

用于物体检测的方法、装置和设备

本申请要求在 2017 年 11 月 22 日提交中国专利局、申请号为 CN 201711175559.8、发明名称为“用于物体检测的方法、装置和设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机视觉技术，尤其是涉及一种用于物体检测的方法、装置、电子设备以及计算机可读存储介质。

背景技术

在图像识别等计算机视觉领域中，往往需要快速准确的实现物体尤其是通用物体的检测。

现有的通用物体检测通常包括：物体定位过程和物体类别识别过程；其中的物体定位过程主要用于确定出图像中的物体的外接框，而物体类别识别过程主要用于确定出物体所属类别。

现有的物体定位的实现方式通常为：利用选择性搜索（Selective Search）算法或者候选区域生成网络（Region Proposal Network, RPN）等方式获得图像中可能是物体的多个候选框，再通过卷积神经网络对各候选框进行回归处理，以使候选框的位置更加精准。

现有的物体类别识别的实现方式通常为：针对物体定位过程所确定出的候选框，将卷积神经网络（Convolutional Neural Network, CNN）作为分类器，从而可以通过 CNN 确定出每一个候选框属于各预设类别的置信度，并将置信度最高的类别作为候选框所属类别。

发明内容

本申请实施方式提供了一种用于物体检测的方法、装置和设备。

根据本申请实施方式的一个方面，提供了一种用于物体检测的方法，所述方法包括：获取待检测图像中的多个待定目标；确定所述多个待定目标分别针对至少一个类别的置信度，根据所述置信度确定多个待定目标所属类别，确定所述多个待定目标各自所属类别所对应的位置偏移值；将所述多个待定目标各自所属类别所对应的位置偏移值作为所述多个待定目标的位置偏移值；根据所述多个待定目标所属类别、所述多个待定目标的位置偏移值以及所述多个待定目标针对所属类别的置信度，确定所述待检测图像中至少一个待定目标的位置信息和所属类别。

在本申请一实施方式中，所述获取待检测图像中的多个待定目标的步骤包括：获取所述待检测图像中基于至少一个预设区域大小形成的多个待定目标。

在本申请又一实施方式中，所述获取所述待检测图像中基于至少一个预设区域大小形成的多个待定目标的步骤包括：获取待检测图像的特征图；针对所述特征图中的一个特征点而言，根据基准框配置信息形成所述特征点的基准框，将所述特征点的基准框作为一个待定目标；获得所述特征图中多个特征点对应的多个待定目标。

在本申请再一实施方式中，所述获取待检测图像的特征图的步骤包括：将所述待检测图像输入卷积神经网络中的主干网络；所述主干网络输出的特征图输入所述卷积神经网络中的滤波层，由所述滤波层根据预设滑窗对所述特征图进行滤波处理，将所述滤波处理后的特征图作为所述待检测图像的特征图。

在本申请再一实施方式中，所述获取所述待检测图像中基于至少一个预设区域大小形成的多个待定目标的步骤包括：获取待检测图像的特征图；基于基准框配置信息对所述特征图进行池化处理，获得多个新的特征图，将所述多个新的特征图作为多个待定目标。

在本申请再一实施方式中，所述预设区域大小包括：基于三种不同的长度和三种不同的长宽比所形

成的九个预设区域大小；或三种不同的宽度和三种不同的长宽比所形成的九个预设区域大小；或基于三种不同的长度和宽度所形成的九个预设区域大小。

在本申请再一实施方式中，所述类别包括：至少两种物体类别以及一个背景类别。

在本申请再一实施方式中，所述确定所述多个待定目标分别针对至少一个类别的置信度，根据所述置信度确定多个待定目标所属类别的步骤包括：针对一个待定目标，计算所述待定目标分别属于所述至少一个类别的置信度，并将最高置信度对应的类别作为所述待定目标所属类别。

在本申请再一实施方式中，所述确定所述多个待定目标各自所属类别所对应的位置偏移值的步骤包括：针对一个待定目标，计算所述待定目标针对所属类别的上偏移值、下偏移值、左偏移值以及右偏移值。

在本申请再一实施方式中，所述待定目标的位置信息包括：待定目标的外接框位置信息。

在本申请再一实施方式中，所述根据所述多个待定目标所属类别、所述多个待定目标的位置偏移值以及所述多个待定目标针对所属类别的置信度，确定所述待检测图像中至少一个待定目标的位置信息和所属类别的步骤包括：从所述多个待定目标中选取获得置信度满足预设置信度要求的多个待定目标；根据所述选取出的多个待定目标的位置偏移值形成待检测图像中的外接框位置信息；将所述选取出的多个待定目标所属类别作为相应外接框所属类别。

在本申请再一实施方式中，所述确定所述多个待定目标分别针对至少一个类别的置信度，根据所述置信度确定多个待定目标所属类别，确定所述多个待定目标各自所属类别所对应的位置偏移值，包括：利用卷积神经网络，确定所述多个待定目标分别针对至少一个类别的置信度，根据所述置信度确定多个待定目标所属类别，确定所述多个待定目标各自所属类别所对应的位置偏移值；所述方法还包括：训练所述卷积神经网络，且所述训练所述卷积神经网络的步骤包括：从图像样本集合中获取标注有至少一个标准位置信息和标准位置所属类别的图像样本；获取图像样本中的多个待定目标；利用卷积神经网络中的一卷积层确定所述多个待定目标分别针对至少一个类别的置信度，根据所述置信度确定多个待定目标所属类别；利用卷积神经网络中的另一卷积层确定所述多个待定目标各自所属类别所对应的位置偏移值；将所述多个待定目标各自所属类别所对应的位置偏移值作为所述多个待定目标的位置偏移值；计算所述多个待定目标相对于相应的标准位置的标准位置偏移值；利用卷积神经网络中的一损失层计算所述多个待定目标中至少一个待定目标针对相应的标准位置的类别的位置偏移值与相应的标准位置偏移值的误差，利用卷积神经网络中的一损失层计算所述多个待定目标中至少一个待定目标针对相应标准位置所属类别的置信度与标准位置所属类别之间的误差，并将所述误差在所述卷积神经网络中反向传播；直到从图像样本集合中获取的图像样本的数量达到预定数量，且在误差属于预定范围的情况下，所述卷积神经网络完成训练。

在本申请再一实施方式中，所述获取图像样本中的多个待定目标的步骤包括：获取图像样本中基于至少一个预设区域大小形成的多个待定目标。

在本申请再一实施方式中，所述标准位置包括：标准外接框。

在本申请再一实施方式中，所述计算所述多个待定目标相对于相应的标准位置的标准位置偏移值的步骤包括：确定所述多个待定目标各自对应的重叠面积最大的标准外接框，并分别计算多个待定目标相对于重叠面积最大的标准外接框的标准位置偏移值。

在本申请再一实施方式中，所述所述多个待定目标中至少一个待定目标包括：从所述多个待定目标中的多个正待定目标中选取出的正待定目标以及从多个待定目标中的多个负待定目标中选取出的负待定目标，且选取出的正待定目标和选取出的负待定目标满足预定比例；针对一待定目标而言，如果所述待定目标与其重叠面积最大的标准外接框之间的重叠面积与所述待定目标和标准外接框共同覆盖的面

积的比值大于第一比例阈值，则所述待定目标为正待定目标；如果所述比值小于第二比例阈值，则所述待定目标为负待定目标。

根据本申请实施方式的另一个方面，提供了一种用于物体检测的装置，包括：获取待定目标模块，用于获取待检测图像中的多个待定目标；预测类别和位置偏移值模块，用于确定所述多个待定目标分别针对至少一个类别的置信度，根据所述置信度确定多个待定目标所属类别，确定所述多个待定目标各自所属类别所对应的位置偏移值；确定位置偏移模块，用于将所述多个待定目标各自所属类别所对应的位置偏移值作为所述多个待定目标的位置偏移值；确定检测结果模块，用于根据所述多个待定目标所属类别、所述多个待定目标的位置偏移值以及所述多个待定目标针对所属类别的置信度，确定所述待检测图像中至少一个待定目标的位置信息和所属类别。

在本申请一实施方式中，所述获取待定目标模块具体用于：获取所述待检测图像中基于至少一个预设区域大小形成的多个待定目标。

在本申请又一实施方式中，所述获取待定目标模块包括：第一获取待定目标子模块，用于获取待检测图像的特征图；针对所述特征图中的一个特征点而言，根据基准框配置信息形成所述特征点的基准框，将所述特征点的一个基准框作为一个待定目标；获得所述特征图中多个特征点对应的多个待定目标。

在本申请再一实施方式中，所述获第一获取待定目标子模块具体用于：将所述待检测图像输入卷积神经网络中的主干网络；所述主干网络输出的特征图被输入所述卷积神经网络中的滤波层，由所述滤波层根据预设滑窗对所述特征图进行滤波处理，将所述滤波处理后的特征图作为所述待检测图像的特征图。

在本申请再一实施方式中，所述获取待定目标模块包括：第二获取待定目标子模块，用于获取待检测图像的特征图；基于基准框配置信息对所述特征图进行池化处理，获得多个新的特征图，将所述多个新的特征图作为多个待定目标。

在本申请再一实施方式中，所述预设区域大小包括：三种不同的长度和三种不同的长宽比所形成的九个预设区域大小；或三种不同的宽度和三种不同的长宽比所形成的九个预设区域大小；或基于三种不同的长度和宽度所形成的九个预设区域大小。

在本申请再一实施方式中，所述类别包括：至少两种物体类别以及一个背景类别。

在本申请再一实施方式中，所述预测类别和位置偏移值模块具体用于：针对一个待定目标，计算所述待定目标分别属于所述至少一个类别的置信度，并将最高置信度对应的类别作为所述待定目标所属类别。

在本申请再一实施方式中，所述预测类别和位置偏移值模块具体用于：针对一个待定目标，计算所述待定目标针对所属类别的上偏移值、下偏移值、左偏移值以及右偏移值。

在本申请再一实施方式中，所述待定目标的位置信息包括：待定目标的外接框位置信息。

在本申请再一实施方式中，所述确定检测结果模块具体用于：从所述多个待定目标中选取获得置信度满足预设置信度要求的多个待定目标；根据所述选取出的多个待定目标的位置偏移值形成待检测图像中的外接框位置信息；将所述选取出的多个待定目标所属类别作为相应外接框所属类别。

在本申请再一实施方式中，所述预测类别和位置偏移值模块利用卷积神经网络实现，所述装置还包括：训练模块，所述训练模块用于：从图像样本集合中获取标注有至少一个标准位置信息和标准位置所属类别的图像样本；获取图像样本中的多个待定目标；利用卷积神经网络中的一卷积层确定所述多个待定目标分别针对至少一个类别的置信度，根据所述置信度确定多个待定目标所属类别；利用卷积神经网络中的另一卷积层确定所述多个待定目标分别各自所属类别所对应的位置偏移值；将所述多个待定目标各自所属类别所对应的位置偏移值作为所述多个待定目标的位置偏移值；计算所述多个待定目标相对于相应的标准位置的标准位置偏移值；利用卷积神经网络中的一损失层计算所述多个待定目标中至少一个

待定目标针对相应的标准位置的类别的位置偏移值与相应的标准位置偏移值的误差,利用卷积神经网络中的一损失层计算所述多个待定目标中至少一个待定目标针对相应标准位置所属类别的置信度与标准位置所属类别之间的误差,并将所述误差在所述卷积神经网络中反向传播;直到从图像样本集合中获取的图像样本的数量达到预定数量,且在误差属于预定范围的情况下,所述卷积神经网络完成训练。

在本申请再一实施方式中,所述训练模块具体用于:获取图像样本中基于至少一个预设区域大小形成的多个待定目标。

在本申请再一实施方式中,所述标准位置包括:标准外接框。

在本申请再一实施方式中,所述训练模块具体用于:确定所述多个待定目标各自对应的重叠面积最大的标准外接框,并分别计算多个待定目标相对于重叠面积最大的标准外接框的标准位置偏移值。

在本申请再一实施方式中,所述部分待定目标包括:从所述多个待定目标中的多个正待定目标中选出的正待定目标以及从多个待定目标中的多个负待定目标中选出的负待定目标,且选出的正待定目标和选出的负待定目标满足预定比例;针对一待定目标而言,如果所述待定目标与其重叠面积最大的标准外接框之间的重叠面积与所述待定目标和标准外接框共同覆盖的面积之比大于第一比例阈值,则所述待定目标为正待定目标;如果所述比值小于第二比例阈值,则所述待定目标为负待定目标。

根据本申请实施方式的还一个方面,提供了一种电子设备,其特征在于,包括处理器,所述处理器包括上述任意一项实施方式所述的用于物体检测的装置。

根据本申请实施方式的再一个方面,提供了一种电子设备,包括:存储器,用于存储计算机程序;处理器,用于执行所述存储器中存储的计算机程序,且所述计算机程序被执行时,实现上述任意一项实施方式所述的用于物体检测的方法中的各步骤。

根据本申请实施方式的再一个方面,提供了一种计算机存储介质,其上存储有计算机程序,该计算机程序被处理器执行时,实现上述任意一项实施方式所述的用于物体检测的方法中的各步骤。

根据本申请实施方式的再一个方面,提供了一种计算机程序产品,包括计算机可读代码,其特征在于,当所述计算机可读代码在设备上运行时,所述设备中的处理器执行用于实现上述任意一项实施方式所述的用于物体检测的方法的指令。

基于本申请上述实施方式提供的用于物体检测的方法、装置和设备,本申请实施方式通过在确定各待定目标所属类别的同时,确定各待定目标分别针对每一类别的位置偏移值,使确定类别的过程和计算位置偏移的过程可以并行执行,从而可以降低两个过程必须先后顺序执行而对物体检测实时性的影响;通过将各待定目标各自所属类别所对应的位置偏移值作为多个待定目标的位置偏移值,可以通过简单的选择操作而快速的确定出各待定目标的位置偏移值,降低了确定位置偏移值耗时较长而对物体检测实时性的影响;通过上述过程本申请可以基于置信度从所有待定目标中快速的选取相应的待定目标,基于选取出的相应的待定目标的位置偏移值快速准确的确定出其位置信息;另外,在本申请预先设置的类别包括三个或者更多的情况下,本申请提供的技术方案可以实现通用物体检测;由此可知,本申请提供的技术方案有利于提高物体检测的实时性,从而使本申请提供的技术方案能够较好的适用于对实时性要求较高的物体检测应用场景中。

下面通过附图和实施例,对本申请的技术方案做进一步的详细描述。

附图说明

构成说明书的一部分的附图描述了本申请的实施例,并且连同描述一起用于解释本申请的原理。

参照附图,根据下面的详细描述,可以更加清楚地理解本申请,其中:

图1为本申请实施例提供的一个应用场景示意图。

图 2 为本申请实施例提供的用于物体检测的方法的一个流程示意图。

图 3 为本申请实施例提供的用于物体检测的方法的另一流程示意图。

图 4 为本申请实施例提供的用于物体检测的方法中对卷积神经网络进行训练的一个示例的流程图。

图 5 为本申请实施例提供的用于物体检测的方法中对卷积神经网络进行训练的一个示例的训练示意图。

图 6 为本申请实施例提供的用于物体检测的装置的一个结构示意图。

图 7 为本申请电子设备一个实施方式的结构示意图。

图 8 为本申请计算机存储介质一个实施方式的示意图。

具体实施方式

现在将参照附图来详细描述本申请的各种示例性实施例。应注意到：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申请的范围。

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。

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，决不作为对本申请及其应用或使用的任何限制。

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。

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。

本申请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。

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。

图 1 为本申请实施例提供的一个应用场景示意图。参考图 1，示意性地示出了根据本申请实施方式的可以在其中实现的应用场景。

图 1 中，汽车 100 中安装有图像采集装置，该图像采集装置能够采集到汽车 100 外部的空间的图像，汽车 100 在处于自动驾驶过程中，需要针对图像采集装置所采集到的图像进行物体检测，以辨识道路上的行人 110、动物、斑马线以及交通信号灯 120 等物体；例如，通过对图像采集装置所采集到的图像进行物体检测，在确定出前方的交通信号灯 120 由绿变红时，通过及时控制汽车 100 的刹车装置等部件，使处于行驶状态的汽车 100 安全地停在前车的后方或者道路上的停车线的后方；后续，在确定出前方的交通信号灯 120 由红变绿、前方斑马线上没有行人 110、且汽车 100 与前车保持安全距离等情况下，通过及时控制汽车 100 的刹车装置等部件，使处于停止状态的汽车 100 继续在道路上前行。

然而，本领域技术人员完全可以理解，本申请实施方式的适用场景不受到该框架任何方面的限制。

本申请实施方式的用于物体检测的方法也可以称为用于通用物体检测的方法。图 2 为本申请实施例提供的用于物体检测的方法的一个流程示意图。该方法可以由任意电子设备执行，例如终端设备、服务器、移动设备、车载设备等等。如图 2 所示，本申请的用于物体检测的方法包括：步骤 S200、步骤 S210、步骤 S220 以及步骤 S230。下面对图 2 中的各步骤分别进行说明。

S200、获取待检测图像中的多个待定目标。

在一个可选示例中，该步骤 S200 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的获取待定目标模块 600 执行。

在一个可选示例中，本申请实施例中的待检测图像可以为图片或者实时摄取的视频中的一帧。本申请实施例获取待检测图像中的多个待定目标可以为：获取待检测图像中的基于至少一个预设区域大小（通常为多个预设区域大小）而形成的多个待定目标。本申请中的预设区域大小可以由预先设置的基准框配置信息确定；待定目标可以为待检测图像的特征图中的特征点的基准框，也可以为在待检测图像的特征图的基础上而形成的小特征图（即局部特征图）。

获取待检测图像中的基于至少一个预设区域大小而形成的多个待定目标的一个例子为：首先，获取待检测图像的特征图；然后，针对该特征图中的任意一个特征点，根据预先设置的基准框配置信息形成该特征点的多个基准框，从而可以获得特征图中的至少一个特征点的基准框，而每一个特征点的每一个基准框均被作为一个待定目标，从而获得多个特征点对应的多个待定目标。

获取待检测图像中的基于至少一个预设区域大小而形成的多个待定目标的另一个例子为：首先，获取待检测图像的特征图；然后，基于预先设置的基准框配置信息对上述特征图进行池化处理，从而可以形成多个新的特征图（即多个小特征图或多个局部特征图），而多个新的特征图被作为多个待定目标，（例如：将每个新的特征图分别作为一个待定目标）。

在上述两个例子中，获取待检测图像的特征图的一种方式可以为：利用现有的 VGG16（Visual Geometry Group，视觉组）网络形成本申请实施方式的卷积神经网络中的主干网络，通过将待检测图像提供给该主干网络，该主干网络即可基于 VGG16 技术输出待检测图像的特征图，从而本申请实施方式可以基于该主干网络的输出获取到待检测图像的特征图。

在上述两个例子中，获取待检测图像的特征图的另一种方式可以为：利用现有的谷歌网络（GoogleNet）形成本申请实施方式的卷积神经网络中的主干网络，通过将待检测图像提供给该主干网络，该主干网络即可基于 GoogleNet 技术输出待检测图像的特征图，从而本申请实施方式可以基于该主干网络的输出获取到待检测图像的特征图。

在上述两个例子中，获取待检测图像的特征图的再一种方式可以为：利用现有的 ResNet 形成本申请实施方式的卷积神经网络中的主干网络，通过将待检测图像提供给该主干网络，该主干网络即可基于 ResNet 技术输出待检测图像的特征图，从而本申请实施方式可以基于该主干网络的输出获取到待检测图像的特征图。

在上述的一个例子中，本申请实施方式可以将主干网络输出的特征图作为最终获取到的特征图，本申请实施方式也可以对主干网络输出的特征图进行滤波处理，从而基于主干网络输出的特征图中的所有特征点形成具有相同特征点数量的新的特征图，并将该新特征图作为最终获取到的待检测图像的特征图。上述滤波处理的实现方式可以为：设定主干网络输出的所有特征点形成一个特征图，卷积神经网络中的滤波层利用预设滑窗（例如，3×3 的滑窗）依次从特征图中选取预定数量的特征点（例如，9 个特征点），并利用每一次选取出的所有特征点形成一个新的特征点，滤波层输出获得的所有新的特征点，且滤波层输出的所有新的特征点同样可以形成一个特征图，然而，新的特征图中的每一个特征点所对应的待检测图像中的区域比主干网络输出的特征图中的相应特征点所对应的待检测图像中的区域更大。本申请实施

方式不限制滤波层对特征图进行滤波处理的具体实现过程。本申请实施方式通过对主干网络输出的特征图进行滤波处理形成新的特征图，有利于提高确定出的外接框位置信息以及外接框所属类别的准确性，从而有利于提高物体检测的准确性。

在一个可选示例中，本申请实施方式通过预先设置的基准框配置信息来定义基准框（即 Anchor，锚点）的数量以及尺寸，即基准框的数量及大小均是已知的，且基准框在图像中的位置与基准框对应的特征点相关。在通常情况下，多个基准框的尺寸并不完全相同，即多个基准框的面积并不完全相同，如多个基准框的长度存在差异，多个基准框的宽度存在差异，且多个基准框的长宽比存在差异。

一个可选的例子，本申请实施方式的基准框配置信息中预先设置了三种不同的长度，并预先设置了三种不同的长宽比，从而利用预先设置的三种不同的长度以及三种不同的长宽比，可以形成九个不同形状的基准框；例如，预先设置的三种不同的长度可以分别为 64、128 以及 256（单位为像素），预先设置的三种不同的长宽比可以分别为 0.5、1.0 以及 2.0，则由上述三种不同的长度以及三种不同的长宽比所形成的九个基准框可以表示为：64×32、64×64、64×128、128×64、128×128、128×256、256×128、256×256 以及 256×512。

另一个可选的例子，本申请实施方式的基准框配置信息中预先设置了三种不同的宽度，并预先设置了三种不同的长宽比，从而利用预先设置的三种不同的宽度以及三种不同的长宽比，可以形成九个不同形状的基准框。

再一个可选的例子，本申请实施方式的基准框配置信息预先针对九个基准框分别设置其各自的长度以及宽度，从而可以明确的确定出九个形状各异的基准框。

本申请中的基准框的数量以及各基准框的尺寸大小可以根据实际经验设置，本申请实施方式不限制基准框的具体数量，也不限制基准框的尺寸。

S210、确定多个待定目标分别针对至少一个类别的置信度，根据置信度确定多个待定目标所属类别，确定多个待定目标各自所属类别所对应的位置偏移值。

在一个可选示例中，该步骤 S210 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的预测类别和位置偏移值模块 610 执行。

在一个可选示例中，本申请可以利用卷积神经网络中的一卷积层确定多个待定目标（例如，所有待定目标）分别针对至少一个类别的置信度，并根据获得的置信度确定多个待定目标分别所属类别。本申请中的卷积神经网络可以是利用大量的带有标注的图像样本训练而成的，该卷积神经网络的训练过程的一个可选例子可以参见下述针对图 4 的描述，在此不再重复说明。本申请实施例中的神经网络可以包括但不限于：主干网络、滤波层和/或池化层以及两个卷积层（例如，第一卷积层和第二卷积层）。

在一个可选示例中，本申请实施例预先设置有至少三种类别，例如，至少两种物体类别以及一个背景类别。例如，预先针对人、狗、火车以及椅子等不同种类的物体设置了 20 种物体类别，并预先针对背景设置了背景类别，总共 21 种类别。类别的数量以及待定目标的数量（例如，基准框的数量）会决定卷积神经网络中的相应卷积层的输出维度。本申请不限制物体类别的数量以及物体类别所针对的物体的具体内容等。

在一个可选示例中，本申请实施方式可以通过卷积层针对置信度的计算来确定待定目标所属类别，例如，通过 RPN 中的用于预测置信度的卷积层等针对置信度的计算来确定待定目标所属类别。例如，设定预先设置有 21 种类别，即第一物体类别、第二物体类别、……、第二十物体类别以及背景类别，在上述设定的情况下，针对第一待定目标而言，可以利用卷积神经网络中的一卷积层计算出第一待定目标属于第一物体类别的置信度、第一待定目标属于第二物体类别的置信度、……、第一待定目标属于第二十物体类别的置信度以及第一待定目标属于背景类别的置信度（总共 21 个置信度），通过比较这

21 个置信度，可以确定出一个最高置信度，从而可以将该最高置信度对应的类别作为第一待定目标所属类别；以此类推，本申请实施方式可以确定出多个待定目标分别所属类别。

在一个可选示例中，本申请实施例可以利用卷积神经网络中的另一卷积层确定多个待定目标（例如，所有待定目标）各自所属类别所对应的位置偏移值。

本申请实施例中，待定目标各自所属类别所对应的位置偏移值可通过待定目标分别针对至少一个类别中每个类别的位置偏移值获得。

在一个可选示例中，本申请中的位置偏移值用于确定待定目标的在属于某一类别时的位置信息。本申请中的位置偏移值通常包括：上偏移值、下偏移值、左偏移值以及右偏移值，即本申请实施方式可以利用卷积层计算出待定目标针对每一类别的上偏移值、下偏移值、左偏移值以及右偏移值；其中的上偏移值表示待定目标的顶边在竖直方向上的偏移量，其中的下偏移值表示待定目标的底边在竖直方向上的偏移量，其中的左偏移值表示待定目标的左侧边在水平方向上的偏移量，其中的右偏移值表示待定目标的右侧边在水平方向上的偏移量。例如，设定预先设置有 21 种类别，即第一物体类别、第二物体类别、……、第二十物体类别以及背景类别，在上述设定的情况下，针对第一待定目标而言，可以利用卷积神经网络中的一卷积层计算出第一待定目标针对第一物体类别的上偏移值、下偏移值、左偏移值和右偏移值、第一待定目标针对第二物体类别的上偏移值、下偏移值、左偏移值和右偏移值、……、第一待定目标针对第二十物体类别的上偏移值、下偏移值、左偏移值和右偏移值以及第一待定目标针对背景类别的上偏移值、下偏移值、左偏移值和右偏移值（总共 4×21 个偏移值）；以此类推，本申请实施方式可以计算出所有待定目标分别针对二十一个类别的上偏移值、下偏移值、左偏移值以及右偏移值。

由上述描述可知，本申请实施例中的确定待定目标所属类别的过程与计算待定目标针对每一类别的位置偏移值的过程，两者之间没有相互依赖的关系，因此，这两个过程完全可以并行执行，从而可以降低两个过程需要先后执行而导致的时间成本增加的现象，有利于提高物体检测的实时性。

S220、将多个待定目标（例如，所有待定目标）各自所属类别所对应的位置偏移值作为多个待定目标的位置偏移值（例如，所有待定目标各自的位置偏移值）。

在一个可选示例中，该步骤 S220 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的确定位置偏移模块 620 执行。

在一个可选示例中，本申请实施方式可以从上述确定出的一个待定目标所属类别以及上述计算出的该待定目标针对至少一个类别的位置偏移值最终确定出该待定目标的位置偏移值，即从已经计算出的该待定目标针对至少一个类别的位置偏移值中挑选出上述确定出的该待定目标所属类别对应的位置偏移值，并将该挑选出的位置偏移值作为该待定目标的位置偏移值，以此类推，本申请实施方式可以确定出多个待定目标的位置偏移值。

一个可选的例子，设定已经利用卷积神经网络中的一卷积层确定出第一待定目标属于第一物体类别，而且已经利用卷积神经网络中的另一卷积层计算出第一待定目标针对第一物体类别的上偏移值、下偏移值、左偏移值和右偏移值、第一待定目标针对第二物体类别的上偏移值、下偏移值、左偏移值和右偏移值、……、第一待定目标针对第二十物体类别的上偏移值、下偏移值、左偏移值和右偏移值以及第一待定目标针对背景类别的上偏移值、下偏移值、左偏移值和右偏移值（总共 4×21 个偏移值），在上述各设定的情况下，本申请实施方式可以将已经计算出的第一待定目标针对第一物体类别的上偏移值、下偏移值、左偏移值和右偏移值作为第一待定目标的位置偏移值；以此类推，本申请实施方式可以确定出所有待定目标的位置偏移值。

S230、根据多个待定目标所属类别（例如，所有待定目标所属类别、或部分待定目标所属类别）、多个待定目标的位置偏移值（例如，所有待定目标的位置偏移值、或部分待定目标的位置偏移值）以及

多个待定目标针对所属类别的置信度（例如，所有待定目标所属类别的置信度、或部分待定目标所属类别的置信度），确定待检测图像中至少一个待定目标的位置信息和所属类别。

在一个可选示例中，本申请中的待定目标的位置信息可以为待定目标的外接框位置信息。本申请通常会从多个待定目标中选取部分待定目标，并利用选取出的待定目标的位置偏移值形成外接框，且选取出的待定目标各自所属类别即为相应的外接框所属类别。

在一个可选示例中，本申请实施方式可以利用预设置信度要求来选取部分待定目标，且该预设置信度要求可以为置信度阈值（置信度阈值如 0.8）等；例如，本申请实施方式可以从多个待定目标中选取置信度超过置信度阈值（如大于等于 0.8）的待定目标，之后，可以利用选取出的待定目标的位置偏移值形成相应外接框位置信息，并将选取出的待定目标所属类别作为相应外接框所属类别。另外，本申请实施方式也可以利用比例或者外接框数量阈值来选取部分待定目标，本申请实施方式不限制从多个待定目标中选取部分待定目标的具体实现方式。

又例如，设定通过上述步骤 S210 和步骤 S220 确定出第一待定目标的置信度为 0.9，第一待定目标属于第一物体类别，且第一待定目标的位置偏移值分别为 y_1 、 y_2 、 x_1 以及 x_2 ，在预设置信度要求为置信度不低于置信度阈值 0.8 的情况下，可以根据第一待定目标对应图像中的区域、 y_1 、 y_2 、 x_1 以及 x_2 确定出第一外接框在图像中的位置信息，并确定出该第一外接框所属类别为第一物体类别。

图 3 为本申请实施例提供的用于物体检测的方法的另一流程示意图。如图 3 所示，本申请的用于物体检测的方法主要包括：步骤 S300、步骤 S310、步骤 S320 以及步骤 S330。下面对图 3 中的各步骤分别进行说明。

S300、获取待检测图像的多个特征点，例如，获取待检测图像的各特征点（即所有特征点）。

在一个可选示例中，本申请实施例中的待处理图像可以为呈现静态的图片或者照片等图像，也可以为呈现动态的视频中的视频帧等。本申请实施方式可以利用现有的多种方式来获取待检测图像的特征点，如利用现有的 VGG16 网络或者 GoogleNet 或者 ResNet 形成本申请实施方式的卷积神经网络中的主干网络，从而通过将待检测图像提供给该主干网络，该主干网络即可基于 VGG16 或者 GoogleNet 或者 ResNet 技术输出待检测图像的多个特征点。

在一个可选示例中，本申请实施例可以将主干网络输出的多个特征点作为最终获取到的待检测图像的多个特征点，本申请实施例也可以对主干网络输出的多个特征点进行滤波处理，从而基于主干网络输出的多个特征点形成具有相同数量的新的特征点，并将多个新的特征点作为最终获取到的待检测图像的多个特征点。上述滤波处理的实现方式可以为：设定主干网络输出的多个特征点形成一个特征图，卷积神经网络中的滤波层利用滑窗（例如， 3×3 的滑窗）依次从特征图中选取预定数量的特征点（例如，9 个特征点），并利用每一次选取出的所有特征点形成一个新的特征点，滤波层输出其所获得的所有新的特征点，且滤波层输出的所有新的特征点同样可以形成一个特征图，然而，新的特征图中的每一个特征点所对应的待检测图像中的区域会更大。本申请不限制滤波层对特征图进行滤波处理的具体实现过程。本申请实施例通过对主干网络输出的多个特征点进行滤波处理形成新的特征点，能够更准确的确定出外接框位置信息以及外接框所属类别，从而有利于提高物体检测的准确性。

S310、针对每一个特征点：确定该特征点的至少一个基准框所属类别及其置信度，并确定该特征点的至少一个基准框针对每一类别的位置偏移值。

在一个可选示例中，本申请实施例可以利用卷积神经网络中的一卷积层确定该特征点的至少一个基准框所属类别及其置信度。本申请实施例中的卷积神经网络是利用大量的带有标注的图像样本训练而成的，该卷积神经网络的训练过程的一个可选例子可以参见下述实施方式中的描述。本申请实施例中的卷积神经网络可以包括但不限于：主干网络、滤波层以及至少两个卷积层（如第一卷积层和第二卷积层）。

在一个可选示例中，本申请实施方式预先定义了基准框（即 Anchor，锚点）的数量及尺寸，即基准框的数量以及大小均是已知的，且基准框在图像中的位置与基准框对应的特征点相关。在通常情况下，多个基准框的尺寸并不完全相同，即多个基准框的面积并不完全相同，如多个基准框的长度存在差异，多个基准框的宽度存在差异，且多个基准框的长宽比存在差异。

一个可选的例子，本申请实施方式可以预先设置三种不同的长度，并预先设置三种不同的长宽比，从而利用预先设置的三种不同的长度以及三种不同的长宽比，可以形成九个不同形状的基准框；例如，预先设置的三种不同的长度分别为 64、128 以及 256（单位为像素），预先设置的三种不同的长宽比分别为 0.5、1.0 以及 2.0，则由上述三种不同的长度以及三种不同的长宽比所形成的九个基准框可以表示为：64×32、64×64、64×128、128×64、128×128、128×256、256×128、256×256 以及 256×512。

另一个可选的例子，本申请实施方式可以预先设置三种不同的宽度，并预先设置三种不同的长宽比，从而利用预先设置的三种不同的宽度以及三种不同的长宽比，可以形成九个不同形状的基准框。

再一个可选的例子，本申请实施方式可以针对九个基准框分别设置其各自的长度以及宽度，从而可以明确地确定出九个形状各异的基准框。

本申请中的基准框的数量及各基准框的尺寸可以根据实际经验设置，本申请实施方式不限制基准框的具体数量，也不限制基准框的尺寸。

在一个可选示例中，本申请实施方式预先设置有至少三种类别，如至少两种物体类别以及一个背景类别。例如，预先针对人、狗、火车以及椅子等不同种类的物体设置了 20 种物体类别，并预先针对背景设置了背景类别，总共 21 种类别。类别的数量以及基准框的数量会决定卷积神经网络中的相应卷积层的输出维度，例如，在预先设置了 21 种类别的情况下，RPN 中的用于预测置信度的卷积层的输出维度可以为： $21 \times 9 \times h \times w$ ，其中，21 表示 21 种类别，9 表示一个特征点的 9 个基准框，h 表示特征图的高度，w 表示特征图的宽度。本申请实施例不限制物体类别的数量以及物体类别所针对的物体的具体内容等。

在一个可选示例中，本申请实施方式可以通过卷积层针对置信度的计算来确定基准框所属类别，例如，通过 RPN 中的用于预测置信度的相应卷积层等针对置信度的计算来确定基准框所属类别。

一个可选例子，针对特征图中的第一特征点而言，设定第一特征点对应九个基准框，即第一基准框至第九基准框，且设定预先设置有 21 种类别，即第一物体类别、第二物体类别、……、第二十物体类别以及背景类别，在上述设定的情况下，可以利用卷积神经网络中的一卷积层计算出第一基准框属于第一物体类别的置信度、第一基准框属于第二物体类别的置信度、……第一基准框属于第二十物体类别的置信度以及第一基准框属于背景类别的置信度（总共 21 个置信度），通过比较这 21 个置信度，可以确定出一个最高置信度，从而可以将该最高置信度对应的类别作为第一基准框所属类别；以此类推，本申请实施方式可以确定出第一特征点的第一基准框至第九基准框分别所属类别（总共 9 个类别），进而本申请实施方式可以确定出特征图中的所有特征点的第一基准框至第九基准框分别所属类别（总共 $N \times 9$ 个类别，其中 N 为特征图所包含的特征点的数量）。

在一个可选示例中，本申请中的位置偏移值通常包括：上偏移值、下偏移值、左偏移值以及右偏移值，本申请实施方式可以利用另一卷积层计算出基准框针对每一类别的上偏移值、下偏移值、左偏移值以及右偏移值；其中的上偏移值表示基准框的顶边在垂直方向上的偏移量，其中的下偏移值表示基准框的底边在垂直方向上的偏移量，其中的左偏移值表示基准框的左侧边在水平方向上的偏移量，其中的右偏移值表示基准框的右侧边在水平方向上的偏移量。

例如，针对特征图中的第一特征点而言，设定第一特征点对应九个基准框，即第一基准框至第九基准框，且设定预先设置有 21 种类别，即第一物体类别、第二物体类别、……、第二十物体类别以及背

景类别,在上述设定的情况下,可以利用卷积神经网络中的一卷积层计算出第一基准框针对第一物体类别的上偏移值、下偏移值、左偏移值和右偏移值、第一基准框针对第二物体类别的上偏移值、下偏移值、左偏移值和右偏移值、……、第一基准框针对第二十物体类别的上偏移值、下偏移值、左偏移值和右偏移值以及第一基准框针对背景类别的上偏移值、下偏移值、左偏移值和右偏移值(总共 4×21 个偏移值);以此类推,本申请实施方式可以计算出第一特征点的第一基准框至第九基准框分别针对二十一个类别的上偏移值、下偏移值、左偏移值以及右偏移值(总共 $9 \times 4 \times 21$ 个偏移值),进而本申请实施方式可以确定出特征图中的所有特征点的第一基准框至第九基准框分别针对二十一个类别的上偏移值、下偏移值、左偏移值以及右偏移值(总共 $N \times 9 \times 4 \times 21$ 个偏移值,其中N为特征图所包含的特征点的数量)。

由上述描述可知,本申请实施例中的确定基准框所属类别的过程(即步骤S310)与计算基准框针对至少一个类别的位置偏移值的过程(即步骤S310),两个过程之间并不存在相互依赖的关系,因此,这两个过程完全可以并行执行,从而可以降低两个过程需要先后执行而导致的时间成本增加的现象,有利于提高物体检测的实时性。

S320、针对多个特征点的多个基准框,将该基准框所属类别所对应的位置偏移值作为该基准框的位置偏移值。

在一个可选示例中,本申请实施方式可以从上述确定出的一个基准框所属类别以及上述计算出的该基准框针对至少一个类别的位置偏移值最终确定出该基准框的位置偏移值,即从上计算出的该基准框针对各类别的位置偏移值中挑选出上述确定出的该基准框所属类别对应的位置偏移值,并将该挑选出的位置偏移值作为该基准框的位置偏移值,从而本申请实施方式可以确定出多个基准框的位置偏移值。

一个可选的例子,设定已经利用卷积神经网络中的一卷积层确定出第一特征点的第一基准框属于第一物体类别,且已经利用卷积神经网络中的另一卷积层计算出第一特征点的第一基准框针对第一物体类别的上偏移值、下偏移值、左偏移值和右偏移值、第一特征点的第一基准框针对第二物体类别的上偏移值、下偏移值、左偏移值和右偏移值、……、第一特征点的第一基准框针对第二十物体类别的上偏移值、下偏移值、左偏移值和右偏移值以及第一特征点的第一基准框针对背景类别的上偏移值、下偏移值、左偏移值和右偏移值(总共 4×21 个偏移值),在上述设定的情况下,本申请实施方式可以将已经计算出的第一特征点的第一基准框针对第一物体类别的上偏移值、下偏移值、左偏移值以及右偏移值作为第一特征点的第一基准框的位置偏移值;以此类推,本申请实施方式可以确定出第一特征点的第一基准框至第九基准框的位置偏移值,进而本申请实施方式可以确定出特征图中的所有特征点的第一基准框至第九基准框的位置偏移值(总共 $N \times 9$ 个位置偏移值,即 $N \times 9 \times 4$ 个偏移值,其中N为特征图所包含的特征点的数量)。

S330、根据多个特征点的基准框所属类别、基准框的位置偏移值以及基准框的置信度确定待检测图像中所包含的外接框位置信息以及外接框所属类别。

在一个可选示例中,本申请通常会从多个特征点的基准框中选取部分基准框,并利用选取出的部分基准框的位置偏移值形成外接框,且选取出的部分基准框各自所属类别即为相应的外接框所属类别。

在一个可选示例中,本申请可以利用预设置信度要求来选取部分基准框,且该预设置信度要求可以为置信度阈值(置信度阈值如0.8)等;一个可选的例子,本申请实施方式可以从所有特征点的各基准框中选取置信度超过置信度阈值(例如,大于等于0.8)的基准框,之后,可以利用选取出的各基准框的位置偏移值形成相应特征点的外接框位置信息,并将选取出的各基准框所属类别作为相应外接框所属类别。另外,本申请实施方式也可以利用比例或者外接框数量阈值来选取部分基准框,本申请实施方式不限制从至少一个特征点的基准框中选取部分基准框的具体实现方式。

例如,设定通过上述步骤S310确定出第一特征点的第一基准框的置信度为0.9,第一特征点的第

一基准框属于第一物体类别，且第一特征点的第一基准框的位置偏移值分别为 y_1 、 y_2 、 x_1 以及 x_2 ，在预置信度要求为置信度不低于置信度阈值 0.8 的情况下，可以根据第一特征点对应图像中的区域、第一基准框的长和宽以及 y_1 、 y_2 、 x_1 和 x_2 确定出第一外接框在图像中的位置信息，并确定出该第一外接框所属类别为第一物体类别。

下面以将特征点的基准框作为待定目标为例，对本申请的卷积神经网络的训练过程进行说明。图 4 为本申请实施例提供的用于物体检测的方法中对卷积神经网络进行训练的一个示例的流程图。如图 4 所示，本申请实施方式的训练过程包括：步骤 S400、步骤 S410、步骤 S420、步骤 S430、步骤 S440 以及步骤 S450。下面结合图 5 对图 4 中的各步骤分别进行说明。

在一个可选示例中，图 4 中包括的步骤 S400、步骤 S410、步骤 S420、步骤 S430、步骤 S440 以及步骤 S450 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的训练模块执行。

S400、从样本集合中获取一个图像样本。到步骤 S410。

在一个可选示例中，本申请实施例中的样本集合通常为非空，且通常包括大量的图像样本。样本集合中的图像样本通常均带有标注，如图像样本均标注有多个外接框位置信息以及标准外接框所属类别等；外接框位置信息可以包括：外接框对角线上的两个顶点在图像样本中的二维坐标信息等。图像样本的标注通常是人工预先标注出来的。本申请实施方式可以从样本集合中安装图像样本的排列次序顺序的选取一个图像样本，也可以从样本集合中随机选取一个图像样本。

S410、通过卷积神经网络中的主干网络生成该图像样本的特征图，即图 5 中的特征图 (Feature Map)。到步骤 S420。图 5 为本申请实施例提供的用于物体检测的方法中对卷积神经网络进行训练的一个示例的训练示意图。图 4 中的一个或多个步骤可参照图 5 进行理解。

S420、通过卷积神经网络中的滤波层对上述特征图进行滤波处理 (例如，利用图 5 中的卷积 (Conv) 层实现滤波处理)，从而可以将主干网络输出的特征图转换为新的特征图，且主干网络输出的特征图和新的特征图所包含的特征点的数量是相同的。到步骤 S430、步骤 S440 以及步骤 S450。

S430、针对滤波处理后的特征图中的每一个特征点的每一个基准框：利用卷积神经网络中的一卷积层 (例如，第一卷积层) 计算该特征点的该基准框分别属于至少一个类别的置信度，例如，图 5 中利用卷积层获得的置信度 (cls_score)，根据一个基准框分别属于至少一个类别的置信度确定出该基准框类别，例如，将最高置信度对应的类别作为该基准框所属类别。到步骤 S460。

S440、针对滤波处理后的特征图中的每一个特征点的每一个基准框：利用卷积神经网络中的一卷积层 (例如，第二卷积层) 计算/预测该特征点的基准框针对至少一个类别的位置偏移值，例如，图 5 中利用卷积层获得位置偏移值 (loc_pred)，该预测出的位置偏移值可以包括：上偏移值、下偏移值、左偏移值以及右偏移值。到步骤 S470。

S450、针对滤波处理后的特征图中的一个特征点的一个基准框：可以利用预先设置的辅助层计算该基准框相对于相应的标准外接框的标准位置偏移值，例如，辅助层计算特征点的基准框与该特征点所对应的标准外接框的上偏移值、下偏移值、左偏移值以及右偏移值。另外，可以利用预先设置的辅助层可以将多个特征点的基准框所属类别设置为相应标准外接框所属类别。到步骤 S460 以及步骤 S470。

可选地，针对一个特征点的一个基准框：辅助层可以计算该基准框与至少一个标准外接框的重叠面积，并将重叠面积较大 (例如：最大) 的标准外接框作为该基准框所对应的标准外接框，并将重叠面积较大 (例如：最大) 的标准外接框所属类别作为该基准框所属类别。

由于辅助层可以根据预先设置的基准框配置信息明确的获知基准框的长宽，因此，本申请中的辅助层可以根据特征点以及基准框的长宽确定出该基准框在图像样本中的位置，而由于图像样本中的至少一个标准外接框均是人工预先标注出来的，因此，辅助层可以获知标准外接框在图像样本中的位置信息 (例

如，标准外接框在其对角线方向上的两个顶点的坐标等），从而本申请中的辅助层可以准确快速的计算出该基准框和与其对应的标准外接框之间的位置偏移值（例如，图 5 中利用辅助层获得的锚目标（Anchor Target）C21），该计算出的位置偏移值即为标准位置偏移值，且该标准位置偏移值可以包括：标准上偏移值、标准下偏移值、标准左偏移值和标准右偏移值。

需要说明的是，本申请实施例可以针对多个特征点的基准框中选取出的部分基准框来进行标准位置偏移值的计算，即不是所有的基准框都会参与位置偏移值之间的误差在卷积神经网络中的反向传递过程；在上述情况下，本申请中的辅助层可以根据基准框与标准外接框的重叠面积来选取需要进行标准位置偏移值的计算的基准框，可选地，针对一个基准框而言，辅助层可以将其与各标准外接框的重叠面积最大的标准外接框作为该基准框对应的标准外接框，从而辅助层可以获得基准框对应的标准外接框，然后，辅助层根据各基准框与其对应的标准外接框的重叠面积是否满足预定要求，可以将多个基准框划分为三类，即正样本类（即重叠面积比例大于第一比例阈值的一类）、负样本类（即重叠面积比例小于第二比例阈值的一类）以及忽略样本类（即重叠面积比例介于第一比例阈值和第二比例阈值之间的一类），之后，辅助层可以从正样本类中选取（例如，随机选取）部分基准框，并从负样本类中选取（例如，随机选取）部分基准框，且辅助层从正样本类中选取出的基准框的数量与从负样本类中选取出的基准框的数量通常应满足预先设置的比例，例如 3:1；之后，辅助层可以针对选取出的基准框进行标记，从而使后续的损失层可以根据该标记针对相应的基准框计算位置偏移值之间的误差。

上述重叠面积比例可以为：针对一待定目标而言，如果该待定目标与其重叠面积最大的标准外接框之间的重叠面积与该待定目标和标准外接框共同覆盖的面积之比，例如，设定待定目标 A 的面积为 3，而待定目标 A 对应的标准外接框 B 的面积为 4，且待定目标 A 与标准外接框 B 之间的重叠面积为 2，则重叠面积比例为： $2/(3+4-2)=2/5$ 。本申请实施例中的第一比例阈值大于第二比例阈值，在一个可选的例子中，第一比例阈值可以为 0.7，而第二比例阈值可以为 0.3。本申请实施例不限制第一比例阈值和第二比例阈值的具体取值。

S460、针对特征图中的特征点：利用一损失层计算该特征点的至少一个基准框属于相应标准外接框所属类别的置信度与标准置信度之间的误差。一个可选的例子，步骤 S430 中的卷积层计算出的特征点的至少一个基准框分别属于至少一个类别的置信度均被传输至图 5 中的交叉熵损失（Softmax With Loss）层，且步骤 S450 中的辅助层针对每一个特征点的每一个基准框所设置的所属类别也被传输至图 5 中的 Softmax With Loss 层，从而 Softmax With Loss 层可以根据辅助层传输来的信息以及卷积层传输来的信息计算出每一个特征点的每一个基准框属于相应标准外接框所属类别的置信度与标准置信度之间的误差，且该误差会通过 Softmax With Loss 层反向传递，直到主干网络。

S470、针对特征图中的一个特征点的一个基准框：利用一损失层（例如，平滑损失（smooth loss）层）计算步骤 S440 预测出的位置偏移值与步骤 S450 计算出的相应的标准位置偏移值之间的误差。可选地，smooth loss 层可以根据辅助层的标记对相应的预测出的位置偏移值与计算出的相应的标准位置偏移值之间的误差，且该误差会通过 smooth loss 层反向传递，直到主干网络（完成一次迭代过程）。到步骤 S480。

S480、误差是否符合要求以及是否继续训练，如果误差符合要求且不需要继续训练，则到步骤 S490；而如果需要继续训练（不论误差是否符合要求），则返回步骤 S400；如果误差不符合要求且不需要继续训练，则到步骤 S481。

本申请实施例中的的是否需要继续训练，可以通过迭代次数（即已经从图像样本集合中获取到的图像样本的数量）来判断，例如，在迭代次数已经达到预定数量的情况下，确定不需要继续训练，而在迭代次数还未达到预定数量的情况下，确定需要继续训练。本申请实施例中的误差是否符合要求可以通过误

差是否属于预定范围来判断,例如,在当前误差属于预定范围的情况下,确定当前误差可以被接受,卷积神经网络成功训练完成,而在当前误差不属于预定范围的情况下,确定当前误差不能够被接受,卷积神经网络训练失败。

S481、未成功完成卷积神经网络的训练,本次训练过程结束。

可选地,如果样本集中的图像样本已经均被用于训练或者读取的样本数量已经达到预定数量,然而,计算出的误差仍不符合要求,则虽然本次训练过程结束了,但是,卷积神经网络并没有训练成功,后续可以继续对卷积神经网络进行训练。

S490、成功完成卷积神经网络的训练,本次训练过程结束。

本申请可以通过交替执行误差的计算过程以及误差的反向传递过程,对卷积神经网络进行训练,直到计算出的误差符合要求,从而成功完成卷积神经网络的训练过程。

还有,本申请可以对成功训练后的卷积神经网络进行测试。对训练后的卷积神经网络进行测试的过程与上述用于物体检测的方法中的步骤相同,只是可以通过计算平均检测准确率(Average Precision, AP)以及平均召回率(Recall,即所有外接框中与标准外接框的重叠率超过预定阈值的百分比)等参数来衡量本申请的卷积神经网络的物体检测性能。

可选地,本申请可以在利用一卷积层确定出多个特征点的基准框所属类别之后,对基准框进行非极大值抑制(non maximum suppression, NMS)处理,即从基准框中去除面积重叠率超过一定阈值的基准框,其中的面积重叠率可以为:两个基准框的相交部分的面积÷(两个基准框的面积之和-两个基准框的相交部分面积),且上述阈值可以为0.3等;之后,本申请可以从去除了面积重叠率超过一定阈值的基准框之后的基准框中选取置信度最高的300个基准框计算AP以及Recall等,以获知训练后的卷积神经网络的物体检测性能。

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。

图6为本申请实施例提供的用于物体检测的装置的一个结构示意图。该实施方式的装置可用于实现本申请上述各方法实施方式。如图6所示,该实施方式的装置主要包括:获取待定目标模块600、预测类别和位置偏移值模块610、确定位置偏移模块620以及确定检测结果模块630。可选的,该装置还可以包括:训练模块(图6中未示出)。

获取待定目标模块600用于获取待检测图像中的多个待定目标,可选地,获取待定目标模块600可以用于获取待检测图像中的基于至少一个预设区域大小而形成的多个待定目标。获取待定目标模块600所使用的至少一个预设区域大小的长度存在差异,至少一个预设区域大小的宽度存在差异,至少一个预设区域大小的长宽比存在差异;例如,预设区域大小包括:三种不同的长度和三种不同的长宽比所形成的九个预设区域大小;或三种不同的宽度和三种不同的长宽比所形成的九个预设区域大小;或基于三种不同的长度和宽度所形成的九个预设区域大小。

在一个实施方式中,获取待定目标模块600可以包括第一获取待定目标子模块,该第一获取待定目标子模块主要用于获取待检测图像的特征图,并针对特征图中的一个特征点而言,根据基准框配置信息形成该特征点的基准框,且该特征点的一个基准框被作为一个待定目标;获得特征图中多个特征点对应的多个待定目标。第一获取待定目标子模块可以通过将待检测图像提供给卷积神经网络中的主干网络,并将主干网络输出的特征图提供给卷积神经网络中的滤波层,以由滤波层根据预设滑窗对特征图进行滤波处理,且第一获取待定目标子模块可以将滤波处理后的特征图作为待检测图像的特征图。

在一个实施方式中,获取待定目标模块600可以包括第二获取待定目标子模块,该第二获取待定目

标子模块主要用于获取待检测图像的特征图，并基于基准框配置信息对特征图进行池化处理，以形成多个新的特征图，将多个新的特征图被作为多个待定目标。

获取待定目标模块 600（如第一获取待定目标子模块以及第二获取待定目标子模块）执行的操作及预设区域大小等可以参见上述方法实施方式中步骤 S200 和步骤 S300 中的相关描述，在此不再重复说明。

预测类别和位置偏移值模块 610 用于确定多个待定目标分别针对至少一个类别的置信度，根据置信度确定多个待定目标所属类别；并且，确定多个待定目标各自所属类别所对应的位置偏移值，例如，预测类别和位置偏移值模块 610 利用卷积神经网络中的一卷积层确定多个待定目标分别针对各类别的置信度，并根据上述置信度确定多个待定目标所属类别，并且预测类别和位置偏移值模块 610 利用卷积神经网络中的另一卷积层确定多个待定目标分别针对每一类别的位置偏移值（例如，上偏移值、下偏移值、左偏移值以及右偏移值）。

在一个可选示例中，本申请实施例中的类别可以包括：至少两种物体类别以及一个背景类别，例如，二十种物体类别以及一个背景类别。可选地，针对一个待定目标，预测类别和位置偏移值模块 610 可以利用卷积神经网络中的一卷积层计算该待定目标分别属于至少一个类别的置信度，并将最高置信度对应的类别作为该待定目标所属类别。预测类别和位置偏移值模块 610 执行的具体操作可以参见上述方法实施方式中步骤 S210 和步骤 S310 中的相关描述，在此不再重复说明。

确定位置偏移模块 620 用于将多个待定目标各自所属类别所对应的位置偏移值作为多个待定目标的位置偏移值。确定位置偏移模块 620 执行的具体操作可以参见上述方法实施方式中步骤 S220 和步骤 S320 中的相关描述，在此不再重复说明。

确定检测结果模块 630 用于根据多个待定目标所属类别、多个待定目标的位置偏移值以及多个待定目标所属类别的置信度确定待检测图像中至少一个待定目标的位置信息和所属类别。一个可选例子，确定检测结果模块 630 可以从所有待定目标中选取置信度满足预设置信度要求的待定目标，并根据选取出的多个待定目标的位置偏移值形成待检测图像中的待定目标的位置信息（如外接框位置信息），将选取出的多个待定目标所属类别作为相应待定目标（如外接框）所属类别。确定检测结果模块 630 执行的具体操作可以参见上述方法实施方式中步骤 S230 和步骤 S330 中的相关描述，在此不再重复说明。

训练模块主要用于训练本申请中的卷积神经网络。可选地，训练模块可以从图像样本集合中获取标注有至少一个标准外接框位置信息和标准外接框所属类别的图像样本，并获取图像样本中的基于至少一个预设区域大小而形成的多个待定目标；之后，训练模块利用卷积神经网络中的一卷积层确定多个待定目标分别针对至少一个类别的置信度，并根据置信度确定多个待定目标所属类别，训练模块利用卷积神经网络中的一卷积层确定多个待定目标分别针对每一类别的位置偏移值，并将多个待定目标各自所属类别所对应的位置偏移值作为多个待定目标的位置偏移值；训练模块计算多个待定目标相对于相应的标准外接框的标准位置偏移值（例如，训练模块确定多个待定目标各自对应的重叠面积最大的标准外接框，并分别计算多个待定目标相对于重叠面积最大的标准外接框的标准位置偏移值），训练模块利用卷积神经网络中的一损失层计算所有或者部分待定目标针对相应的标准外接框的类别的位置偏移值与相应的标准位置偏移值的误差；训练模块利用卷积神经网络中的一损失层计算所有或者部分待定目标针对相应标准外接框所属类别的置信度与标准外接框所属类别之间的误差，并将误差在卷积神经网络中反向传播；训练模块控制是否继续训练，例如，训练模块在从图像样本集合中获取的图像样本的数量达到预定数量，且在误差属于预定范围的情况下，训练模块确定卷积神经网络成功训练完成。

上述部分待定目标可以包括但不限于：正待定目标和负待定目标，例如，训练模块从所有待定目标中的所有正待定目标中选取正待定目标，并从所有待定目标中的所有负待定目标中选取负待定目标，

且训练模块选取出的正待定目标和选取出的负待定目标满足预定比例。训练模块执行的具体操作、正待定目标以及负待定目标等内容可以参见上述针对图 4 和图 5 的相关描述，在此不再重复说明。

本申请实施方式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 7，其示出了适于用来实现本申请实施方式的终端设备或服务器的电子设备 700 的结构示意图：如图 7 所示，电子设备 700 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）701，和/或 一个或多个图像处理器（GPU）713 等，处理器可以根据存储在只读存储器（ROM）702 中的可执行指令或者从存储部分 708 加载到随机访问存储器（RAM）703 中的可执行指令而执行各种适当的动作和处理。通信部 712 可以包括但不限于网卡，所述网卡可以包括但不限于 IB（Infiniband）网卡。处理器可与只读存储器 702 和/或随机访问存储器 703 中通信以执行可执行指令，通过总线 704 与通信部 712 相连、并经通信部 712 与其他目标设备通信，从而完成本申请实施方式提供的任一项方法对应的操作。

此外，在 RAM 703 中，还可以存储有装置操作所需的各种程序以及数据。CPU701、ROM702 及 RAM703 通过总线 704 彼此相连。在有 RAM703 的情况下，ROM702 为可选模块。RAM703 存储可执行指令，或在运行时向 ROM702 中写入可执行指令，可执行指令使 CPU 701 执行上述通信方法对应的操作。输入/输出（I/O）接口 705 也连接至总线 704。通信部 712 可集成设置，也可以设置为具有多个子模块（如多个 IB 网卡），并在总线链接上。

以下部件连接至 I/O 接口 705：包括键盘、鼠标等的输入部分 706；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 707；包括硬盘等的存储部分 708；以及包括诸如 LAN 卡、调制解调器等网络接口卡的通信部分 709。通信部分 709 经由诸如因特网的网络执行通信处理。驱动器 710 也需要连接至 I/O 接口 705。可拆卸介质 711，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 710 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 708。

需要说明的，如图 7 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 7 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 GPU713 和 CPU701 可分离设置或者可将 GPU713 集成在 CPU701 上，通信部可分离设置，也可集成设置在 CPU701 或 GPU713 上，等等。这些可替换的实施方式均落入本申请公开的保护范围。

特别地，根据本申请的实施方式，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，本申请公开的实施方式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方式提供的方法步骤对应的指令。在这样的实施方式中，计算机程序可以通过通信部分 709 从网络上被下载和安装，和/或从可拆卸介质 711 被安装。在该计算机程序被中央处理单元（CPU）701 执行时，执行本申请的方法中限定的上述功能。

可能以许多方式来实现本申请的方法和装置、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申请的方法和装置、设备。用于方法的步骤的上述顺序仅是为了进行说明，本申请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，在一些实施方式中，还可将本申请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申请的方法的机器可读指令。因而，本申请还覆盖存储用于执行根据本申请的方法的程序的记录介质，例如，图 8 所示的计算机可读存储介质 800。

本说明书中各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之

处，各个实施例之间相同或相似的部分相互参见即可。对于系统实施例而言，由于其与方法实施例基本对应，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

可能以许多方式来实现本申请的方法和装置。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申请的方法和装置。用于所述方法的步骤的上述顺序仅是为了进行说明，本申请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，还可将本申请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申请的方法的机器可读指令。因而，本申请还覆盖存储用于执行根据本申请的方法的程序的记录介质。

本申请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申请限于所公开的形式。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申请的原理和实际应用，并且使本领域的普通技术人员能够理解本申请从而设计适于特定用途的带有各种修改的各种实施例。