



(12)发明专利申请

(10)申请公布号 CN 109389640 A  
(43)申请公布日 2019.02.26

(21)申请号 201811149818.4

(22)申请日 2018.09.29

(71)申请人 北京字节跳动网络技术有限公司  
地址 100041 北京市石景山区实兴大街30  
号院3号楼2层B-0035房间

(72)发明人 胡耀全

(74)专利代理机构 北京英赛嘉华知识产权代理  
有限责任公司 11204  
代理人 王达佐 马晓亚

(51)Int.Cl.  
G06T 7/70(2017.01)

权利要求书2页 说明书10页 附图5页

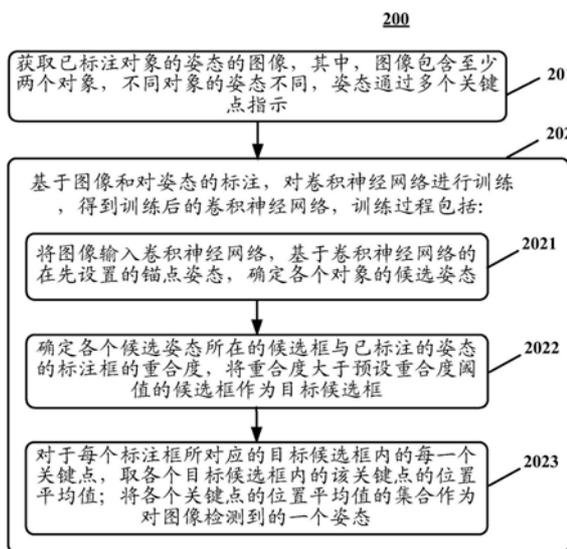
(54)发明名称

图像处理方法和装置

(57)摘要

本申请实施例公开了图像处理方法和装置。该方法的一具体实施方式包括：获取已标注对象的姿态的图像；基于该图像和对姿态的标注，对卷积神经网络进行训练，得到训练后的卷积神经网络，训练过程包括：将该图像输入卷积神经网络，基于该卷积神经网络的在先设置的锚点姿态，确定各个对象的候选姿态；将重合度大于预设重合度阈值的候选框作为目标候选框；对于每个标注框所对应的目标候选框内的每一个关键点，取各个目标候选框内的该关键点的位置平均值；将各个关键点的位置平均值的集合作为对该图像检测到的一个姿态。本实施例通过重合度对各个候选姿态进行筛选并取得关键点的平均值，以准确地分辨出图像中的各个姿态。

CN 109389640 A



1. 一种图像处理方法,包括:

获取已标注对象的姿态的图像,其中,所述图像包含至少两个对象,不同对象的姿态不同,姿态通过多个关键点指示;

基于所述图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括:

将所述图像输入卷积神经网络,基于所述卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态;

确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框;

对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值;将各个关键点的位置平均值的集合作为对所述图像检测到的一个姿态。

2. 根据权利要求1所述的方法,其中,在所述将所述图像输入卷积神经网络,基于所述卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态之前,所述方法还包括:

对目标图像中的多个预设姿态进行聚类,得到关键点集合;

将各个关键点集合确定为锚点姿态,其中,不同关键点集合所包括的关键点在所述目标图像中的位置不同。

3. 根据权利要求2所述的方法,其中,所述对目标图像中的多个预设姿态进行聚类,得到关键点集合,包括:

对各个预设姿态所对应的多维向量进行聚类,其中,预设姿态所对应的多维向量的维度数量与预设姿态的关键点数量相同;

将聚类中心的多维向量所对应的姿态的各个关键点组成关键点集合。

4. 根据权利要求1所述的方法,其中,所述对于每个标注框所对应的目标候选框的每一个关键点,取各个目标候选框内候选姿态中的该关键点的位置平均值,包括:

对于每个标注框所对应的每个目标候选框内的每一个关键点,响应于确定该关键点的位置在该标注框以外,将预设的第一预设权重作为该目标候选框内的该关键点的权重;响应于确定该关键点的位置在该标注框以内,将预设的第二预设权重作为该目标候选框内的该关键点的权重,所述第一预设权重小于第二预设权重;基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

5. 根据权利要求1述的方法,其中,所述对于每个标注框所对应的目标候选框的每一个关键点,取各个目标候选框内候选姿态中的该关键点的位置平均值,包括:

对于每个标注框所对应的每个目标候选框内的每一个关键点,确定该关键点与已标注的姿态中该关键点的距离是否小于或等于预设距离阈值;响应于确定小于或者等于,基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

6. 一种图像处理装置,包括:

获取单元,被配置成获取已标注对象的姿态的图像,其中,所述图像包含至少两个对象,不同对象的姿态不同,姿态通过多个关键点指示;

训练单元,被配置成基于所述图像和对姿态的标注,对卷积神经网络进行训练,得到训

练后的卷积神经网络,训练过程包括:

将所述图像输入卷积神经网络,基于所述卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态;确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框;对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值;将各个关键点的位置平均值的集合作为对所述图像检测到的一个姿态。

7. 根据权利要求6所述的装置,其中,所述装置还包括:

聚类单元,被配置成对目标图像中的多个预设姿态进行聚类,得到关键点集合;

确定单元,被配置成将各个关键点集合确定为锚点姿态,其中,不同关键点集合所包括的关键点在所述目标图像中的位置不同。

8. 根据权利要求7所述的装置,其中,所述聚类单元,进一步被配置成:

对各个预设姿态所对应的多维向量进行聚类,其中,预设姿态所对应的多维向量的维度数量与预设姿态的关键点数量相同;

将聚类中心的多维向量所对应的预设姿态的各个关键点组成关键点集合。

9. 根据权利要求6所述的装置,其中,所述训练单元,进一步被配置成:

对于每个标注框所对应的每个目标候选框内的每一个关键点,响应于确定该关键点的位置在该标注框以外,将预设的第一预设权重作为该目标候选框内的该关键点的权重;响应于确定该关键点的位置在该标注框以内,将预设的第二预设权重作为该目标候选框内的该关键点的权重,所述第一预设权重小于第二预设权重;基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

10. 根据权利要求6的装置,其中,所述训练单元,进一步被配置成:

对于每个标注框所对应的每个目标候选框内的每一个关键点,确定该关键点与已标注的姿态中该关键点的距离是否小于或等于预设距离阈值;响应于确定小于或者等于,基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

11. 一种电子设备,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-5中任一所述的方法。

12. 一种计算机可读存储介质,其上存储有计算机程序,其中,该程序被处理器执行时实现如权利要求1-5中任一所述的方法。

## 图像处理方法和装置

### 技术领域

[0001] 本申请实施例涉及计算机技术领域,具体涉及互联网技术领域,尤其涉及图像处理和装置。

### 背景技术

[0002] 在确认人体关键点时,有时需要确认单人的关键点,有时则需要确认多个人中每个人的关键点。在相关技术中,检测多个人中每个人的关键点时,通常难以准确地检测出结果。

### 发明内容

[0003] 本申请实施例提出了图像处理方法和装置。

[0004] 第一方面,本申请实施例提供了一种图像处理方法,包括:获取已标注对象的姿态的图像,其中,图像包含至少两个对象,不同对象姿态不同,姿态通过多个关键点指示;基于图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括:将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态;确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框;对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值;将各个关键点的位置平均值的集合作为对图像检测到的一个姿态。

[0005] 在一些实施例中,在将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态之前,方法还包括:对目标图像中的多个预设姿态进行聚类,得到关键点集合;将各个关键点集合确定为锚点姿态,其中,不同关键点集合所包括的关键点在目标图像中的位置不同。

[0006] 在一些实施例中,对目标图像中的多个预设姿态进行聚类,得到关键点集合,包括:对各个预设姿态所对应的多维向量进行聚类,其中,预设姿态所对应的多维向量的维度数量与预设姿态的关键点数量相同;将聚类中心的多维向量所对应的姿态的各个关键点组成关键点集合。

[0007] 在一些实施例中,对于每个标注框所对应的目标候选框的每一个关键点,取各个目标候选框内候选姿态中的该关键点的位置平均值,包括:对于每个标注框所对应的每个目标候选框内的每一个关键点,响应于确定该关键点的位置在该标注框以外,将预设的第一预设权重作为该目标候选框内的该关键点的权重;响应于确定该关键点的位置在该标注框以内,将预设的第二预设权重作为该目标候选框内的该关键点的权重,第一预设权重小于第二预设权重;基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

[0008] 在一些实施例中,对于每个标注框所对应的目标候选框的每一个关键点,取各个目标候选框内候选姿态中的该关键点的位置平均值,包括:对于每个标注框所对应的每个

目标候选框内的每一个关键点,确定该关键点与已标注的姿态中该关键点的距离是否小于或等于预设距离阈值;响应于确定小于或者等于,基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

[0009] 第二方面,本申请实施例提供了一种图像处理装置,包括:获取单元,被配置成获取已标注对象的姿态的图像,其中,图像包含至少两个对象,不同对象的姿态不同,姿态通过多个关键点指示;训练单元,被配置成基于图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括:将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态;确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框;对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值;将各个关键点的位置平均值的集合作为对图像检测到的一个姿态。

[0010] 在一些实施例中,装置还包括:聚类单元,被配置成对目标图像中的多个预设姿态进行聚类,得到关键点集合;确定单元,被配置成将各个关键点集合确定为锚点姿态,其中,不同关键点集合所包括的关键点在目标图像中的位置不同。

[0011] 在一些实施例中,聚类单元,进一步被配置成:对各个预设姿态所对应的多维向量进行聚类,其中,预设姿态所对应的多维向量的维度数量与预设姿态的关键点数量相同;将聚类中心的多维向量所对应的预设姿态的各个关键点组成关键点集合。

[0012] 在一些实施例中,训练单元,进一步被配置成:被配置成对于每个标注框所对应的每个目标候选框内的每一个关键点,响应于确定该关键点的位置在该标注框以外,将预设的第一预设权重作为该目标候选框内的该关键点的权重;响应于确定该关键点的位置在该标注框以内,将预设的第二预设权重作为该目标候选框内的该关键点的权重,第一预设权重小于第二预设权重;基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

[0013] 在一些实施例中,训练单元,进一步被配置成:对于每个标注框所对应的每个目标候选框内的每一个关键点,确定该关键点与已标注的姿态中该关键点的距离是否小于或等于预设距离阈值;响应于确定小于或者等于,基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

[0014] 第三方面,本申请实施例提供了一种电子设备,包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如图像处理方法中任一实施例的方法。

[0015] 第四方面,本申请实施例提供了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现如图像处理方法中任一实施例的方法。

[0016] 本申请实施例提供的图像处理方案,首先,获取已标注对象的姿态的图像,其中,图像包含至少两个对象,不同对象的姿态不同,姿态通过多个关键点指示。之后,基于图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括:将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态。然后,确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框。而后,对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值。最后,将各个关键

点的位置平均值的集合作为对图像检测到的一个姿态。本实施例能够从包括至少两个对象的图像中,通过重合度对各个候选姿态进行筛选,以选取指示对象更加准确的目标候选框。并且,取得关键点的平均值准确地分辨出图像中的各个姿态。

### 附图说明

[0017] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本申请的其它特征、目的和优点将会变得更明显:

[0018] 图1是本申请可以应用于其中的示例性系统架构图;

[0019] 图2是根据本申请的图像处理方法的一个实施例的流程图;

[0020] 图3是根据本申请的图像处理方法的一个应用场景的示意图;

[0021] 图4是根据本申请的图像处理方法的又一个实施例的流程图;

[0022] 图5是根据本申请的图像处理装置的一个实施例的结构示意图;

[0023] 图6是适于用来实现本申请实施例的电子设备的计算机系统的结构示意图。

### 具体实施方式

[0024] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释相关发明,而非对该发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与有关发明相关的部分。

[0025] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0026] 图1示出了可以应用本申请的图像处理方法或图像处理装置的实施例的示例性系统架构100。

[0027] 如图1所示,系统架构100可以包括终端设备101、102、103,网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0028] 用户可以使用终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用,例如图像处理应用、视频类应用、直播应用、即时通信工具、邮箱客户端、社交平台软件等。

[0029] 这里的终端设备101、102、103可以是硬件,也可以是软件。当终端设备101、102、103为硬件时,可以是具有显示屏的各种电子设备,包括但不限于智能手机、平板电脑、电子书阅读器、膝上型便携计算机和台式计算机等等。当终端设备101、102、103为软件时,可以安装在上述所列举的电子设备中。其可以实现成多个软件或软件模块(例如用来提供分布式服务的多个软件或软件模块),也可以实现成单个软件或软件模块。在此不做具体限定。

[0030] 服务器105可以是提供各种服务的服务器,例如对终端设备101、102、103提供支持的后台服务器。后台服务器可以对获取到的已标注对象的姿态的图像等数据进行分析等处理,并将处理结果(例如对图像检测到的一个姿态)反馈给终端设备。

[0031] 需要说明的是,本申请实施例所提供的图像处理方法可以由服务器105或者终端设备101、102、103执行,相应地,图像处理装置可以设置于服务器105或者终端设备101、102、103中。

[0032] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器的。

[0033] 继续参考图2,示出了根据本申请的图像处理方法的一个实施例的流程200。该图像处理方法,包括以下步骤:

[0034] 步骤201,获取已标注对象的姿态的图像,其中,图像包含至少两个对象,不同对象的姿态不同,姿态通过多个关键点指示。

[0035] 在本实施例中,图像处理方法的执行主体(例如图1所示的服务器或终端设备)可以获取已标注对象的姿态的图像。在图像中,对象的姿态被标注出来。这里的对象可以是人、人脸、猫、物品等等。具体地,可以通过关键点的坐标来表示姿态。举例来说,人在呈站立姿态和下蹲姿态时,鼻尖关键点的坐标与脚尖关键点的坐标的距离不同。

[0036] 步骤202,基于图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括步骤2021、步骤2022和步骤2023,如下:

[0037] 步骤2021,将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态。

[0038] 在本实施例中,上述执行主体可以将获取的图像输入卷积神经网络,从而基于卷积神经网络中的在先设置的锚点姿态,由卷积神经网络得到各个对象的候选姿态(proposal)。具体地,卷积神经网络中包括区域候选网络(Region Proposal Network, RPN)。卷积神经网络中的锚点姿态(anchor)在图像中的尺寸和位置是固定的。上述执行主体可以将上述图像输入区域候选网络,区域候选网络可以确定候选姿态与锚点姿态之间的尺寸的差值和位置的差值,并利用上述的尺寸的差值与位置的差值来表示各个候选姿态的尺寸和位置。这里的尺寸可以采用面积或宽、高或长、宽等来表示,位置可以采用坐标来表示。对于图像中的每一个对象,上述执行主体可以确定多个候选姿态。

[0039] 在训练的过程中,上述执行主体可以获得卷积神经网络输出的姿态作为对上述图像检测到的姿态,并基于预设的损失函数,确定该姿态与标注的姿态的损失值。之后利用该损失值进行训练,以得到训练后的卷积神经网络。

[0040] 步骤2022,确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框。

[0041] 在本实施例中,上述执行主体可以确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度(Intersection over Union, IOU)。之后,上述执行主体可以选取重合度大于预设重合度阈值的候选框,并将所选取的候选框作为目标候选框。具体地,姿态的框的宽度和高度可以是姿态所包括的关键点的最左坐标、最右坐标所生成的宽度(或者长度),和最上坐标、最下坐标所生成的高度(或者宽度)。重合度可以是候选框和标注框之间的交集与候选框和标注框之间的并集的比值。如果候选框与标注框的重合度较大,表明该候选框对对象进行框定的准确度较高,这样,通过该候选框能够更准确地划分对象和非对象。

[0042] 步骤2023,对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值;将各个关键点的位置平均值的集合作为对图像检测到的一个姿态。

[0043] 在本实施例中,上述执行主体可以对于每个标注框所对应的目标候选框内的每一

个关键点,取该标注框所对应的各个目标候选框内候选姿态中的该关键点的位置平均值。从而,上述执行主体可以将该标注框所对应的目标候选框的各个关键点的位置平均值的集合作为对图像检测到的一个姿态。相对应的标注框与目标候选框指示相同的对象。

[0044] 具体地,可以对各个关键点的位置设置相同的权重,以计算位置平均值。此外,对各个关键点的位置所设置的权重也可以存在不同。

[0045] 需要说明的是,虽然对目标候选框内姿态的每一个关键点都取位置平均值,但是本实施例存在部分目标候选框内的该关键点不参与取得位置平均值的可能。

[0046] 在本实施例的一些可选的实现方式中,步骤2023中的对于每个标注框所对应的目标候选框内的每一个关键点,取至少两个目标候选框内的该关键点的位置平均值,可以包括:

[0047] 对于每个标注框所对应的每个目标候选框内的每一个关键点,响应于确定该关键点的位置在该标注框以外,将预设的第一预设权重作为该目标候选框内的该关键点的权重;对于每个标注框所对应的每个目标候选框内的每一个关键点,响应于确定该关键点的位置在该标注框以内,将预设的第二预设权重作为该目标候选框内的该关键点的权重,第一预设权重小于第二预设权重;基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

[0048] 在这些可选的实现方式中,上述执行主体可以在计算位置平均值的时候,对在标注框以外的位置的坐标采用较小的权重,并对在标注框以内的位置的坐标采用较大的权重。举例来说,关键点A、关键点B和关键点C的权重分别在标注框内、标注框内和标注框外,可以分别对关键点A、关键点B和关键点C采用权重1、1和0.5来计算位置平均值。那么取得的位置平均值为 $(1 \times \text{关键点A位置} + 1 \times \text{关键点B位置} + 0.5 \times \text{关键点C位置}) / (1 + 1 + 0.5)$ 。

[0049] 这些实现方式可以差别化地获取不同目标候选框的权重。因为在标注框外的关键点往往准确度较低,这样的权重设定方式可以降低这些关键点的权重以获取更准确的关键点的位置平均值,进而准确地确定姿态。

[0050] 在本实施例的一些可选的实现方式中,步骤2023中的对于每个标注框所对应的目标候选框内的每一个关键点,取至少两个目标候选框内的该关键点的位置平均值,可以包括:

[0051] 对于每个标注框所对应的每个目标候选框内的每一个关键点,确定该关键点与已标注的姿态中该关键点的距离是否小于或等于预设距离阈值;响应于确定小于或者等于,基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

[0052] 在这些可选的实现方式中,上述执行主体可以确定每个标注框所对应的每个目标候选框内的每个关键点与该标注框内已标注的姿态中该关键点的距离是否小于或等于预设距离阈值,并由此对该标注框所对应的各个目标候选框内的该关键点进行取舍。也即在这些实现方式中存在部分目标候选框内的该关键点不参与取得位置平均值。具体地,如果该标注框所对应的某个目标候选框内的某个关键点与所标注的该关键点之间的距离较小,则确定该关键点可以参与计算位置平均值。如果该标注框所对应的某个目标候选框内存在某个关键点与所标注的该关键点之间的距离较大,则表示卷积神经网络得到的候选姿态中的该关键点准确度较差,可以确定该关键点不参与计算位置平均值。

[0053] 举例来说,一个标注框M所对应的三个目标候选框a、b和c内都有鼻尖关键点,a、b和c内的鼻尖关键点与标注框M内所标注的鼻尖关键点的距离分别为1、2和3。如果预设距离阈值为2.5,那么目标候选框a、b所分别对应的距离1、2均小于该预设距离阈值,所以目标候选框a、b内的鼻尖关键点可以参与计算位置平均值。

[0054] 这些实现方式能够从标注框对应的各个目标候选框内的某个关键点中选取距离标注框较近的关键点确定位置平均值,能够避免偏差较大的关键点参与位置平均值的计算,进而提高确定姿态的准确度。

[0055] 继续参见图3,图3是根据本实施例的图像处理方法的应用场景的一个示意图。在图3的应用场景中,执行主体301可以获取已标注对象的姿态的图像302,其中,图像包含至少两个对象,不同对象的姿态不同,姿态通过多个关键点指示。基于图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括:将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态303,确定各个对象的候选姿态304。确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框305。对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值306。将各个关键点的位置平均值的集合作为对图像检测到的一个姿态307。

[0056] 本实施例能够从包括至少两个对象的图像中,通过重合度对各个候选姿态进行筛选,以选取指示对象更加准确的目标候选框。并且,取得关键点的平均值准确地分辨出图像中的各个姿态。

[0057] 进一步参考图4,其示出了图像处理方法的又一个实施例的流程400。该图像处理方法的流程400,包括以下步骤:

[0058] 步骤401,对目标图像中的多个预设姿态进行聚类,得到关键点集合。

[0059] 在本实施例中,图像处理方法运行于其上的执行主体(例如图1所示的服务器或终端设备)可以获取目标图像,并对目标图像中的多个预设姿态进行聚类,以得到关键点集合。具体地,上述执行主体可以采用多种方式对多个预设姿态进行聚类。比如,可以对每一个关键点的位置的坐标进行聚类,以得到各个关键点的聚类结果。

[0060] 在本实施例的一些可选的实现方式中,上述步骤401可以包括以下步骤:

[0061] 对各个预设姿态所对应的多维向量进行聚类,其中,预设姿态所对应的多维向量的维度数量与预设姿态的关键点数量相同;将聚类中心的多维向量所对应的姿态的各个关键点组成关键点集合。

[0062] 在这些实现方式中,预设姿态可以采用多维向量表示。多维向量中的每一个维度的向量都对应预设姿态中的一个关键点的位置的坐标。通过聚类可以得到一个或多个聚类中心。这里的聚类中心也是一个多维向量。上述执行主体可以将这个多维向量所指示的姿态的各个关键点组成关键点集合。

[0063] 步骤402,将各个关键点集合确定为锚点姿态,其中,不同关键点集合所包括的关键点在目标图像中的位置不同。

[0064] 在本实施例中,上述执行主体可以将得到的各个关键点集合确定为锚点姿态。这样,所得到的各个锚点姿态的位置是更加差别化的。同时,本实施例也能够对多个预设姿态进行聚类,以得到准确的锚点姿态。这样,在检测姿态的过程中,能够减小检测得到的候选

姿态与锚点姿态的偏差。

[0065] 步骤403,获取已标注对象的姿态的图像,其中,图像包含至少两个对象,不同对象的姿态不同,姿态通过多个关键点指示。

[0066] 在本实施例中,上述执行主体可以获取已标注对象的姿态的图像。在图像中,对象的姿态被标注出来。这里的对象可以是人、人脸、猫、物品等等。具体地,可以通过关键点的坐标来表示姿态。

[0067] 步骤404,基于图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括步骤4041、步骤4042和步骤4043,如下:

[0068] 步骤4041,将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态。

[0069] 在本实施例中,上述执行主体可以将获取的图像输入卷积神经网络,从而基于卷积神经网络的在先设置的锚点姿态,由卷积神经网络得到各个对象的候选姿态。具体地,卷积神经网络中包括区域候选网络。锚点姿态在图像中的尺寸和位置是固定的。

[0070] 步骤4042,确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框。

[0071] 在本实施例中,上述执行主体可以确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度。之后,上述执行主体可以选取重合度大于预设重合度阈值的候选框,并将所选取的候选框作为目标候选框。

[0072] 步骤4043,对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值;将各个关键点的位置平均值的集合作为对图像检测到的一个姿态。

[0073] 在本实施例中,上述执行主体可以对于每个标注框所对应的目标候选框内的每一个关键点,取该标注框所对应的各个目标候选框内候选姿态中的该关键点的位置平均值。从而,上述执行主体可以将该标注框所对应的目标候选框的各个关键点的位置平均值的集合作为对图像检测到的一个姿态。

[0074] 本实施例所得到的各个锚点姿态是更加差别化的,有利于在获得丰富的锚点姿态的同时,控制锚点姿态的数量。这样,既能够提高区域候选网络的运算速度,也能够确保检测得到的候选姿态与锚点姿态的偏差较小。并且,本实施例也能够对多个预设姿态进行聚类,得到准确的锚点姿态,从而进一步减小检测得到的候选姿态与锚点姿态的偏差。

[0075] 进一步参考图5,作为对上述各图所示方法的实现,本申请提供了一种图像处理装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0076] 如图5所示,本实施例的图像处理装置500包括:获取单元501和训练单元502。其中,获取单元501,被配置成获取已标注对象的姿态的图像,其中,图像包含至少两个对象,不同对象的姿态不同,姿态通过多个关键点指示;训练单元502,被配置成基于图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括:将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态;确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框;对于每个标注框所对应的目标候选框内的每一个关键

点,取各个目标候选框内的该关键点的位置平均值;将各个关键点的位置平均值的集合作为对图像检测到的一个姿态。

[0077] 在一些实施例中,图像处理装置500的获取单元501可以获取已标注对象的姿态的图像。在图像中,对象的姿态被标注出来。这里的对象可以是人、人脸、猫、物品等等。具体地,可以通过关键点的坐标来表示姿态。举例来说,人在呈站立姿态和下蹲姿态时,鼻尖关键点的坐标与脚尖关键点的坐标的距离不同。

[0078] 在一些实施例中,训练单元502可以将获取的图像输入卷积神经网络,从而基于卷积神经网络中的在先设置的锚点姿态,由卷积神经网络得到各个对象的候选姿态。之后,选取重合度大于预设重合度阈值的候选框,并将所选取的候选框作为目标候选框。上述执行主体还可以对于每个标注框所对应的目标候选框内的每一个关键点,取该标注框所对应的各个目标候选框内候选姿态中的该关键点的位置平均值。

[0079] 在本实施例的一些可选的实现方式中,装置还包括:聚类单元,被配置成对目标图像中的多个预设姿态进行聚类,得到关键点集合;确定单元,被配置成将各个关键点集合确定为锚点姿态,其中,不同关键点集合所包括的关键点在目标图像中的位置不同。

[0080] 在一些实施例中,聚类单元,进一步被配置成:对各个预设姿态所对应的多维向量进行聚类,其中,预设姿态所对应的多维向量的维度数量与预设姿态的关键点数量相同;将聚类中心的多维向量所对应的预设姿态的各个关键点组成关键点集合。

[0081] 在本实施例的一些可选的实现方式中,训练单元,进一步被配置成:对于每个标注框所对应的每个目标候选框内的每一个关键点,响应于确定该关键点的位置在该标注框以外,将预设的第一预设权重作为该目标候选框内的该关键点的权重;响应于确定该关键点的位置在该标注框以内,将预设的第二预设权重作为该目标候选框内的该关键点的权重,第一预设权重小于第二预设权重;基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

[0082] 在本实施例的一些可选的实现方式中,训练单元,进一步被配置成:对于每个标注框所对应的每个目标候选框内的每一个关键点,确定该关键点与已标注的姿态中该关键点的距离是否小于或等于预设距离阈值;响应于确定小于或者等于,基于该标注框所对应的各个目标候选框内的该关键点的权重,确定各个目标候选框内的该关键点的位置平均值。

[0083] 下面参考图6,其示出了适于用来实现本申请实施例的电子设备的计算机系统600的结构示意图。图6示出的电子设备仅仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0084] 如图6所示,计算机系统600包括中央处理单元(CPU和/或GPU)601,其可以根据存储在只读存储器(ROM)602中的程序或者从存储部分608加载到随机访问存储器(RAM)603中的程序而执行各种适当的动作和处理。在RAM 603中,还存储有系统600操作所需的各种程序和数据。中央处理单元601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0085] 以下部件连接至I/O接口605:包括键盘、鼠标等的输入部分606;包括诸如阴极射线管(CRT)、液晶显示屏(LCD)等以及扬声器等的输出部分607;包括硬盘等的存储部分608;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609。通信部分609经由诸如因特网的网络执行通信处理。驱动器610也根据需要连接至I/O接口605。可拆卸介质611,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610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608。

[0086] 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,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分609从网络上被下载和安装,和/或从可拆卸介质611被安装。在该计算机程序被中央处理单元601执行时,执行本申请的方法中限定的上述功能。需要说明的是,本申请的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一—但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0087] 附图中的流程图和框图,图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0088] 描述于本申请实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中,例如,可以描述为:一种处理器包括获取单元和训练单元。其中,这些单元的名称在某种情况下并不构成对该单元本身的限定,例如,获取单元还可以被描述为“获取已标注对象的姿态的图像的单元”。

[0089] 作为另一方面,本申请还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的装置中所包含的;也可以是单独存在,而未装配入该装置中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被该装置执行时,使得该装置:获取已标注对象的姿态的图像,其中,图像包含至少两个对象,不同对象的姿态不同,

姿态通过多个关键点指示;基于图像和对姿态的标注,对卷积神经网络进行训练,得到训练后的卷积神经网络,训练过程包括:将图像输入卷积神经网络,基于卷积神经网络的在先设置的锚点姿态,确定各个对象的候选姿态;确定各个候选姿态所在的候选框与已标注的姿态的标注框的重合度,将重合度大于预设重合度阈值的候选框作为目标候选框;对于每个标注框所对应的目标候选框内的每一个关键点,取各个目标候选框内的该关键点的位置平均值;将各个关键点的位置平均值的集合作为对图像检测到的一个姿态。

[0090] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本申请中所涉及的发明范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

100

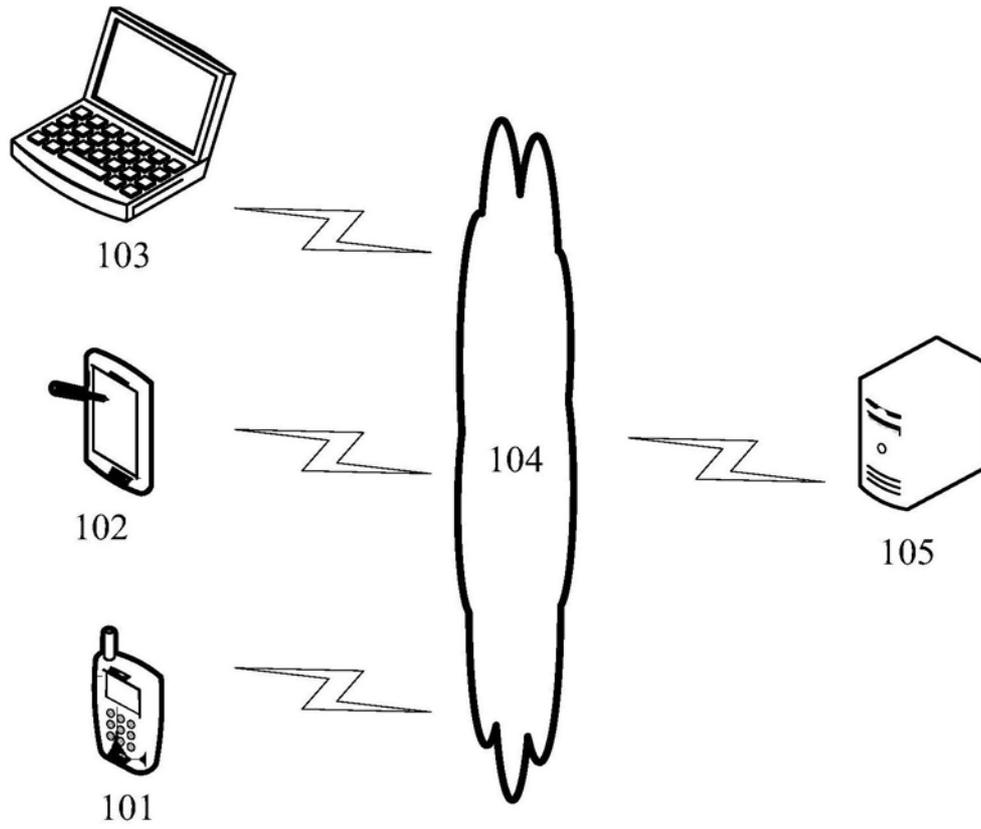


图1

200

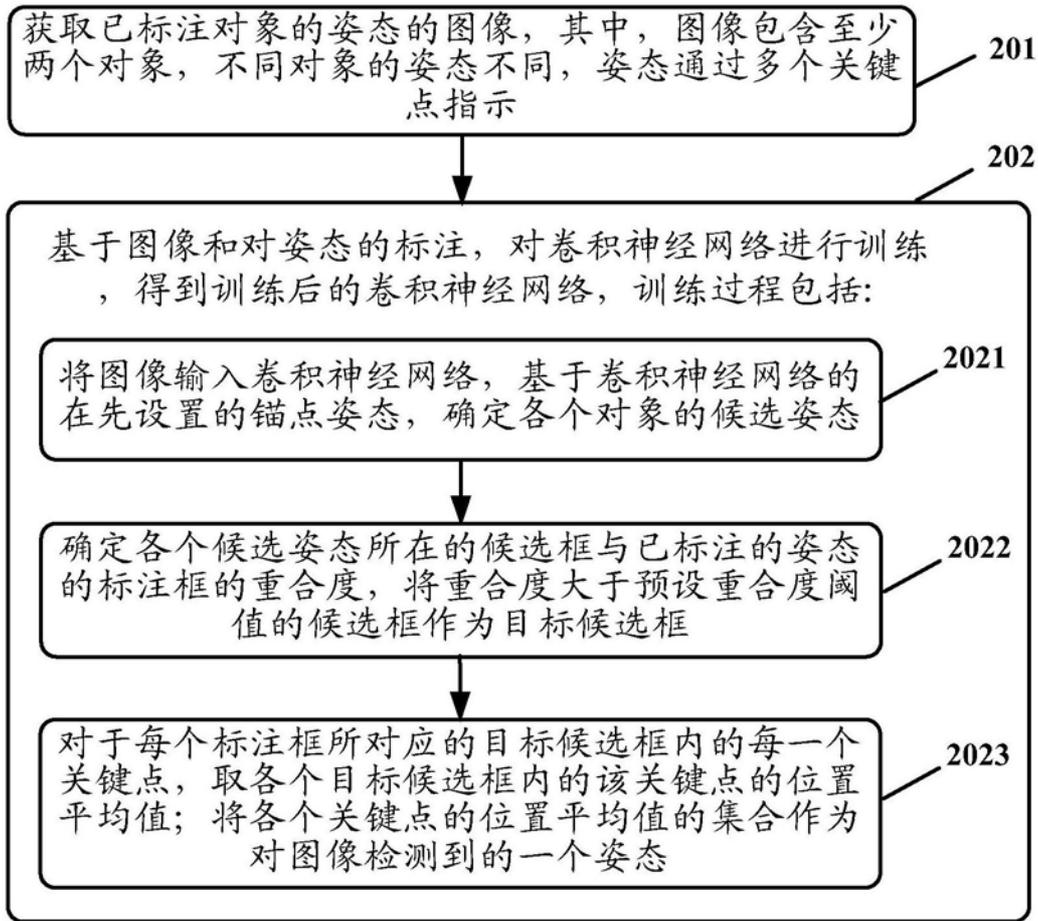


图2

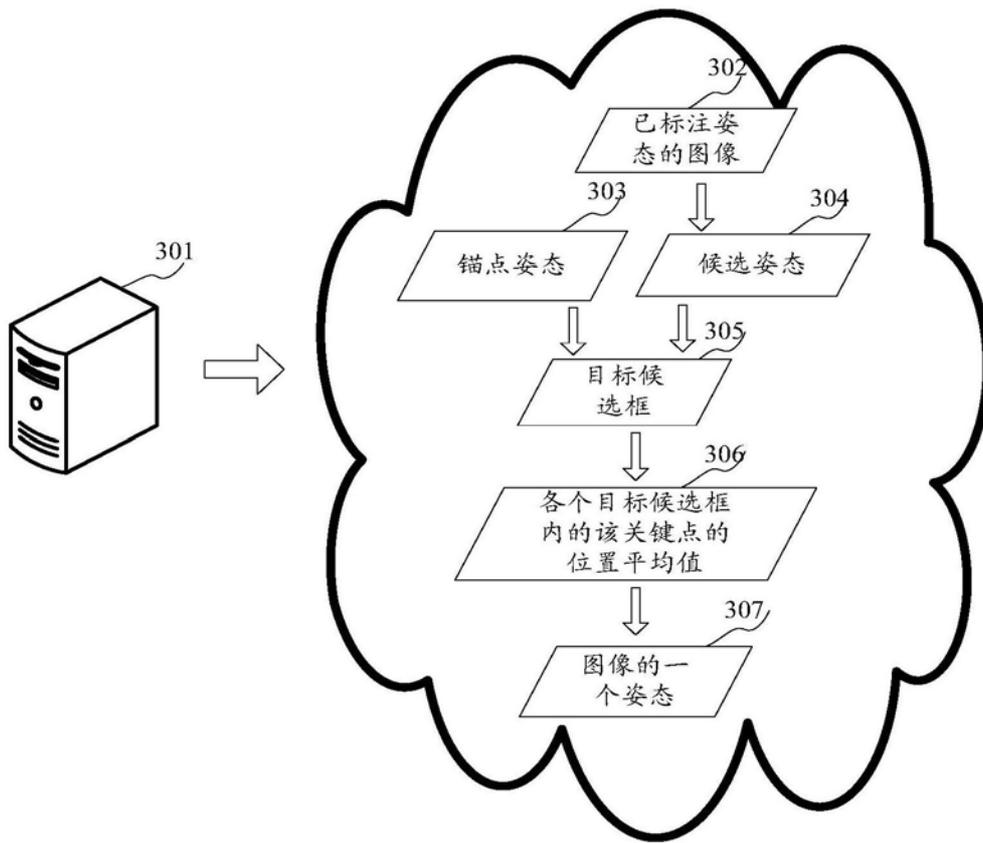


图3

400

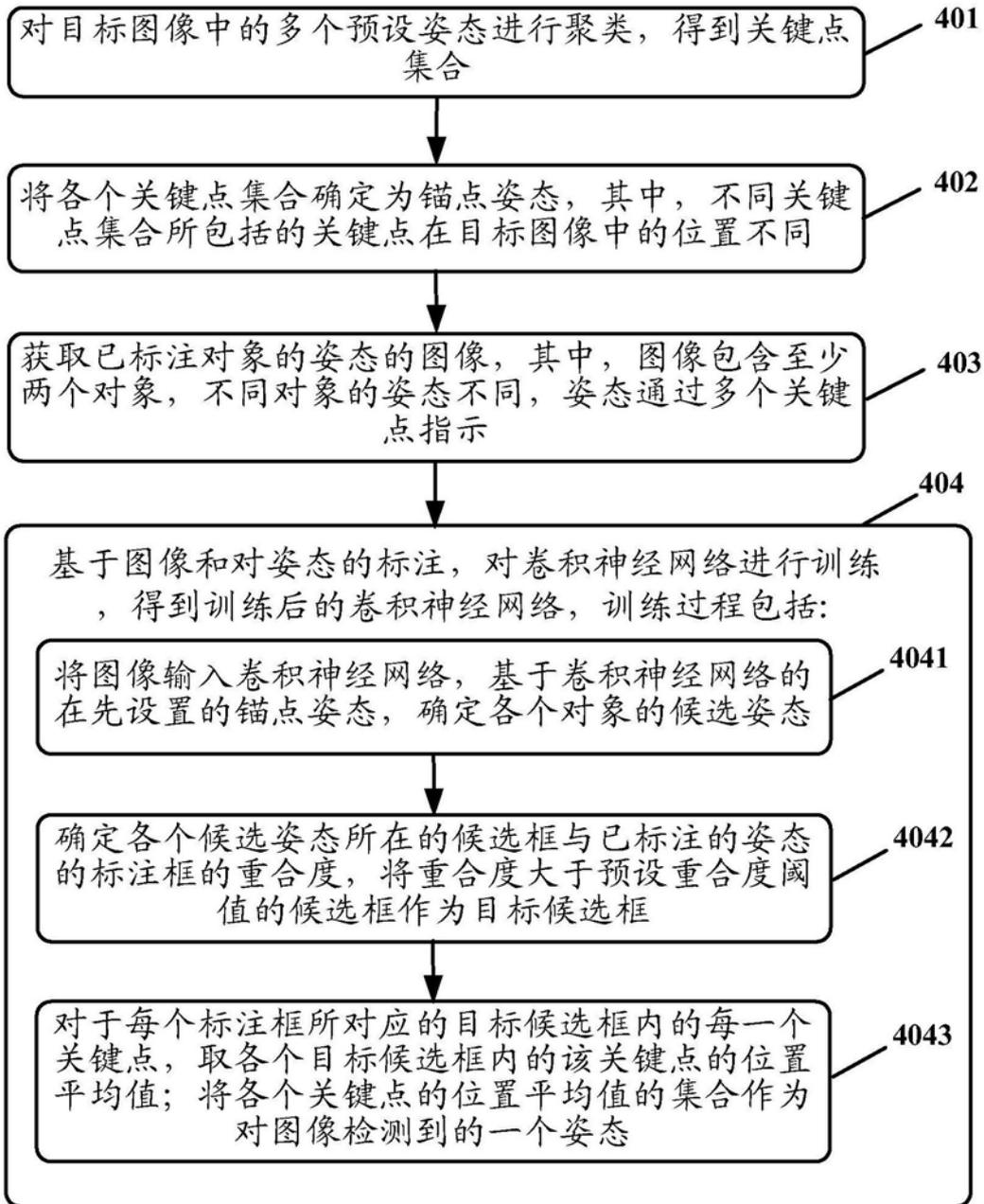


图4

500

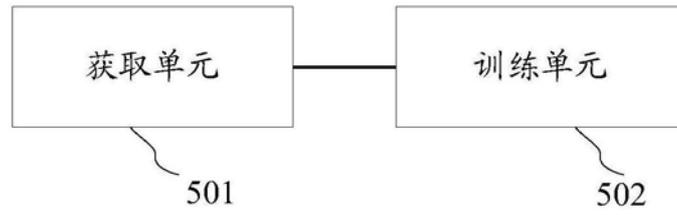


图5

600

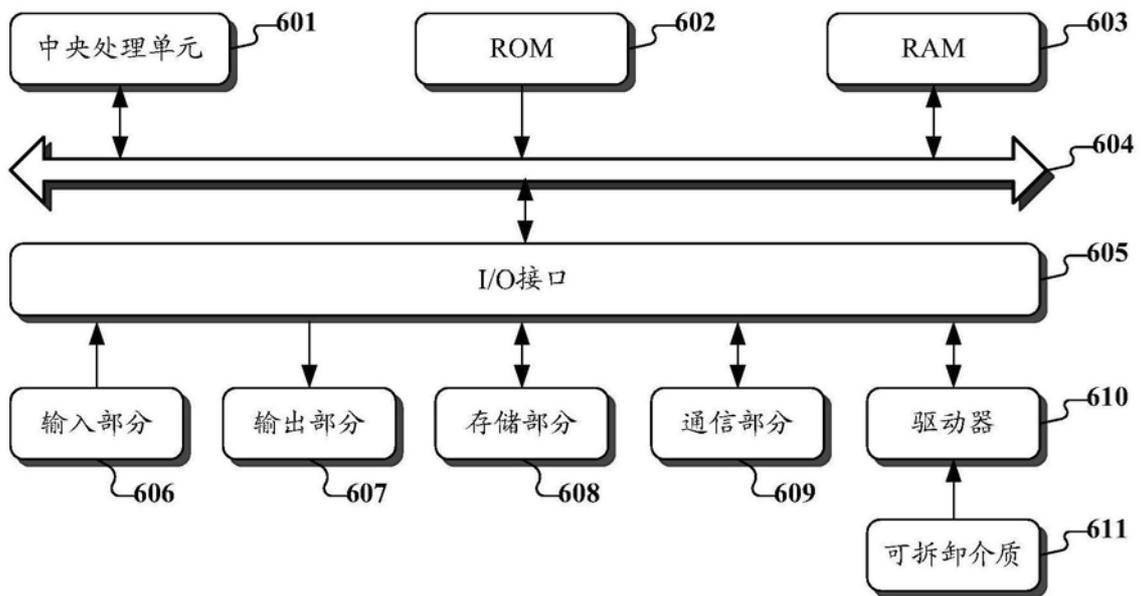


图6