



(12)发明专利

(10)授权公告号 CN 108229479 B

(45)授权公告日 2019.12.31

(21)申请号 201710648545.7

G06K 9/62(2006.01)

(22)申请日 2017.08.01

G06N 3/04(2006.01)

(65)同一申请的已公布的文献号

G06N 3/08(2006.01)

申请公布号 CN 108229479 A

(56)对比文件

(43)申请公布日 2018.06.29

CN 106022221 A, 2016.10.12,

(73)专利权人 北京市商汤科技开发有限公司

WO 2017091833 A1, 2017.06.01,

地址 100084 北京市海淀区中关村东路1号
院3号楼7层710-712房间

CN 105787482 A, 2016.07.20,

审查员 李原野

(72)发明人 詹晓航 刘子纬 罗平 吕健勤
汤晓鸥

(74)专利代理机构 北京思源智汇知识产权代理
有限公司 11657

代理人 毛丽琴

(51)Int.Cl.

G06K 9/34(2006.01)

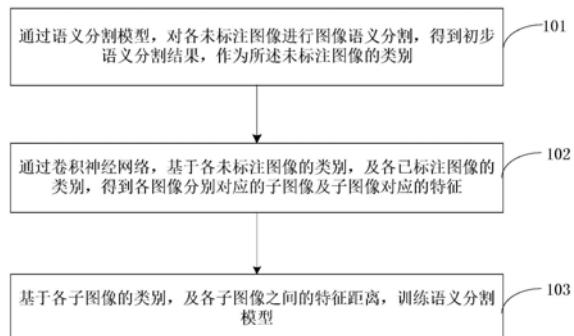
权利要求书4页 说明书11页 附图3页

(54)发明名称

语义分割模型的训练方法和装置、电子设备、存储介质

(57)摘要

本发明实施例公开了一种语义分割模型的训练方法和装置、电子设备、存储介质，其中方法包括：通过语义分割模型，对各未标注图像进行图像语义分割，得到初步语义分割结果，作为所述未标注图像的类别；通过卷积神经网络，基于各所述未标注图像的类别，及各已标注图像的类别，得到各图像分别对应的子图像及子图像对应的特征，所述各图像包括各所述未标注图像及各所述已标注图像，所述各子图像携带有对应图像的类别；基于各子图像的类别，及各子图像之间的特征距离，训练语义分割模型。本发明上述实施例训练得到的语义分割模型，在语义分割上能获得较高的准确率。



1. 一种语义分割模型的训练方法,其特征在于,包括:

通过语义分割模型,对各未标注图像进行图像语义分割,得到初步语义分割结果,作为所述未标注图像的类别;

通过卷积神经网络,基于各所述未标注图像的类别,及各已标注图像的类别,得到各图像分别对应的子图像及子图像对应的特征,所述各图像包括各所述未标注图像及各所述已标注图像,所述各子图像携带有对应图像的类别;

基于各子图像的类别,及各子图像之间的特征距离,训练语义分割模型。

2. 根据权利要求1所述的方法,其特征在于,基于各子图像的类别,及各子图像之间的特征距离,训练语义分割模型,包括:

根据子图像之间的类别关系建立面片图,所述面片图包括节点和边,所述节点包括所述子图像,所述边包括任意两个所述子图像之间的特征距离;

对所述语义分割模型进行训练,使得所述面片图中,类别相同的两个子图像之间的特征距离小于第一预设值,类别不相同的两个子图像之间的特征距离大于第二预设值。

3. 根据权利要求2所述的方法,其特征在于,根据子图像之间的类别关系建立面片图,包括:

选择至少一个子图像作为基准节点,分别针对各基准节点:

将与所述基准节点相同类别的子图像作为正相关节点,将与所述基准节点不同类别的子图像作为负相关节点,分别在所述基准节点与各所述正相关节点之间建立正相关连接,分别在所述基准节点与各所述负相关节点建立负相关连接;

由各所述基准节点及其所述正相关节点、所述负相关节点、所述正相关连接以及所述负相关连接形成稀疏连接的所述面片图。

4. 根据权利要求2所述的方法,其特征在于,所述对所述语义分割模型进行训练,包括:

通过梯度反向传播算法,对所述语义分割模型进行训练,以使所述卷积神经网络的误差达到最小化,所述误差为基于所述卷积神经网络获得的对应子图像的特征的三重损失。

5. 根据权利要求4所述的方法,其特征在于,所述通过梯度反向传播算法,对所述语义分割模型进行训练,包括:

根据所述建立的面片图中的子图像之间的特征距离通过损失函数计算得到最大误差;

将最大误差通过梯度反向传播,计算所述卷积神经网络中每一层的误差;

根据所述每一层的误差计算出每一层参数的梯度,根据所述梯度修正所述卷积神经网络中对应层的参数;

根据优化参数后的卷积神经网络输出的子图像之间的距离计算得到误差,将所述误差作为最大误差;

迭代执行将最大误差通过梯度反向传播,计算所述卷积神经网络中每一层的误差;根据所述每一层的误差计算出每一层参数的梯度,根据所述梯度修正所述卷积神经网络中对应层的参数,直到所述最大误差小于或等于预设值。

6. 根据权利要求4所述的方法,其特征在于,对所述语义分割模型进行训练,包括:

基于所述卷积神经网络的训练结果获得所述卷积神经网络的参数;

基于获得的所述卷积神经网络的参数初始化所述语义分割模型中的参数。

7. 根据权利要求1-6任一所述的方法,其特征在于,基于所述未标注图像的类别,及已

标注图像的类别,得到各图像分别对应的子图像,包括:

响应于预设大小的选择框在所述各图像上移动,对选择框内的像素进行判断,当所述选择框内的像素中同一类别的像素所占比例大于或等于预设值时,将所述选择框内的图像作为一个子图像输出,并对所述子图像标注为所述类别。

8.根据权利要求7所述的方法,其特征在于,还包括:当所述选择框内的像素中同一类别的像素所占比例小于预设值时,丢弃所述选择框。

9.根据权利要求7所述的方法,其特征在于,通过卷积神经网络得到所述子图像对应的特征,包括:

通过卷积神经网络分别对所述未标注图像和所述已标注图像进行特征提取,获取对应所述未标注图像和所述已标注图像的特征图;

基于所述子图像对应的选择框的位置和大小,从对应的特征图中获得对应所述选择框内的特征。

10.根据权利要求8所述的方法,其特征在于,通过卷积神经网络得到所述子图像对应的特征,包括:

通过卷积神经网络分别对所述未标注图像和所述已标注图像进行特征提取,获取对应所述未标注图像和所述已标注图像的特征图;

基于所述子图像对应的选择框的位置和大小,从对应的特征图中获得对应所述选择框内的特征。

11.根据权利要求1-6任一所述的方法,其特征在于,在通过语义分割模型对未标注图像进行图像语义分割之前,还包括:

使用随机梯度下降法对所述语义分割模型进行训练,直至满足预设收敛条件。

12.根据权利要求1-6任一所述的方法,其特征在于,在通过卷积神经网络,对所述未标注图像和所述已标注图像分别进行特征提取之前,还包括:

使用随机梯度下降法对所述卷积神经网络进行训练,直至满足预设收敛条件。

13.一种语义分割模型的训练装置,其特征在于,包括:

分割单元,用于通过语义分割模型,对各未标注图像进行图像语义分割,得到初步语义分割结果,作为所述未标注图像的类别;

子图像提取单元,用于通过卷积神经网络,基于各所述未标注图像的类别,及各已标注图像的类别,得到各图像分别对应的子图像及子图像对应的特征,所述各图像包括各所述未标注图像及各所述已标注图像,所述各子图像携带有对应图像的类别所述已标注图像;

所述已标注图像训练单元,用于基于各子图像的类别,及各子图像之间的特征距离,训练语义分割模型。

14.根据权利要求13所述的装置,其特征在于,所述训练单元,包括:

面片图建立模块,用于根据子图像之间的类别关系建立面片图,所述面片图包括节点和边,所述节点包括所述子图像,所述边包括任意两个所述子图像之间的特征距离;

模型训练模块,用于对所述语义分割模型进行训练,使得所述面片图中,类别相同的两个子图像之间的特征距离小于第一预设值,类别不相同的两个子图像之间的特征距离大于第二预设值。

15.根据权利要求14所述的装置,其特征在于,所述面片图建立模块,包括:

基准选择模块,用于选择至少一个子图像作为基准节点;

连接关系建立模块,用于分别针对各基准节点:将与所述基准节点相同类别的子图像作为正相关节点,将与所述基准节点不同类别的子图像作为负相关节点,分别在所述基准节点与各所述正相关节点之间建立正相关连接,分别在所述基准节点与各所述负相关节点建立负相关连接;

连接图建立模块,用于由各所述基准节点及其所述正相关节点、所述负相关节点、所述正相关连接以及所述负相关连接形成稀疏连接的所述图。

16. 根据权利要求14所述的装置,其特征在于,所述模型训练模块,包括:

网络训练模块,用于通过梯度反向传播算法,对所述语义分割模型进行训练,以使所述卷积神经网络的误差达到最小化,所述误差为基于所述卷积神经网络获得的对应子图像的特征的三重损失。

17. 根据权利要求16所述的装置,其特征在于,所述网络训练模块,具体用于:

根据所述建立的面片图中的子图像之间的特征距离通过损失函数计算得到最大误差;
将最大误差通过梯度反向传播,计算所述卷积神经网络中每一层的误差;

根据所述每一层的误差计算出每一层参数的梯度,根据所述梯度修正所述卷积神经网络中对应层的参数;

根据优化参数后的卷积神经网络输出的子图像之间的距离计算得到误差,将所述误差作为最大误差;

迭代执行将最大误差通过梯度反向传播,计算所述卷积神经网络中每一层的误差;根据所述每一层的误差计算出每一层参数的梯度,根据所述梯度修正所述卷积神经网络中对应层的参数,直到所述最大误差小于或等于预设值。

18. 根据权利要求16所述的装置,其特征在于,所述模型训练模块,还包括:

分割模型训练模块,用于基于所述卷积神经网络的训练结果获得所述卷积神经网络的参数;基于获得的所述卷积神经网络的参数初始化所述语义分割模型中的参数。

19. 根据权利要求13-18任一所述的装置,其特征在于,所述子图像提取单元,包括:图像选择模块,用于响应于预设大小的选择框在所述图像上移动,对选择框内的像素进行判断,当所述选择框内的像素中同一类别的像素所占比例大于或等于预设值时,将所述选择框内的图像作为一个子图像输出,并对所述子图像标注为所述类别。

20. 根据权利要求19所述的装置,其特征在于,所述子图像提取单元,还包括:

框丢弃模块,用于当所述选择框内的像素中同一类别的像素所占比例小于预设值时,丢弃所述选择框。

21. 根据权利要求19所述的装置,其特征在于,所述子图像提取单元,还包括:

特征提取模块,用于通过卷积神经网络分别对所述未标注图像和所述已标注图像进行特征提取,获取对应所述未标注图像和所述已标注图像的特征图;

子图像特征模块,用于基于所述子图像对应的选择框的位置和大小,从对应所述已标注图像的特征图中获得对应所述选择框内的特征。

22. 根据权利要求20所述的装置,其特征在于,所述子图像提取单元,还包括:

特征提取模块,用于通过卷积神经网络分别对所述未标注图像和所述已标注图像进行特征提取,获取对应所述未标注图像和所述已标注图像的特征图;

子图像特征模块,用于基于所述子图像对应的选择框的位置和大小,从对应所述已标注图像的特征图中获得对应所述选择框内的特征。

23.根据权利要求13-18任一所述的装置,其特征在于,还包括:模型微调单元,用于使用随机梯度下降法对所述语义分割模型进行训练,直至满足预设收敛条件。

24.根据权利要求13-18任一所述的装置,其特征在于,还包括:网络微调单元,用于使用随机梯度下降法对所述卷积神经网络进行训练,直至满足预设收敛条件。

25.一种电子设备,其特征在于,包括处理器,所述处理器包括权利要求13至24任意一项所述的语义分割模型的训练装置。

26.一种电子设备,其特征在于,包括:存储器,用于存储可执行指令;
以及处理器,用于与所述存储器通信以执行所述可执行指令从而完成权利要求1至12任意一项所述语义分割模型的训练方法的操作。

27.一种计算机存储介质,用于存储计算机可读取的指令,其特征在于,所述指令被执行时执行权利要求1至12任意一项所述语义分割模型的训练方法的操作。

语义分割模型的训练方法和装置、电子设备、存储介质

技术领域

[0001] 本发明涉及计算机视觉技术,尤其是一种语义分割模型的训练方法和装置、电子设备、存储介质。

背景技术

[0002] 图像语义分割通过对输入图像的每一个像素在输出上都分配对应的判断标注,标明这个像素最可能是属于一个什么物体或类别。是计算机视觉领域的一个重要任务,其应用包括机器场景理解、视频分析等。

[0003] 为了使图像语义分割的结果更准确,需要对图像语义分割模型进行训练。

发明内容

[0004] 本发明实施例所要解决的一个技术问题是:提供一种语义分割模型的训练技术。

[0005] 本发明实施例提供的一种语义分割模型的训练方法,包括:

[0006] 通过语义分割模型,对各未标注图像进行图像语义分割,得到初步语义分割结果,作为所述未标注图像的类别;

[0007] 通过卷积神经网络,基于各所述未标注图像的类别,及各已标注图像的类别,得到各图像分别对应的子图像及子图像对应的特征,所述各图像包括各所述未标注图像及各所述已标注图像,所述各子图像携带有对应图像的类别;

[0008] 基于各子图像的类别,及各子图像之间的特征距离,训练语义分割模型。

[0009] 根据本发明实施例的一个方面,提供的一种语义分割模型的训练装置,其特征在于,包括:

[0010] 分割单元,用于通过语义分割模型,对各未标注图像进行图像语义分割,得到初步语义分割结果,作为所述未标注图像的类别;

[0011] 子图像提取单元,用于通过卷积神经网络,基于各所述未标注图像的类别,及各已标注图像的类别,得到各图像分别对应的子图像及子图像对应的特征,所述各图像包括各所述未标注图像及各所述已标注图像,所述各子图像携带有对应图像的类别;

[0012] 训练单元,用于基于各子图像的类别,及各子图像之间的特征距离,训练语义分割模型。

[0013] 根据本发明实施例的一个方面,提供的一种电子设备,包括处理器,所述处理器包括如上所述的语义分割模型的训练装置。

[0014] 根据本发明实施例的一个方面,提供的一种电子设备,包括:存储器,用于存储可执行指令;

[0015] 以及处理器,用于与所述存储器通信以执行所述可执行指令从而完成如上所述语义分割模型的训练方法的操作。

[0016] 根据本发明实施例的一个方面,提供的一种计算机存储介质,用于存储计算机可读取的指令,其特征在于,所述指令被执行时执行如上所述语义分割模型的训练方法的操

作。

[0017] 基于本发明上述实施例提供的一种语义分割模型的训练方法和装置、电子设备和存储介质,通过语义分割模型对未标注图像进行图像语义分割,使未标注图像能够得到一个带噪声的类别,基于未标注图像的类别,及已标注图像的类别,得到各图像分别对应的子图像,将标注图像和未标注图像都应用到训练中,实现了自监督训练;通过卷积神经网络,实现对子图像进行特征提取,基于各子图像的类别,及各子图像之间的特征距离,实现对语义分割模型的训练,通过训练得到具有较强的语义区分能力的自监督学习的语义分割模型,在语义分割上能获得较高的准确率。

[0018] 下面通过附图和实施例,对本发明的技术方案做进一步的详细描述。

附图说明

[0019] 构成说明书的一部分的附图描述了本发明的实施例,并且连同描述一起用于解释本发明的原理。

[0020] 参照附图,根据下面的详细描述,可以更加清楚地理解本发明,其中:

[0021] 图1为本发明语义分割模型的训练方法一个实施例的流程图。

[0022] 图2为本发明语义分割模型的训练方法建立面片图的一个具体示例示意图。

[0023] 图3为本发明语义分割模型的训练方法建立面片图的另一个具体示例示意图。

[0024] 图4为本发明语义分割模型的训练装置一个实施例的结构示意图。

[0025] 图5为本发明电子设备一个实施例的结构示意图。

具体实施方式

[0026] 现在将参照附图来详细描述本发明的各种示例性实施例。应注意:除非另外具体说明,否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本发明的范围。

[0027] 同时,应当明白,为了便于描述,附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

[0028] 以下对至少一个示例性实施例的描述实际上仅仅是说明性的,决不作为对本发明及其应用或使用的任何限制。

[0029] 对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论,但在适当情况下,所述技术、方法和设备应当被视为说明书的一部分。

[0030] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步讨论。

[0031] 本发明实施例可以应用于计算机系统/服务器,其可与众多其它通用或专用计算系统环境或配置一起操作。适于与计算机系统/服务器一起使用的众所周知的计算系统、环境和/或配置的例子包括但不限于:个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境,等等。

[0032] 计算机系统/服务器可以在由计算机系统执行的计算机系统可执行指令(诸如程序模块)的一般语境下描述。通常,程序模块可以包括例程、程序、目标程序、组件、逻辑、数

据结构等等,它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施,分布式云计算环境中,任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中,程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

[0033] 为了让图像语义分割的模型收敛到较好的结果,采用深度学习的语义分割一般需要一个预训练模型,例如ImageNet的分类任务作为预训练模型。

[0034] 但在实现本发明的过程中,发明人发现,这种基于分类任务的预训练模型需要大量的类别标注和较长的训练时间。为了避免使用分类任务作为预训练,自监督学习被提出。

[0035] 自监督学习是用不带标注的图像本身进行训练,得到图像描述子,图像描述子是可以用来描述图像语义信息的高纬向量;然后用这些图像描述子进行语义分割的训练。通过上述原理,可以利用大量的未标注图片帮助语义分割。然而,目前现有技术中基于自监督学习的语义分割方法表现结果都不够理想,与基于分类任务作为预训练的语义分割方法差距较大。

[0036] 图1为本发明语义分割模型的训练方法一个实施例的流程图。如图1所示,该实施例方法包括:

[0037] 步骤101,通过语义分割模型,对各未标注图像进行图像语义分割,得到初步语义分割结果,作为所述未标注图像的类别。

[0038] 其中,未标注图像是指该图像中所有像素的类别(例如:语义类别)是不确定的,在本实施例中示例性地,可以通过一个已知的语义分割模型对未标注图像进行图像语义分割,获得具有噪声的语义分割结果。

[0039] 步骤102,通过卷积神经网络,基于各未标注图像的类别,及各已标注图像的类别,得到各图像分别对应的子图像及子图像对应的特征。

[0040] 其中,各图像包括各未标注图像及各已标注图像,各子图像携带有对应图像的类别。示例性地,通过可设置大小的选择框在图像中移动,再根据图像中各像素的类别判断选择框内的像素是否是同一类别的,当一个选择框中的超出设定比例的像素都属于同一类别,就可以将这个选择框作为一个子图像输出。

[0041] 步骤103,基于各子图像的类别,及各子图像之间的特征距离,训练语义分割模型。

[0042] 基于本发明上述实施例提供的一种语义分割模型的训练方法,通过语义分割模型对未标注图像进行图像语义分割,使未标注图像能够得到一个带噪声的类别,基于未标注图像的类别,及已标注图像的类别,得到各图像分别对应的子图像,将标注图像和未标注图像都应用到训练中,实现了自监督训练;通过卷积神经网络,实现对子图像进行特征提取,基于各子图像的类别,及各子图像之间的特征距离,实现对语义分割模型的训练,通过训练得到具有较强的语义区分能力的自监督学习的语义分割模型,在语义分割上能获得较高的准确率。

[0043] 本发明语义分割模型的训练方法的另一个实施例中,在上述实施例的基础上,操作103包括:

[0044] 根据子图像之间的类别关系建立patch graph面片图,该面片图包括节点和边,节点包括子图像,边包括任意两个子图像之间的特征距离;

[0045] 对语义分割模型进行训练,使得该面片图中,类别相同的两个子图像之间的特征

距离小于第一预设值,类别不相同的两个子图像之间的特征距离大于第二预设值。

[0046] 该实施例中,图2为本发明语义分割模型的训练方法建立面片图的一个具体示例示意图。如图2所示,为了建立patch graph面片图22,首先需要确定节点221,本实施例中将子图像作为节点221,在任意一个已知类别图像21中通过选择框211选择至少一个子图像,将具有连接关系的子图像之间的特征距离作为边222(图2中中层特征中选取的选择框中的特征为子图像的特征),其中子图像之间的连接关系是根据子图像对应的类别决定的;子图像的特征是通过对应的选择框在卷积神经网络的输出层输出的特征图中选出的特征;可选地,该输出层为卷积神经网络中的中层或深层中的任意一层;选择卷积神经网络中层或深层中的一层作为输出层,其中图像浅层特征一般表征图像中物体的一些edge边缘、角点等信息,图像中层特征一般表征物体的一些部件信息(比如:车辆的轮子、人脸的鼻子等),图像深层特征一般表征图像整体的类别信息(比如:人、车、马等);为了通过子图像建立图并对参数进行优化,选择中层或深层中的一层作为已标注图像和未标注图像的输出层,并且,经过多次实践证明,中层特征的优化效果优于深层特征;其中,第一预设值和第二预设值是预先设定的,通常第二预设值大于第一预设值,通过第一预设值和第二预设值使类别相同的两个子图像之间的特征距离越小,类别不相同的两个子图像之间的特征距离越大。

[0047] 图3为本发明语义分割模型的训练方法建立面片图的另一个具体示例示意图。该实施例方法包括:通过卷积神经网络,基于各未标注图像的类别(该未标注图像的类别可基于已知语义分割模型获得),及各已标注图像的类别,得到各图像分别对应的子图像及子图像对应的特征(图3中中层特征中的对应子图像位置的特征);根据子图像之间的类别关系建立patch graph面片图,该面片图包括节点和边(图3中面片图中圆代表节点,连接两个圆之间的线条代表边),节点包括子图像,边包括任意两个子图像之间的特征距离。

[0048] 在本发明语义分割模型的训练方法上述各实施例的一个具体示例中,根据子图像之间的类别关系建立面片图,包括:

[0049] 选择至少一个子图像作为基准节点,分别针对各基准节点:

[0050] 将与基准节点相同类别的子图像作为正相关节点,将与基准节点不同类别的子图像作为负相关节点,分别在基准节点与各正相关节点之间建立正相关连接,分别在基准节点与各负相关节点建立负相关连接;

[0051] 由各基准节点及其正相关节点、负相关节点、正相关连接以及负相关连接形成稀疏连接的图。

[0052] 该实施例中,建立面片图的过程就是从所有子图像中随机选择多个子图像,分别以随机选择的子图像作为anchor锚,基于语义类别,随机选择与anchor同类别的一个子图像作为positive积极图,随机选择与anchor不同语义类别的一个子图像作为negative消极图,此时就基于一个子图像建立了两个连接:anchor-positive和anchor-negative;基于这些连接,就建立了一个稀疏连接的面片图。

[0053] 在本发明语义分割模型的训练方法上述各实施例的一个具体示例中,对语义分割模型进行训练,包括:

[0054] 通过梯度反向传播算法,对所述语义分割模型进行训练,以使卷积神经网络的误差达到最小化,误差为基于卷积神经网络获得的对应子图像的特征的三重损失。

[0055] 本实施例中通过梯度反向传播算法缩小卷积神经网络中的误差,使卷积神经网络

从第一层到输出层的所有层的参数得到优化,梯度反向传播算法(BP算法)是在有导师指导下,适合于多层神经元网络的一种学习算法,它建立在梯度下降法的基础上。BP网络的输入输出关系实质上是一种映射关系:一个n输入m输出的BP神经网络所完成的功能是从n维欧氏空间向m维欧氏空间中一有限域的连续映射,这一映射具有高度非线性。BP算法的学习过程由正向传播过程和反向传播过程组成。在正向传播过程中,输入信息通过输入层经隐含层,逐层处理并传向输出层。如果在输出层得不到期望的输出值,则取输出与期望的误差的平方和作为目标函数,转入反向传播,逐层求出目标函数对各神经元权值的偏导数,构成目标函数对权值向量的梯量,作为修改权值的依据,网络的学习在权值修改过程中完成。误差达到所期望值时,网络学习结束。

[0056] 面片图中的边是由输出层输出的子图像之间的特征距离获得,其中输出层是从中层或深层中选择的一层,因此,优化的不是卷积神经网络的所有层的参数,而是从第一层到该输出层的参数,因此,在误差计算过程中,也同样是计算从输出层到第一层中每一层的误差。

[0057] 在本发明语义分割模型的训练方法上述各实施例的一个具体示例中,通过梯度反向传播算法,对语义分割模型进行训练,包括:

[0058] 根据建立的面片图中的子图像的特征之间的距离通过损失函数计算得到最大误差;

[0059] 将最大误差通过梯度反向传播,计算卷积神经网络中每一层的误差;

[0060] 根据每一层的误差计算出每一层参数的梯度,根据梯度修正所述卷积神经网络中对应层的参数;

[0061] 根据优化参数后的卷积神经网络输出的子图像之间的距离计算得到误差,将误差作为最大误差;

[0062] 迭代执行将最大误差通过梯度反向传播,计算卷积神经网络中每一层的误差;根据每一层的误差计算出每一层参数的梯度,根据梯度修正所述卷积神经网络中对应层的参数,直到最大误差小于或等于预设值。

[0063] 在本实施例中,首先定义一个损失函数,卷积神经网络通过最小化这个损失函数来优化网络参数,该损失函数公式如公式(1)所示:

[0064]

$$\min \sum_i^N \sum_{c \in C} -\log q(c|z^{(i)}) \simeq \sum_i^N \max(0, \|z_{anc}^{(i)} - z_{pos}^{(i)}\| - \|z_{anc}^{(i)} - z_{neg}^{(i)}\| + m) \quad \text{公式(1)}$$

[0065] 其中, $\|z_{anc}^{(i)} - z_{pos}^{(i)}\|$ 表示基于子图像建立的面片图中anchor与positive之间的距离, $\|z_{anc}^{(i)} - z_{neg}^{(i)}\|$ 表示基于子图像建立的面片图中anchor与negative之间的距离, m表示一个常数,该公式是基于现有技术中triplet loss三重损失的公式获得的,通过计算出的误差,结合梯度反向传播算法就可以实现对卷积神经网络中每层的参数的优化。

[0066] 在本发明语义分割模型的训练方法上述各实施例的一个具体示例中,对语义分割模型进行训练的过程可以包括:

[0067] 基于卷积神经网络的训练结果获得卷积神经网络的参数;

[0068] 基于获得的卷积神经网络的参数初始化语义分割模型中的参数。

[0069] 在本实施例中,由于语义分割模型也属于卷积神经网络,经过训练得到的卷积神经网络的参数具有较强的语义类别区分性,在语义分割上能获得较高的准确率,将该卷积神经网络的参数替换掉原始语义分割模型中的参数,就获得训练完成的语义分割模型。

[0070] 本发明语义分割模型的训练方法的另一个实施例中,在上述各实施例的基础上,操作102可以包括:

[0071] 响应于预设大小的选择框在各图像上移动,对选择框内的像素进行判断,当选择框内的像素中同一语义类别的像素所占比例大于或等于预设值时,将选择框内的图像作为一个子图像输出,并对子图像标注为类别。

[0072] 在本实施例中,通过一个大小可变化的选择框对各图像进行分割,其中,各图像包括未标注图像和已标注图像,当选择框内的像素属于一个类别(例如:语义类别)的像素所占比例大于或等于预设值时,可以将该选择框分为该类别,并将该选择框内的所有像素输出作为一个子图像,对于选择框的大小是可调的,当通过一个大小的选择框在图像中没有获得子图像时,可以通过调整选择框的大小,重新进行分割,直到得到一定数量的子图像。

[0073] 在本发明语义分割模型的训练方法上述各实施例的一个具体示例中,操作102还可以包括:当选择框内的像素中同一类别的像素所占比例小于预设值时,丢弃该选择框。

[0074] 在本示例中,对于已设定大小的选择框需要在一个图像中完成逐像素的移动,以避免漏掉可选子图像,当一个选择框中存在多个类别,但每个语义类别对应的像素比例都小于预设值,那么该选择框是无法确定类别的,此时需要将选择框移动到下一个位置,在下一个位置继续判断;当通过一个设定大小的选择框在一个图像中未获得任一子图像时,需要调整选择框的大小,重新对该图像进行选择。

[0075] 在本发明语义分割模型的训练方法上述各实施例的一个具体示例中,通过卷积神经网络得到子图像对应的特征,包括:

[0076] 通过卷积神经网络分别对未标注图像和标注图像的进行特征提取,获得对应未标注图像和已标注图像的特征图;

[0077] 基于子图像对应的选择框的位置和大小,从对应的特征图中获得对应选择框内的特征。

[0078] 在本实施例中,通过获得子图像的选择框的位置和大小,在对应的卷积神经网络的输出层的特征图中通过同样位置和大小的选择框选择出对应子图像的特征,进而通过子图像的特征获得任意两个子图像之间的特征距离。

[0079] 在本发明语义分割模型的训练方法上述各实施例的一个具体示例中,还可以包括,在操作102之前,基于语义分割模型的参数初始化卷积神经网络的参数。

[0080] 示例性地,为了得到更准确的特征,采用语义分割模型的参数对卷积神经网络的参数进行初始化。

[0081] 在本发明语义分割模型的训练方法的还有一个实施例中,在上述各实施例的基础上,在通过语义分割模型对未标注图像进行图像语义分割之前,还包括:

[0082] 使用随机梯度下降法对语义分割模型进行训练,直至满足预设收敛条件。

[0083] 本实施例实现的是分别对语义分割模型的微调,具体示例中,微调过程可包括:1. 使用VGG-16网络结构的语义分割模型。2. 设置语义分割模型的初始学习率为0.01,每30000

轮迭代下降10倍。3. 使用随机梯度下降算法微调并优化语义分割任务,此过程利用8个GPU分布计算。4. 随机梯度下降算法:随机选择一批数据(本案例为16张图片),输入到网络中,前向传播得到结果,计算其与标注结果的误差,利用反向传播得到每一层的误差。根据每一层的误差算出每一层参数的梯度,根据梯度修正参数值;在不断修正的过程中让模型收敛。5. 迭代到第60000轮左右模型收敛。6. 利用这个语义分割模型在现有公开数据集上进行测试。

[0084] 在本发明语义分割模型的训练方法的再一个实施例中,在上述各实施例的基础上,在利用卷积神经网络,对未标注图像和已标注图像分别进行特征提取之前,还包括:

[0085] 使用随机梯度下降法对卷积神经网络进行训练,直至满足预设收敛条件。

[0086] 本实施例实现的是分别对卷积神经网络的微调,具体实施时,微调过程可包括:1. 使用VGG-16网络结构的卷积神经网络。2. 设置卷积神经网络的初始学习率为0.01,每30000轮迭代下降10倍。3. 使用随机梯度下降算法微调并优化语义分割任务,此过程利用8个GPU分布计算。4. 随机梯度下降算法:随机选择一批数据(本案例为16张图片),输入到网络中,前向传播得到结果,计算其与标注结果的误差,利用反向传播得到每一层的误差。根据每一层的误差算出每一层参数的梯度,根据梯度修正参数值;在不断修正的过程中让模型收敛。5. 迭代到第60000轮左右模型收敛。6. 利用这个卷积神经网络在现有公开数据集上进行测试。

[0087] 本领域普通技术人员可以理解:实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成,前述的程序可以存储于一计算机可读取存储介质中,该程序在执行时,执行包括上述方法实施例的步骤;而前述的存储介质包括:ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0088] 图4为本发明语义分割模型的训练装置一个实施例的结构示意图。该实施例的装置可用于实现本发明上述各方法实施例。如图4所示,该实施例的装置包括:

[0089] 分割单元41,用于通过语义分割模型,对各未标注图像进行图像语义分割,得到初步语义分割结果,作为未标注图像的类别。

[0090] 子图像提取单元42,用于通过卷积神经网络,基于各未标注图像的类别,及各已标注图像的类别,得到各图像分别对应的子图像及子图像对应的特征。

[0091] 其中,各图像包括各未标注图像及各已标注图像,各子图像携带有对应图像的类别。

[0092] 训练单元43,用于基于各子图像的类别,及各子图像之间的特征距离,训练语义分割模型。

[0093] 基于本发明上述实施例提供的一种语义分割模型的训练装置,通过语义分割模型对未标注图像进行图像语义分割,使未标注图像能够得到一个带噪声的类别,基于未标注图像的类别,及已标注图像的类别,得到各图像分别对应的子图像,将标注图像和未标注图像都应用到训练中,实现了自监督训练;通过卷积神经网络,实现对子图像进行特征提取,基于各子图像的类别,及各子图像之间的特征距离,实现对语义分割模型的训练,通过训练得到具有较强的语义区分能力的自监督学习的语义分割模型,在语义分割上能获得较高的准确率。

[0094] 本发明语义分割模型的训练装置的另一个实施例中,在上述实施例的基础上,训

练单元43,包括:

[0095] 面片图建立模块,用于根据子图像之间的类别关系建立面片图,该面片图包括节点和边,节点包括子图像,边包括任意两个子图像之间的特征距离;

[0096] 模型训练模块,用于对语义分割模型进行训练,使得面片图中,类别相同的两个子图像之间的特征距离小于第一预设值,类别不相同的两个子图像之间的特征距离大于第二预设值。

[0097] 该实施例中,为了建立patch graph面片图,首先需要确定节点,本实施例中将子图像作为节点,将具有连接关系的子图像之间的特征距离作为边,其中子图像之间的连接关系是根据子图像对应的类别决定的;子图像的特征是通过对的选择框在卷积神经网络的输出层输出的特征图中选出的特征;可选地,该输出层为卷积神经网络中的中层或深层中的任意一层;选择卷积神经网络中层或深层中的一层作为输出层,其中图像浅层特征一般表征图像中物体的一些edge边缘、角点等信息,图像中层特征一般表征物体的一些部件信息(比如:车辆的轮子、人脸的鼻子等),图像深层特征一般表征图像整体的类别信息(比如:人、车、马等);为了通过子图像建立图并对参数进行优化,选择中层或深层中的一层作为已标注图像和未标注图像的输出层,并且,经过多次实践证明,中层特征的优化效果优于深层特征;其中,第一预设值和第二预设值是预先设定的,通常第二预设值大于第一预设值,通过第一预设值和第二预设值使类别相同的两个子图像之间的特征距离越小,类别不相同的两个子图像之间的特征距离越大。

[0098] 在本发明语义分割模型的训练装置上述各实施例的一个具体示例中,面片图建立模块,包括:

[0099] 基准选择模块,用于选择至少一个子图像作为基准节点;

[0100] 连接关系建立模块,用于分别针对各基准节点:将与基准节点相同类别的子图像作为正相关节点,将与基准节点不同类别的子图像作为负相关节点,分别在基准节点与各正相关节点之间建立正相关连接,分别在基准节点与各所述负相关节点建立负相关连接;

[0101] 连接图建立模块,用于由各基准节点及其正相关节点、负相关节点、正相关连接以及负相关连接形成稀疏连接的所述图。

[0102] 在本发明语义分割模型的训练装置上述各实施例的一个具体示例中,模型训练模块包括:

[0103] 网络训练模块,用于通过梯度反向传播算法,对语义分割模型进行训练,以使卷积神经网络的误差达到最小化,误差为基于卷积神经网络获得的对应子图像的特征的三重损失。

[0104] 在本发明语义分割模型的训练装置上述各实施例的一个具体示例中,网络训练模块,具体用于:

[0105] 根据建立的面片图中的子图像之间的特征距离通过损失函数计算得到最大误差;

[0106] 将最大误差通过梯度反向传播,计算卷积神经网络中每一层的误差;

[0107] 根据每一层的误差计算出每一层参数的梯度,根据梯度修正卷积神经网络中对应层的参数;

[0108] 根据优化参数后的卷积神经网络输出的子图像之间的距离计算得到误差,将误差作为最大误差;

[0109] 迭代执行将最大误差通过梯度反向传播,计算卷积神经网络中每一层的误差;根据每一层的误差计算出每一层参数的梯度,根据梯度修正卷积神经网络中对应层的参数,直到最大误差小于或等于预设值。

[0110] 在本发明语义分割模型的训练装置上述各实施例的一个具体示例中,模型训练模块,还包括:

[0111] 分割模型训练模块,用于基于卷积神经网络的训练结果获得卷积神经网络的参数;基于获得的卷积神经网络的参数初始化语义分割模型中的参数。

[0112] 本发明语义分割模型的训练装置的另一个实施例中,在上述各实施例的基础上,子图像提取单元,包括:

[0113] 图像选择模块,用于响应于预设大小的选择框在图像上移动,对选择框内的像素进行判断,当选择框内的像素中同一类别的像素所占比例大于或等于预设值时,将选择框内的图像作为一个子图像输出,并对子图像标注为类别。

[0114] 在本实施例中,通过一个大小可变化的选择框对各图像进行分割,其中,各图像包括未标注图像和已标注图像,当选择框内的像素属于一个类别(例如:语义类别)的像素所占比例大于或等于预设值时,可以将该选择框分为该类别,并将该选择框内的所有像素输出作为一个子图像,对于选择框的大小是可调的,当通过一个大小的选择框在图像中没有获得子图像时,可以通过调整选择框的大小,重新进行分割,直到得到一定数量的子图像。

[0115] 在本发明语义分割模型的训练装置上述各实施例的一个具体示例中,子图像提取单元,还包括:

[0116] 框丢弃模块,用于当选择框内的像素中同一类别的像素所占比例小于预设值时,丢弃该选择框。

[0117] 在本发明语义分割模型的训练装置上述各实施例的一个具体示例中,子图像提取单元,还包括:

[0118] 特征提取模块,用于通过卷积神经网络分别对未标注图像和已标注图像进行特征提取,获取对应未标注图像和已标注图像的特征图;

[0119] 子图像特征模块,用于基于子图像对应的选择框的位置和大小,从对应已标注图像的特征图中获得对应选择框内的特征。

[0120] 在本发明语义分割模型的训练装置的另一个实施例中,在上述各实施例的基础上,还包括:模型微调单元,用于使用随机梯度下降法对语义分割模型进行训练,直至满足预设收敛条件。

[0121] 本实施例实现的是分别对语义分割模型的微调,具体示例中,微调过程可包括:1. 使用VGG-16网络结构的语义分割模型。2. 设置语义分割模型的初始学习率为0.01,每30000轮迭代下降10倍。3. 使用随机梯度下降算法微调并优化语义分割任务,此过程利用8个GPU分布计算。4. 随机梯度下降算法:随机选择一批数据(本案例为16张图片),输入到网络中,前向传播得到结果,计算其与标注结果的误差,利用反向传播得到每一层的误差。根据每一层的误差算出每一层参数的梯度,根据梯度修正参数值;在不断修正的过程中让模型收敛。5. 迭代到第60000轮左右模型收敛。6. 利用这个语义分割模型在现有公开数据集上进行测试。

[0122] 在本发明语义分割模型的训练装置的另一个实施例中,在上述各实施例的基础

上,还包括:网络微调单元,用于使用随机梯度下降法对卷积神经网络进行训练,直至满足预设收敛条件。

[0123] 本实施例实现的是分别对卷积神经网络的微调,具体实施时,微调过程可包括:1. 使用VGG-16网络结构的卷积神经网络。2. 设置卷积神经网络的初始学习率为0.01,每30000轮迭代下降10倍。3. 使用随机梯度下降算法微调并优化语义分割任务,此过程利用8个GPU分布计算。4. 随机梯度下降算法:随机选择一批数据(本案例为16张图片),输入到网络中,前向传播得到结果,计算其与标注结果的误差,利用反向传播得到每一层的误差。根据每一层的误差算出每一层参数的梯度,根据梯度修正参数值;在不断修正的过程中让模型收敛。5. 迭代到第60000轮左右模型收敛。6. 利用这个卷积神经网络在现有公开数据集上进行测试。

[0124] 根据本发明实施例的一个方面,提供的一种电子设备,包括处理器,处理器包括本发明语义分割模型的训练装置各实施例中的任意一项。

[0125] 根据本发明实施例的一个方面,提供的一种电子设备,包括:存储器,用于存储可执行指令;

[0126] 以及处理器,用于与存储器通信以执行可执行指令从而完成本发明语义分割模型的训练方法各实施例中的任意一项的操作。

[0127] 根据本发明实施例的一个方面,提供的一种计算机存储介质,用于存储计算机可读取的指令,其特征在于,所述指令被执行时执行本发明语义分割模型的训练方法各实施例中的任意一项的操作。

[0128] 本发明实施例还提供了一种电子设备,例如可以是移动终端、个人计算机(PC)、平板电脑、服务器等。下面参考图5,其示出了适于用来实现本申请实施例的终端设备或服务器的电子设备500的结构示意图:如图5所示,计算机系统500包括一个或多个处理器、通信部等,所述一个或多个处理器例如:一个或多个中央处理单元(CPU)501,和/或一个或多个图像处理器(GPU)513等,处理器可以根据存储在只读存储器(ROM)502中的可执行指令或者从存储部分508加载到随机访问存储器(RAM)503中的可执行指令而执行各种适当的动作和处理。通信部512可包括但不限于网卡,所述网卡可包括但不限于IB(Infiniband)网卡,

[0129] 处理器可与只读存储器502和/或随机访问存储器530中通信以执行可执行指令,通过总线504与通信部512相连、并经通信部512与其他目标设备通信,从而完成本申请实施例提供的任一项方法对应的操作,例如,通过语义分割模型,对各未标注图像进行图像语义分割,得到初步语义分割结果,作为所述未标注图像的类别;通过卷积神经网络,基于各未标注图像的类别,及各已标注图像的类别,得到各图像分别对应的子图像及子图像对应的特征,各图像包括各未标注图像及各已标注图像,各子图像携带有对应图像的类别;基于各子图像的类别,及各子图像之间的特征距离,训练语义分割模型。

[0130] 此外,在RAM 503中,还可存储有装置操作所需的各种程序和数据。CPU501、ROM502以及RAM503通过总线504彼此相连。在有RAM503的情况下,ROM502为可选模块。RAM503存储可执行指令,或在运行时向ROM502中写入可执行指令,可执行指令使处理器501执行上述通信方法对应的操作。输入/输出(I/O)接口505也连接至总线504。通信部512可以集成设置,也可以设置为具有多个子模块(例如多个IB网卡),并在总线链接上。

[0131] 以下部件连接至I/O接口505:包括键盘、鼠标等的输入部分506;包括诸如阴极射

线管 (CRT)、液晶显示器 (LCD) 等以及扬声器等的输出部分 507；包括硬盘等的存储部分 508；以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 509。通信部分 509 经由诸如因特网的网络执行通信处理。驱动器 510 也根据需要连接至 I/O 接口 505。可拆卸介质 511，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器 510 上，以便于从其上读出的计算机程序根据需要被安装入存储部分 508。

[0132] 需要说明的，如图 5 所示的架构仅为一种可选实现方式，在具体实践过程中，可根据实际需要对上述图 5 的部件数量和类型进行选择、删减、增加或替换；在不同功能部件设置上，也可采用分离设置或集成设置等实现方式，例如 GPU 和 CPU 可分离设置或者可将 GPU 集成在 CPU 上，通信部可分离设置，也可集成设置在 CPU 或 GPU 上，等等。这些可替换的实施方式均落入本发明公开的保护范围。

[0133] 特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括有形地包含在机器可读介质上的计算机程序，计算机程序包含用于执行流程图所示的方法的程序代码，程序代码可包括对应执行本申请实施例提供的方法步骤对应的指令，例如，通过语义分割模型，对各未标注图像进行图像语义分割，得到初步语义分割结果，作为所述未标注图像的类别；通过卷积神经网络，基于各未标注图像的类别，及各已标注图像的类别，得到各图像分别对应的子图像及子图像对应的特征，各图像包括各未标注图像及各已标注图像，各子图像携带有对应图像的类别；基于各子图像的类别，及各子图像之间的特征距离，训练语义分割模型。在这样的实施例中，该计算机程序可以通过通信部分 509 从网络上被下载和安装，和/或从可拆卸介质 511 被安装。在该计算机程序被中央处理单元 (CPU) 501 执行时，执行本申请的方法中限定的上述功能。

[0134] 可能以许多方式来实现本发明的方法和装置、设备。例如，可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本发明的方法和装置、设备。用于方法的步骤的上述顺序仅是为了进行说明，本发明的方法的步骤不限于以上具体描述的顺序，除非以其它方式特别说明。此外，在一些实施例中，还可将本发明实施为记录在记录介质中的程序，这些程序包括用于实现根据本发明的方法的机器可读指令。因而，本发明还覆盖存储用于执行根据本发明的方法的程序的记录介质。

[0135] 本发明的描述是为了示例和描述起见而给出的，而并不是无遗漏的或者将本发明限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施例是为了更好说明本发明的原理和实际应用，并且使本领域的普通技术人员能够理解本发明从而设计适于特定用途的带有各种修改的各种实施例。

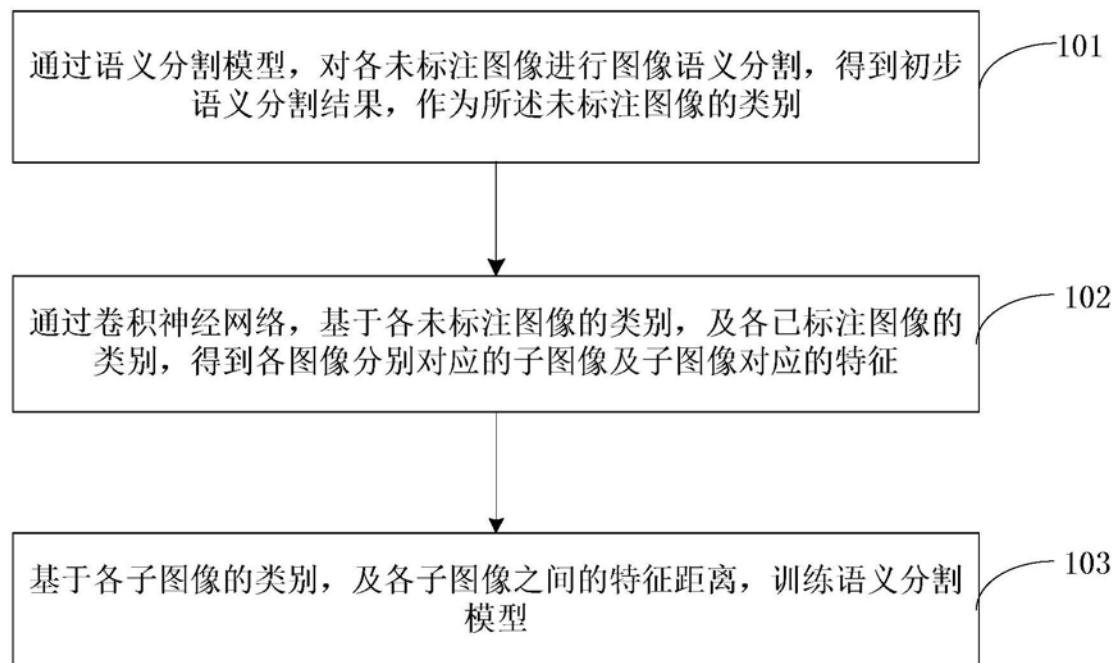


图1

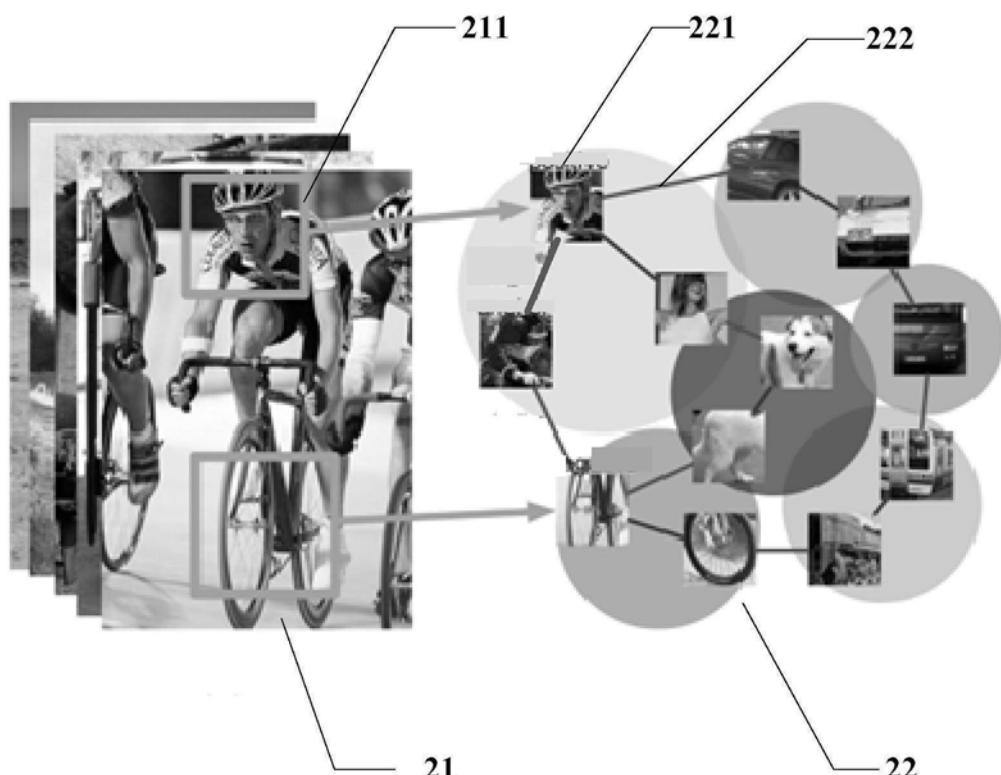


图2

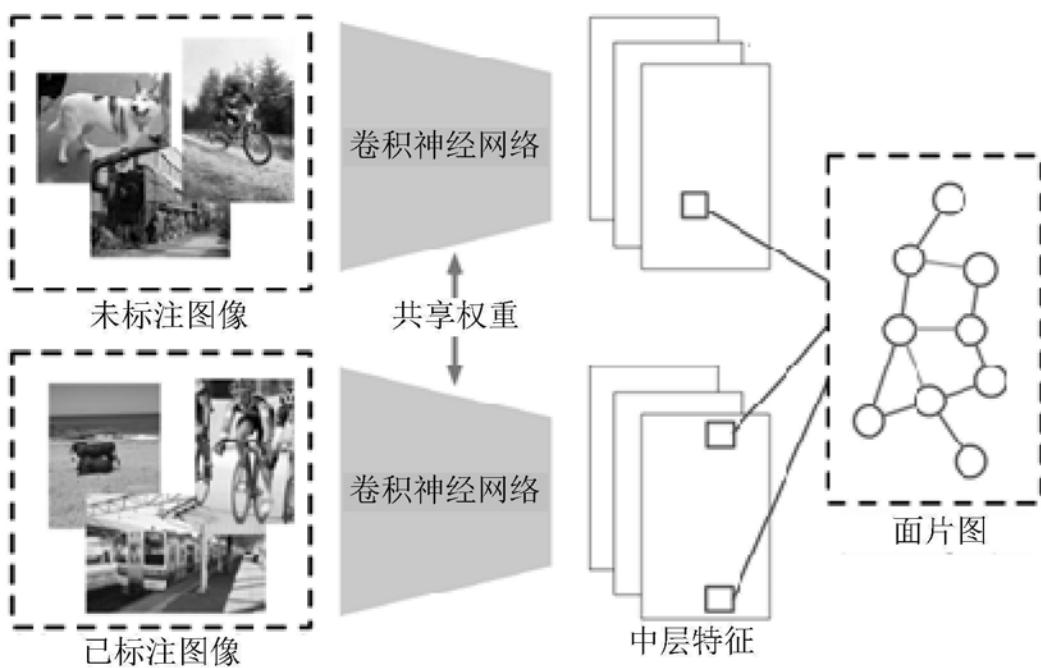


图3

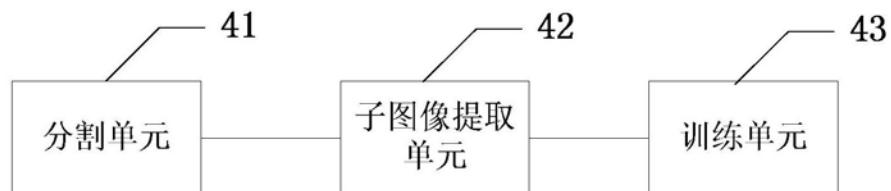


图4

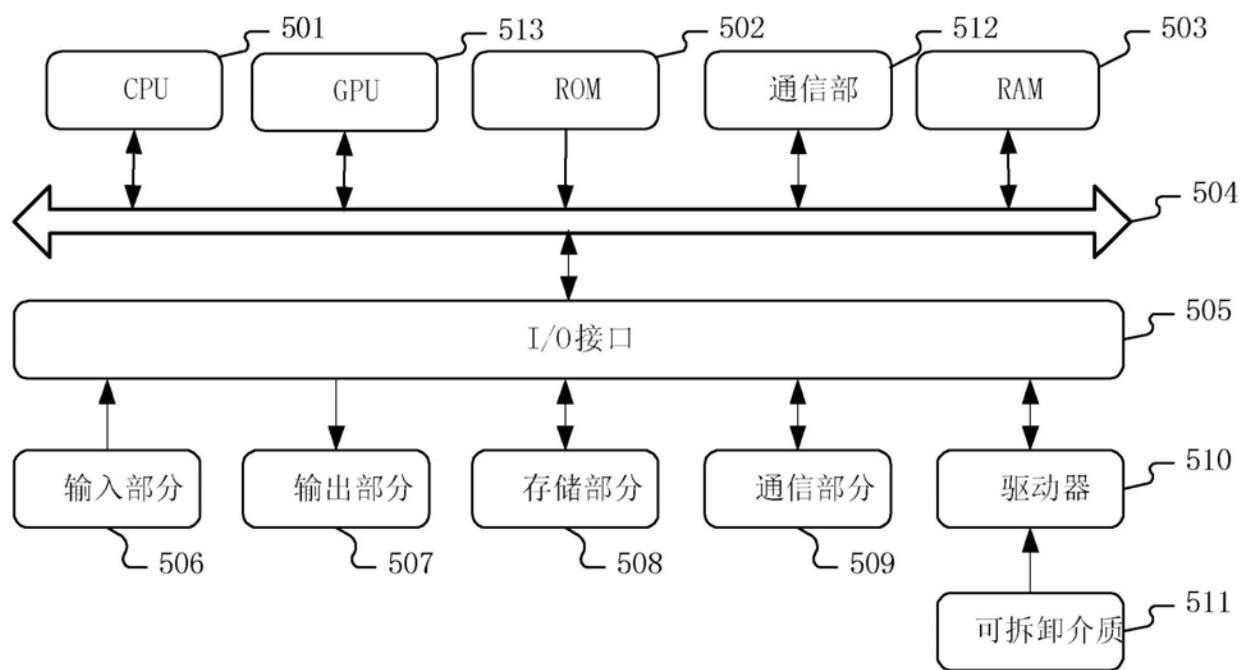
500

图5