分割模型的训练装置,包括:

[0023] 第一构建单元,用于构建训练样本集,所述训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,所述第二类别物体标记有边界框;

[0024] 训练单元,用于将所述训练样本集输入至深度网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数;

[0025] 第二构建单元,用于将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至所述深度网络模型以及所述权重传递函数中,构建出语义分割模型。

[0026] 本申请还提供一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现上述任一项所述方法的步骤。

[0027] 本申请还提供一种计算机存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现上述任一项所述的方法的步骤。

[0028] 本申请中提供的语义分割模型的训练方法、装置、计算机设备和存储介质,具有以下有益效果:

[0029] 本申请中提供的语义分割模型的训练方法、装置、计算机设备和存储介质,构建的训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,第二类别物体标记有边界框,无需在所有类别物体上均标记分割掩码,大幅度降低训练样本集的标记成本,且降低后续训练样本和训练计算时的压力,提升训练效率;将训练样本集输入至深度网络模型以及权重传递函数中进行训练,训练出边界框预测掩码参数,便可以根据某一个类别物体的边界框预测出它的掩码参数,从而对该类别物体进行语义分割,分割成本低。

附图说明

- [0030] 图1是本申请一实施例中语义分割模型的训练方法步骤示意图；
- [0031] 图2是本申请另一实施例中语义分割模型的训练方法步骤示意图；
- [0032] 图3是本申请一实施例中语义分割模型的训练装置结构框图；
- [0033] 图4是本申请另一实施例中语义分割模型的训练装置结构框图；
- [0034] 图5是本申请另一实施例中的分割单元结构框图；
- [0035] 图6为本申请一实施例的计算机设备的结构示意图。
- [0036] 本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

[0037] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0038] 参照图1，本申请实施例中提供了一种语义分割模型的训练方法，包括以下步骤：

[0039] 步骤S1，构建训练样本集，所述训练样本集中包括第一类别物体以及第二类别物体，其中第一类别物体标记有边界框以及分割掩码，所述第二类别物体标记有边界框。

[0040] 标记边界框就是用一个方框把目标物体注释出来，标记分割掩码（掩码注释）则是把一个物体沿着其轮廓标注出来。相比较而言，边界框注释因为容易标注，所以成本低；而掩码注释难标注，所以标注成本高。

[0041] 由现有的语义分割现状可知，若要对一张图像进行语义分割，则需要得到该图像中每一个类别物体的分割掩码，而要得到分割掩码，则需要预先训练得到该类别物体对应的掩码参数，掩码参数可通过在训练样本集的图像标注分割掩码并将训练样本集输入至深度网络模型中训练得到。通俗的说，若待语义分割的图像中具有物体类别是人和树，则需要在训练样本集中注释有人和树的分割掩码，然后进行训练，得到人和树的掩码参数，然后再去待分割图像中对人和树进行分割。也就是说，待语义分割的图像中的类别有多少，则需要在制作训练样本时标注多少种分割掩码。然而，自然界的物体类别多种多样，掩码注释的成本太高，不能对所有类别的物体进行标注（目前最多只能标注80个类别）。

[0042] 本步骤S1中，所述训练样本集中包括有大量的物体类别，例如，一张图像中具有物体类别包括：人、树、汽车、火车等大量类别的物体，可以理解的是，训练样本集中的物体类别越多，越有利于后续语义分割模型对图像进行语义分割。应当注意的是，本实施例中的第一类别物体以及第二类别物体并不是单单指分别只有一种类别的物体，其表示的是一系列的物体，即第一类别物体以及第二类别物体的数量都为多种，其只需要满足第一类别物体标记有边界框以及分割掩码，所述第二类别物体只标记有边界框。

[0043] 在一实施例中，所述第二类别物体的数量远多于所述第一类别物体的数量，少量的第一类别物体，降低训练成本。如在具体的一个实施例中，训练样本集中包括一百种类别的物体，按照步骤S1的构建训练样本集的方法，其只需要对其中20种类别的物体标注有分割掩码以及边界框，而剩下的80种类别的物体只需要标注有边界框，相对于目前的技术中一百种类别的物体全部需要标注有分割掩码以及边界框而言，少标注80种类别物体的分割掩码，大幅度降低标注成本。

[0044] 本实施例中,构建的训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,第二类别物体只标记有边界框。该训练样本集是完整注释数据(带分割掩码的类别)和弱注释数据(仅带边界框的类别)的混合,因此将基于该训练样本集进行训练的任务称为偏监督任务。那么,本申请实施例中的语义分割模型的训练方法实则是一种基于偏监督学习的语义分割训练的方法。偏监督学习训练样本集的主要好处是它允许我们通过利用两种类型的现有数据集来构建一个大规模的实例分割模型。

[0045] 综上,本实施例中构建的训练样本集中只有部分类别的物体标记有分割掩码,无需在所有类别物体上均标记分割掩码,大幅度降低标记构建训练样本集的成本,且降低后续训练样本和训练计算的压力,从而提升训练效率。

[0046] 步骤S2,将所述训练样本集输入至深度网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数。

[0047] 本步骤S2中,将上述构建好的训练样本集输入深度网络模型中进行训练,根据第一类别物体标记的边界框以及分割掩码,训练得出其对应的第一边界框参数以及第一掩码参数,根据第二类别物体标记的边界框训练出其对应的第二边界框参数。本实施例中在训练深度网络模型的同时,还训练了一个权重传递函数,权重传递函数被训练成根据图片类别的边界框参数来预测类别的分割掩码参数(即实例分割参数)。权重传递函数可以在Mask R-CNN中使用带有掩码注释的类作为监督学习的数据来进行端到端的训练。在推理时,权重传递函数用于预测每个类别的实例分割参数,从而使语义分割模型能够分割所有目标的类别,包括在训练时没有掩码注释的目标类别(如第二类别物体)。

[0048] 具体地,上述权重传递函数根据上述第一边界框参数以及第一掩码参数训练出边界框预测掩码参数,该边界框预测掩码参数指的是通过某个类别物体的边界框预测出其对应掩码的一个参数。在训练得出该参数之后,则在对待分割图像进行语义分割时,则可以先预测出其对应的边界框,在通过边界框预测出其掩码参数,从而实现语义分割,具体实现过程将在下述实施例中进行阐述,在此不做过多赘述。

[0049] 由于上述构建的训练样本集只有部分类别的物体标记有分割掩码,因此,在本步骤的训练过程中,可以大幅度降低训练样本和训练计算时的压力,提升训练效率。

[0050] 步骤S3,将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型。

[0051] 上述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数即为本实施例中的语义分割模型的训练参数。若要使用语义分割模型对一张图像进行语义分割,则需要得到该图像中每一个类别物体的分割掩码,而要得到分割掩码,则需要预先训练得到该类别物体对应的掩码参数(即现有语义分割模型中的训练参数);也就是说,若要得到对图像进行语义分割的语义分割模型,则需要先得到掩码参数,而如上述步骤S1中所述的,直接进行掩码注释以获得对应的掩码参数的训练成本较高。因此,本实施例中训练得到的训练参数为上述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数,将上述训练参数对应输入至深度网络模型以及权重传递函数中,便构建出适用于对图像进行语义分割的语义分割模型。

[0052] 参照图2,在一实施例中,上述将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型的步骤S3之后,包括:

[0053] 步骤S4,将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果。

[0054] 具体地,上述将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果的步骤S4,包括:

[0055] a、将所述待分割图像输入至所述语义分割模型中,通过所述第一边界框参数预测出待分割图像中第一类别物体的边界框,以及通过所述第二边界框参数预测出待分割图像中第二类别物体的边界框。

[0056] 本步骤a中,待分割图像输入至语义分割模型中,可以直接预测出待分割图像中第一类别物体以及第二类别物体的边界框。

[0057] b、根据所述第一类别物体的边界框、第二类别物体的边界框以及所述边界框预测掩码参数,分别预测出所述待分割图像中第一类别物体以及第二类别物体的掩码参数。

[0058] 本步骤b中,将上述步骤a中预测出的第一类别物体的边界框、第二类别物体的边界框输入至权重传递函数中进行计算,则可以预测出所述待分割图像中第一类别物体以及第二类别物体的掩码参数。

[0059] c、根据所述待分割图像中第一类别物体以及第二类别物体的掩码参数对所述待分割图像中的第一类别物体以及第二类别物体进行图像语义分割。

[0060] 在上述步骤b中得出待分割图像中第一类别物体以及第二类别物体的掩码参数之后,在本步骤c中则可以得出待分割图像中第一类别物体以及第二类别物体对应的分割掩码,然后根据分割掩码对待分割图像进行图像语义分割。

[0061] 在一实施例中,所述深度网络模型为Mask-RCNN网络模型。Mask-RCNN网络模型是一种现有模型,Mask R-CNN是在Faster R-CNN上的扩展--在其已有的用于边框识别的分支上添加一个并行的用于预测目标掩码的分支。Mask R-CNN包括两个分支,一个分支是边界框预测分支,另一个分支是掩码预测分支。在Mask R-CNN中,边界框预测分支的最后一层以及掩码预测分支的最后一层均包含对每个类别执行边界框分类和实例掩码预测任务时所用的类别参数(包括边界框参数、掩码参数)。例如,通过边界框预测分支可以根据边界框预测出其边界框参数,通过掩码预测分支可以根据分割掩码预测出对应的掩码参数。

[0062] 上述Mask R-CNN方法实质是分别学习某一类别的边界框参数和掩码参数。而在本实施例中,训练Mask-RCNN网络模型的同时,还训练了一个权重传递函数。具体地,在边界框分支的最后一层以及掩码分支的最后一层之间训练上述权重传递函数,将上述训练样本集输入到Mask R-CNN网络中,同时使用权重传递函数进行训练,直到损失函数收敛,则可以训练出本实施例中的语义分割模型的训练参数。

[0063] 具体地,上述权重传递函数的表达式为:

$$[0064] \quad w_{\text{seg}}^c = \mathcal{T}(w_{\text{det}}^c; \theta)$$

$$[0065] \quad w_{\text{det}}^c = [w_{\text{cls}}^c, w_{\text{box}}^c]$$

[0066] 其中, \mathcal{T} 为传递函数, ω_{cls} 为类的权重, ω_{box} 为边界框的权重, ω_{det} 为合并向量, θ 为类别不可知的学习参数, ω_{seg} 为边界框预测掩码参数。

[0067] 在本实施例中, 上述权重传递函数为两层全连接神经网络, 其中两层全连接层的神经元个数分别是5120以及256, 其使用的激活函数为LeakyReLU, 该激活函数具有非线性、可微性、单调性等特点。

[0068] 在一实施例中, 使用语义分割模型对待分割图像进行语义分割之后, 即把物体轮廓描绘出来之后, 使用不同的RGB颜色对不同类别的物体进行标注, 以进行突出显示, 且在相邻物体之间填充差异性较大的RGB颜色; 或者可以根据物体轮廓以及场景, 识别该物体的类别, 并从数据库中调用相应(相近/相似)的图片填充在该分割出的物体上。

[0069] 在另一实施例中, 使用语义分割模型对待分割图像进行语义分割之后, 即把物体轮廓描绘出来之后, 将每一个物体对应的部分单独截取出来, 生成一个新的图层, 即将每一个物体对应生成一张图片。最后还可以选择一个与图层相应的背景与图层进行合成。

[0070] 综上所述, 为本申请实施例中提供的语义分割模型的训练方法, 构建的训练样本集中包括第一类别物体以及第二类别物体, 其中第一类别物体标记有边界框以及分割掩码, 第二类别物体只标记有边界框, 无需在所有类别物体上均标记分割掩码, 大幅度降低训练样本集的标记成本, 且降低后续训练样本和训练计算时的压力, 提升训练效率; 将训练样本集输入至深度网络模型以及权重传递函数中进行训练, 训练出边界框预测掩码参数, 便可以根据某一个类别物体的边界框预测出它的掩码参数, 从而对该类别物体进行语义分割, 分割成本低。

[0071] 参照图3, 本申请实施例中还提供了一种语义分割模型的训练装置, 包括:

[0072] 第一构建单元10, 用于构建训练样本集, 所述训练样本集中包括第一类别物体以及第二类别物体, 其中第一类别物体标记有边界框以及分割掩码, 所述第二类别物体标记有边界框;

[0073] 标记边界框就是用一个方框把目标物体注释出来, 标记分割掩码(掩码注释)则是把一个物体沿着其轮廓标注出来。相比较而言, 边界框注释因为容易标注, 所以成本低; 而掩码注释难标注, 所以标注成本高。

[0074] 由现有的语义分割现状可知, 若要对一张图像进行语义分割, 则需要得到该图像中每一个类别物体的分割掩码, 而要得到分割掩码, 则需要预先训练得到该类别物体对应的掩码参数, 掩码参数可通过在训练样本集的图像标注分割掩码并将训练样本集输入至深度网络模型中训练得到。通俗的说, 若待语义分割的图像中具有物体类别是人和树, 则需要在训练样本集中注释有人和树的分割掩码, 然后进行训练, 得到人和树的掩码参数, 然后再去待分割图像中对人和树进行分割。也就是说, 待语义分割的图像中的类别有多少, 则需要在制作训练样本时标注多少种分割掩码。然而, 自然界的物体类别多种多样, 掩码注释的成本太高, 不能对所有类别的物体进行标注(目前最多只能标注80个类别)。

[0075] 本实施例中, 第一构建单元10构建的所述训练样本集中包括有大量的物体类别, 例如, 一张图像中具有物体类别包括: 人、树、汽车、火车等大量类别的物体, 可以理解的是, 训练样本集中的物体类别越多, 越有利于后续语义分割模型对图像进行语义分割。应当注意的是, 本实施例中的第一类别物体以及第二类别物体并不是单单指分别只有一种类别的物体, 其表示的是一系列的物体, 即第一类别物体以及第二类别物体的数量都为多种, 其

只需要满足第一类别物体标记有边界框以及分割掩码,所述第二类别物体只标记有边界框。

[0076] 在一实施例中,所述第二类别物体的数量远多于所述第一类别物体的数量,少量的第一类别物体,降低训练成本。如在具体的一个实施例中,训练样本集中包括一百种类别的物体,按照第一构建单元10构建训练样本集的过程,其只需要对其中20种类别的物体标注有分割掩码以及边界框,而剩下的80种类别的物体只需要标注有边界框,相对于目前的技术中一百种类别的物体全部需要标注有分割掩码以及边界框而言,少标注80种类别物体的分割掩码,大幅度降低标注成本。

[0077] 本实施例中,构建的训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,第二类别物体只标记有边界框。该训练样本集是完整注释数据(带分割掩码的类别)和弱注释数据(仅带边界框的类别)的混合,因此将基于该训练样本集进行训练的任务称为偏监督任务。那么,本申请实施例中的语义分割模型的训练方法实则是一种基于偏监督学习的语义分割训练的方法。偏监督学习训练样本集的主要好处是它允许我们通过利用两种类型的现有数据集来构建一个大规模的实例分割模型。

[0078] 综上,本实施例中构建的训练样本集中只有部分类别的物体标记有分割掩码,无需在所有类别物体上均标记分割掩码,大幅度降低标记构建训练样本集的成本,且降低后续训练样本和训练计算的压力,从而提升训练效率。

[0079] 训练单元20,用于将所述训练样本集输入至深度网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数;

[0080] 在本实施例中,训练单元20将上述第一构建单元10构建好的训练样本集输入深度网络模型中进行训练,根据第一类别物体标记的边界框以及分割掩码,训练得出其对应的第一边界框参数以及第一掩码参数,根据第二类别物体标记的边界框训练出其对应的第二边界框参数。本实施例中在训练深度网络模型的同时,还训练了一个权重传递函数,权重传递函数被训练成根据图片类别的边界框参数来预测类别的分割掩码参数(即实例分割参数)。权重传递函数可以在Mask R-CNN中使用带有掩码注释的类作为监督学习的数据来进行端到端的训练。在推理时,权重传递函数用于预测每个类别的实例分割参数,从而使语义分割模型能够分割所有目标类别,包括在训练时没有掩码注释的目标类别(如第二类别物体)。

[0081] 具体地,上述权重传递函数根据上述第一边界框参数以及第一掩码参数训练出边界框预测掩码参数,该边界框预测掩码参数指的是通过某个类别物体的边界框预测出其对应掩码的一个参数。在训练得出该参数之后,则在对待分割图像进行语义分割时,则可以先预测出其对应的边界框,在通过边界框预测出其掩码参数,从而实现语义分割,具体实现过程将在下述实施例中进行阐述,在此不做过多赘述。

[0082] 由于上述构建的训练样本集只有部分类别的物体标记有分割掩码,因此,在本步骤的训练过程中,可以大幅度降低训练样本和训练计算时的压力,提升训练效率。

[0083] 第二构建单元30,用于将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型。

[0084] 上述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数即为本实施例中的语义分割模型的训练参数。若要使用语义分割模型对一张图像进行语义分割,则需要得到该图像中每一个类别物体的分割掩码,而要得到分割掩码,则需要预先训练得到该类别物体对应的掩码参数(即语义分割模型中的训练参数);也就是说,若要得到对图像进行语义分割的语义分割模型,则需要先得到掩码参数,而如上述实施例所述的,直接进行掩码注释以获得对应的掩码参数的训练成本较高。因此,本实施例中训练得到的训练参数为上述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数,将上述参数对应输入至深度网络模型以及权重传递函数中,便构建出适用于对图像进行语义分割的语义分割模型。

[0085] 参照图4,在一实施例中,上述语义分割模型的训练装置还包括:

[0086] 分割单元40,用于将待分割图像输入至所述语义分割模型中输出待分割图像的语义分割结果。

[0087] 具体地,参照图5,上述分割单元40包括:

[0088] 第一预测模块401,用于将所述待分割图像输入至所述语义分割模型中,通过所述第一边界框参数预测出待分割图像中第一类别物体的边界框,以及通过所述第二边界框参数预测出待分割图像中第二类别物体的边界框;

[0089] 第一预测模块401待分割图像输入至语义分割模型中,可以直接预测出待分割图像中第一类别物体以及第二类别物体的边界框。

[0090] 第二预测模块402,用于根据所述第一类别物体的边界框、第二类别物体的边界框以及所述边界框预测掩码参数,分别预测出所述待分割图像中第一类别物体以及第二类别物体的掩码参数;

[0091] 第二预测模块402将上述第一预测模块401预测出的第一类别物体的边界框、第二类别物体的边界框输入至权重传递函数中进行计算,则可以预测出所述待分割图像中第一类别物体以及第二类别物体的掩码参数。

[0092] 分割模块403,用于根据所述待分割图像中第一类别物体以及第二类别物体的掩码参数对所述待分割图像中的第一类别物体以及第二类别物体进行图像语义分割。

[0093] 在上述第二预测模块402得出待分割图像中第一类别物体以及第二类别物体的掩码参数之后,分割模块403则可以得出待分割图像中第一类别物体以及第二类别物体对应的分割掩码,然后根据分割掩码对待分割图像进行图像语义分割。

[0094] 在一实施例中,所述深度网络模型为Mask-RCNN网络模型。Mask-RCNN网络模型是一种现有模型,Mask R-CNN是在Faster R-CNN上的扩展——在其已有的用于边框识别的分支上添加一个并行的用于预测目标掩码的分支。Mask R-CNN包括两个分支,一个分支是边界框预测分支,另一个分支是掩码预测分支。在Mask R-CNN中,边界框预测分支的最后一层以及掩码预测分支的最后一层均包含对每个类别执行边界框分类和实例掩码预测任务时所用的类别参数(包括边界框参数、掩码参数)。例如,通过边界框预测分支可以根据边界框预测出其边界框参数,通过掩码预测分支可以根据分割掩码预测出对应的掩码参数。

[0095] 上述Mask R-CNN方法实质是分别学习某一类别的边界框参数和掩码参数。而在本实施例中,训练Mask-RCNN网络模型的同时,还训练了一个权重传递函数。具体地,在边界框分支的最后一层以及掩码分支的最后一层之间训练上述权重传递函数,将上述训练样本集

输入到Mask R-CNN网络中,同时使用权重传递函数进行训练,直到损失函数收敛,则可以训练出本实施例中的语义分割模型的训练参数。

[0096] 具体地,上述权重传递函数的表达式为:

$$[0097] \quad w_{\text{seg}}^c = \mathcal{T}(w_{\text{det}}^c; \theta)$$

$$[0098] \quad w_{\text{det}}^c = [w_{\text{cls}}^c, w_{\text{box}}^c]$$

[0099] 其中, \mathcal{T} 为传递函数, ω_{cls} 为类的权重, ω_{box} 为边界框的权重, ω_{det} 为合并向量, θ 为类别不可知的学习参数, ω_{seg} 为边界框预测掩码参数。

[0100] 在本实施例中,上述权重传递函数为两层全连接神经网络,其中两层全连接层的神经元个数分别是5120以及256,其使用的激活函数为LeakyReLU,该激活函数具有非线性、可微性、单调性等特点。

[0101] 在一实施例中,使用语义分割模型对待分割图像进行语义分割之后,即把物体轮廓描绘出来之后,使用不同的RGB颜色对不同类别的物体进行标注,以进行突出显示,且在相邻物体之间填充差异性较大的RGB颜色;或者可以根据物体轮廓以及场景,识别该物体的类别,并从数据库中调用相应(相近/相似)的图片填充在该分割出的物体上。

[0102] 在另一实施例中,使用语义分割模型对待分割图像进行语义分割之后,即把物体轮廓描绘出来之后,将每一个物体对应的部分单独截取出来,生成一个新的图层,即将每一个物体对应生成一张图片。最后还可以选择一个与图层相应的背景与图层进行合成。

[0103] 综上所述,为本申请实施例中提供的语义分割模型的训练装置,构建的训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,第二类别物体只标记有边界框,无需在所有类别物体上均标记分割掩码,大幅度降低训练样本集的标记成本,且降低后续训练样本和训练计算时的压力,提升训练效率;将训练样本集输入至深度网络模型以及权重传递函数中进行训练,训练出边界框预测掩码参数,便可以根据某一个类别物体的边界框预测出它的掩码参数,从而对该类别物体进行语义分割,分割成本低。

[0104] 参照图6,本申请实施例中还提供一种计算机设备,该计算机设备可以是服务器,其内部结构可以如图6所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中,该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储语义分割模型等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种语义分割模型的训练方法。

[0105] 上述处理器执行上述语义分割模型的训练方法的步骤:构建训练样本集,所述训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,所述第二类别物体标记有边界框;

[0106] 将所述训练样本集输入至深度网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框

参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数;

[0107] 将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型。

[0108] 在一实施例中,上述处理器将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型的步骤之后,包括:

[0109] 将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果。

[0110] 在一实施例中,上述处理器将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果的步骤,包括:

[0111] 将所述待分割图像输入至所述语义分割模型中,通过所述第一边界框参数预测出待分割图像中第一类别物体的边界框,以及通过所述第二边界框参数预测出待分割图像中第二类别物体的边界框;

[0112] 根据所述第一类别物体的边界框、第二类别物体的边界框以及所述边界框预测掩码参数,分别预测出所述待分割图像中第一类别物体以及第二类别物体的掩码参数;

[0113] 根据所述待分割图像中第一类别物体以及第二类别物体的掩码参数对所述待分割图像中的第一类别物体以及第二类别物体进行图像语义分割。

[0114] 在一实施例中,上述深度网络模型为Mask-RCNN网络模型。

[0115] 在一实施例中,上述权重传递函数的表达式为:

$$[0116] \quad w_{\text{seg}}^c = \mathcal{T}(w_{\text{det}}^c; \theta)$$

$$[0117] \quad w_{\text{det}}^c = [w_{\text{cls}}^c, w_{\text{box}}^c]$$

[0118] 其中, \mathcal{T} 为传递函数, ω_{cls} 为类的权重, ω_{box} 为边界框的权重, ω_{det} 为合并向量, θ 为类别不可知的学习参数, ω_{seg} 为边界框预测掩码参数。

[0119] 在一实施例中,上述权重传递函数为两层全连接神经网络,其中两层全连接层的神经元个数分别是5120以及256,其使用的激活函数为LeakyReLU。

[0120] 在一实施例中,上述第二类别物体的数量多于所述第一类别物体的数量。

[0121] 本领域技术人员可以理解,图6中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定。

[0122] 本申请一实施例还提供一种计算机存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现一种语义分割模型的训练方法,具体为:构建训练样本集,所述训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,所述第二类别物体标记有边界框;

[0123] 将所述训练样本集输入至深度网络模型中进行训练,训练出所述第一类别物体的第一边界框参数、第一掩码参数以及第二类别物体的第二边界框参数;将所述第一边界框参数以及第一掩码参数输入至权重传递函数中进行训练,训练出边界框预测掩码参数;

[0124] 将所述第一边界框参数、第一掩码参数、第二边界框参数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型。

[0125] 在一实施例中,上述处理器将所述第一边界框参数、第一掩码参数、第二边界框参

数以及边界框预测掩码参数输入至深度网络模型以及权重传递函数中,构建出语义分割模型的步骤之后,包括:

[0126] 将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果。

[0127] 在一实施例中,上述处理器将待分割图像输入至所述语义分割模型中以输出待分割图像的语义分割结果的步骤,包括:

[0128] 将所述待分割图像输入至所述语义分割模型中,通过所述第一边界框参数预测出待分割图像中第一类别物体的边界框,以及通过所述第二边界框参数预测出待分割图像中第二类别物体的边界框;

[0129] 根据所述第一类别物体的边界框、第二类别物体的边界框以及所述边界框预测掩码参数,分别预测出所述待分割图像中第一类别物体以及第二类别物体的掩码参数;

[0130] 根据所述待分割图像中第一类别物体以及第二类别物体的掩码参数对所述待分割图像中的第一类别物体以及第二类别物体进行图像语义分割。

[0131] 在一实施例中,上述深度网络模型为Mask-RCNN网络模型。

[0132] 在一实施例中,上述权重传递函数的表达式为:

$$[0133] \quad w_{\text{seg}}^c = \mathcal{T}(w_{\text{det}}^c; \theta)$$

$$[0134] \quad w_{\text{det}}^c = [w_{\text{cls}}^c, w_{\text{box}}^c]$$

[0135] 其中, \mathcal{T} 为传递函数, ω_{cls} 为类的权重, ω_{box} 为边界框的权重, ω_{det} 为合并向量, θ 为类别不可知的学习参数, ω_{seg} 为边界框预测掩码参数。

[0136] 在一实施例中,上述权重传递函数为两层全连接神经网络,其中两层全连接层的神经元个数分别是5120以及256,其使用的激活函数为LeakyReLU。

[0137] 在一实施例中,上述第二类别物体的数量多于所述第一类别物体的数量。

[0138] 综上所述,为本申请实施例中提供的语义分割模型的训练方法、装置、计算机设备和存储介质,构建的训练样本集中包括第一类别物体以及第二类别物体,其中第一类别物体标记有边界框以及分割掩码,第二类别物体只标记有边界框,无需在所有类别物体上均标记分割掩码,大幅度降低训练样本集的标记成本,且降低后续训练样本和训练计算时的压力,提升训练效率;将训练样本集输入至深度网络模型以及权重传递函数中进行训练,训练出边界框预测掩码参数,便可以根据某一个类别物体的边界框预测出它的掩码参数,从而对该类别物体进行语义分割,分割成本低。

[0139] 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流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的和实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可以包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM通过多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双速据率SDRAM(SSRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0140] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其它变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、装置、物品或者方法不仅包括那些要素,而且还包括没有明确列出的其它要素,或者是还包括为这种过程、装置、物品或者方法所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

[0141] 以上所述仅为本申请的优选实施例,并非因此限制本申请的专利范围,凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其它相关的技术领域,均同理包括在本申请的专利保护范围内。

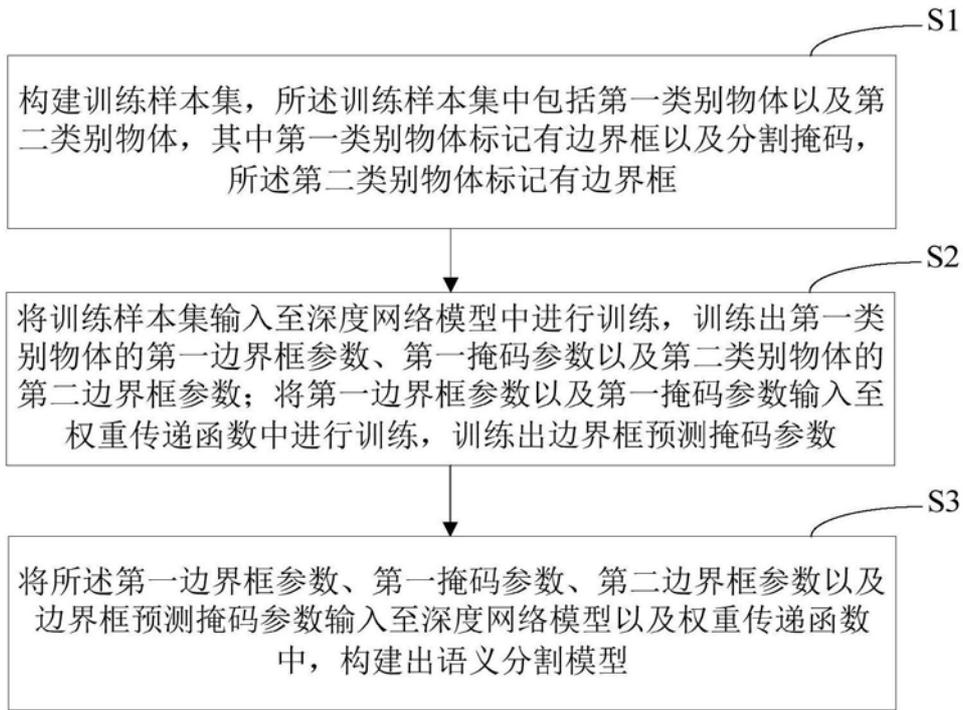


图1

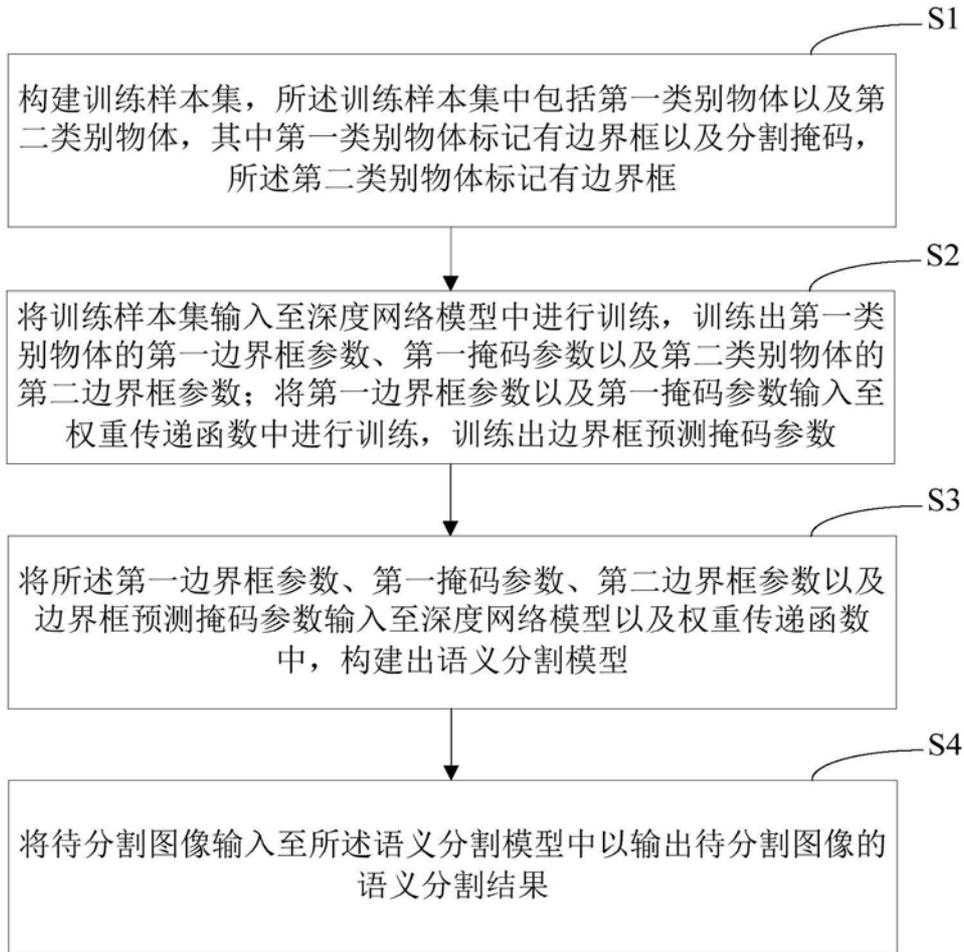


图2

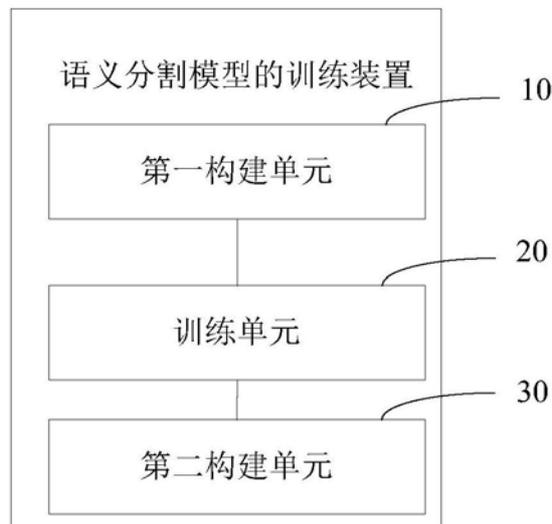


图3

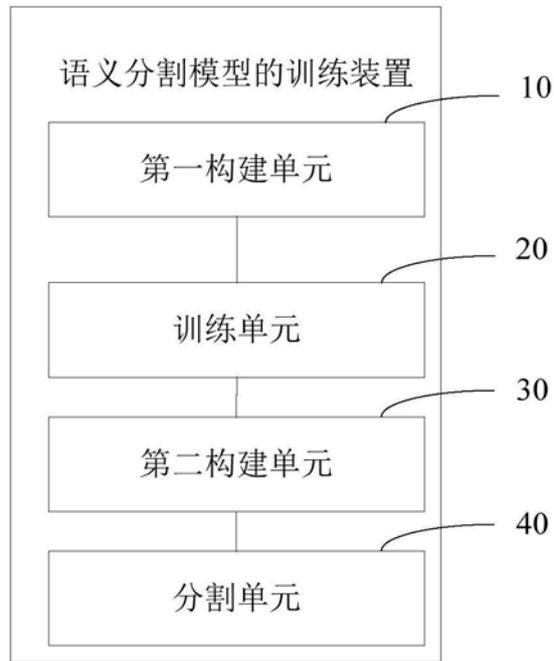


图4

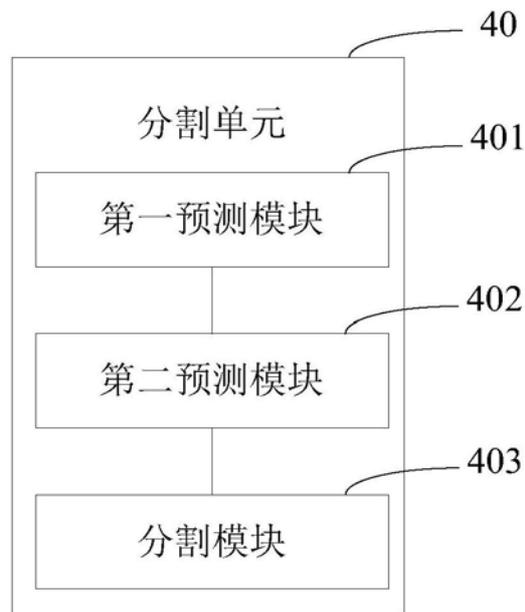


图5

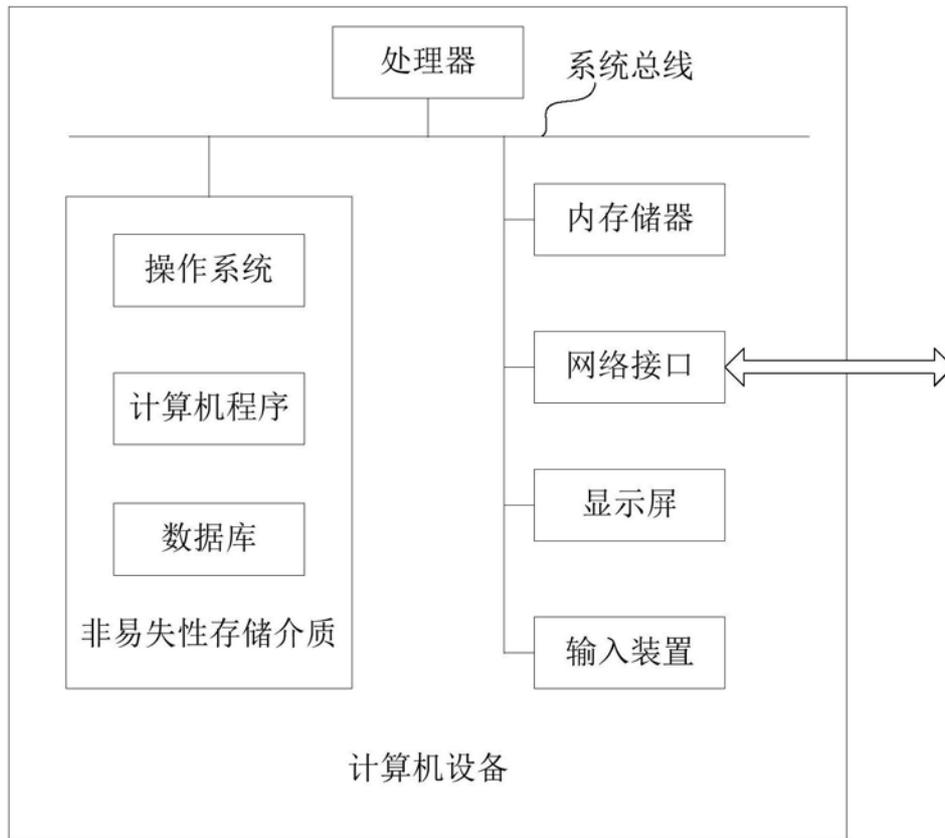


图6