

物体检测方法、神经网络的训练方法、装置和电子设备

本申请要求在 2017 年 2 月 23 日提交中国专利局、申请号为 CN201710100676.1、发明名称为“物体检测方法、神经网络的训练方法、装置和电子设备”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

本申请涉及图像处理技术，尤其涉及一种物体检测方法和装置、神经网络的训练方法和装置、电子设备。

背景技术

目标区域框检测的目的是从图像检测出若干可能存在物体的矩形框。在目前常规的使用卷积神经网络执行检测的技术中，通过卷积神经网络中的池化层逐渐减小特征图的大小，从而最终确定可能存在物体的矩形框，这种网络结构被称作“缩小网络”(zoom-out structure)。

发明内容

本申请提供一种基于图像进行目标区域框检测的技术。

根据本申请实施例的一方面，提供一种物体检测方法，包括：通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图；其中，所述深度卷积神经网络包括第一子网和第二子网，所述第一子网具有至少一个下采样层，所述第二子网具有至少一个上采样层；所述融合特征图通过第一特征图和第二特征图得到，所述第一特征图从第一子网获取得到，所述第二特征图从第二子网获取得到；根据所述多个融合特征图获取目标区域框数据。

在本申请的一种实现方式中，所述第二子网设置在所述第一子网的末端，所述第一子网具有多个第一卷积层和所述至少一个下采样层，所述下采样层设置在所述多个第一卷积层之间，所述第二子网具有多个第二卷积层和所述至少一个上采样层，所述上采样层设置在所述多个第二卷积层之间，所述第一卷积层和所述第二卷积层对称设置，所述至少一个下采样层和所述至少一个上采样层分别对称地设置。

在本申请的一种实现方式中，在至少一个所述第一卷积层设有用于输出所述第一特征图的第一输出分支，在第二卷积层设有用于输出所述第二特征图的第二输出分支。

在本申请的一种实现方式中，所述第二子网还具有多个第三卷积层，所述第三卷积层的输入包括所述第一输出分支和所述第二输出分支；所述预测获取多个融合特征图包括：从所述多个第三卷积层的输出端分别获取所述融合特征图。

在本申请的一种实现方式中，所述融合特征图中的至少一个点具有与多个物体探测框对应的框融合检测数据以及预测准确信息，所述根据所述多个融合特征图获取目标区域框数据包括：根据至少一个所述融合特征图中的框融合检测数据以及预测准确信息分别获取与所述融合特征图各自对应的目标区域框数据。

在本申请的一种实现方式中，所述根据所述多个融合特征图获取目标区域框数据包括：分别获取所述融合特征图各自对应的初选目标区域框数据；迭代地执行以下物体区域框回归操作，直到所述迭代满足迭代终止条件为止：通过调整所述融合特征图，从经过调整的融合特征图获取新的初选目标区域框数据；将经过所述迭代得到的所述初选目标区域框数据作为所述待处理的图像中的目标区域框数据。

在本申请的一种实现方式中，所述深度卷积神经网络还包括第三子网，所述第三子网具有多组第四卷积层和多个池化层，所述多组第四卷积层分别与所述第三卷积层对应，所述多个池化层分别与所述多组第四卷积层对应，并且每个所述池化层的输入包括所述经过调整的融合特征图和所述初选目标区域框的数据。

在本申请的一种实现方式中，所述物体区域框回归操作包括：通过所述第四卷积层分别对当前的所述融合特征图进行卷积，获取调整融合特征图；根据当前的初选目标区域框数据，通过所述池化层对所述调整融合特征图进行区域池化，获取新的融合特征图；从所述新的融合特征图获取所述新的初选目标区域框数据。

在本申请的一种实现方式中,所述第三子网还具有设置在所述池化层输出端的第五卷积层,所述从所述新的融合特征图获取所述新的初选目标区域框数据包括:通过所述第五卷积层对所述新的融合特征图进行规范化卷积,从经过规范化卷积的融合特征图获取所述新的初选目标区域框数据。

在本申请的一种实现方式中,所述第一子网和所述第二子网均为认知一样本归一化(Inception-BN)网络结构,所述第三子网为残差网络(ResNet)结构。

根据本申请的第二方面,提供一种神经网络的训练方法,包括:将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络,检测获取多个融合特征图,所述深度卷积神经网络包括第一子网和第二子网,所述第一子网具有至少一个下采样层,所述第二子网具有至少一个上采样层;所述融合特征图通过第一特征图和第二特征图得到,所述第一特征图从第一子网获取得到,所述第二特征图从第二子网获取得到;根据所述多个融合特征图获取所述样本图像的目标区域框数据;根据获取到的所述样本图像的目标区域框数据以及所述目标区域框标注信息确定物体框检测的第一差异数据;根据所述第一差异数据调整所述深度卷积神经网络的网络参数。

在本申请的一种实现方式中,所述第二子网设置在所述第一子网的末端,所述第一子网具有多个第一卷积层和所述至少一个下采样层,所述下采样层设置在所述多个第一卷积层之间,所述第二子网具有多个第二卷积层和所述至少一个上采样层,所述上采样层设置在所述多个第二卷积层之间,所述第一卷积层和所述第二卷积层对称设置,所述至少一个下采样层和所述至少一个上采样层分别对称地设置。

在本申请的一种实现方式中,在至少一个所述第一卷积层设有用于输出所述第一特征图的第一输出分支,在第二卷积层设有用于输出所述第二特征图的第二输出分支。

在本申请的一种实现方式中,所述第二子网还具有多个第三卷积层,所述第三卷积层的输入包括所述第一输出分支和所述第二输出分支;所述检测获取多个融合特征图包括:从所述多个第三卷积层的输出端分别获取所述融合特征图。

在本申请的一种实现方式中,所述融合特征图中的至少一个点具有与多个物体探测框对应的框融合检测数据以及预测准确信息。

在本申请的一种实现方式中,所述深度卷积神经网络还包括第三子网,所述第三子网具有多组第四卷积层和多个池化层,所述多组第四卷积层分别与所述第三卷积层对应,所述多个池化层分别与所述多组第四卷积层对应,并且每个所述池化层的输入包括所述经过调整的融合特征图和所述初选目标区域框的数据。

在本申请的一种实现方式中,所述方法还包括:迭代地执行以下目标区域框回归训练操作,直到所述迭代满足迭代终止条件为止:通过所述第四卷积层分别对当前的所述融合特征图进行卷积,获取调整融合特征图;根据当前的初选目标区域框数据,通过所述池化层对所述调整融合特征图进行区域池化,获取新的融合特征图;从所述新的融合特征图获取所述新的初选目标区域框数据;根据未经过调整的初选目标区域框数据和新的初选目标区域框数据之间的框回归数据、新的初选目标区域框数据和相应的目标区域框标注信息确定物体框检测的第二差异数据;根据所述第二差异数据调整所述第三子网的网络参数。

在本申请的一种实现方式中,所述第三子网还具有设置在所述池化层输出端的第五卷积层,所述从所述新的融合特征图获取所述新的初选目标区域框数据包括:通过所述第五卷积层对所述新的融合特征图进行规范化卷积,从经过规范化卷积的融合特征图获取所述新的初选目标区域框数据。

在本申请的一种实现方式中,在将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络,检测获取多个融合特征图之前,所述方法还包括:缩放所述样本图像,使得至少一个物体区域框的真值被物体探测框覆盖。

在本申请的一种实现方式中,所述样本图像的目标区域框标注信息包括正样本区域框的标注信息和负样本区域框的标注信息;所述正样本区域框与物体区域框的真值的重叠率不低于第一重叠比率值,所述负样本区域框与物体区域框的真值的重叠率不高于第二重叠比率值,所述第一重叠比率值大于所述第二重叠比率值。

在本申请的一种实现方式中,所述样本图像的目标区域框标注信息还包括中性样本区域框的标注信息,所述中性样本区域框与物体区域框的真值的重叠率在所述第一重叠比率值和所述第二重叠比率值之间。

在本申请的一种实现方式中,在全部所述样本图像当中,标注的正样本区域框的总和在所述正样本区域框、负样本区域框以及中性样本区域框的框总数中的占比不小于预定的第一比值,所述第一比值大于50%;

标注的负样本区域框的总和在框总数中的占比不大于预定的第二比值；标注的中性样本区域框的总和在框总数中的占比不大于预定的第三比值，所述第三比例不大于第一比值和第二比值之和的一半。

在本申请的一种实现方式中，所述第一子网和所述第二子网均为认知一样本归一化网络结构，所述第三子网为残差网络结构。

根据本申请的第三方面，提供一种物体检测装置，包括：融合特征图预测模块，用于通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图；其中，所述深度卷积神经网络包括第一子网和第二子网，所述第一子网具有至少一个下采样层，所述第二子网具有至少一个上采样层；所述融合特征图通过第一特征图和第二特征图得到，所述第一特征图从第一子网获取得到，所述第二特征图从第二子网获取得到；目标区域框预测模块，用于根据所述融合特征图预测模块获取的多个融合特征图获取目标区域框数据。

在本申请的一种实现方式中，所述第二子网设置在所述第一子网的末端，所述第一子网具有多个第一卷积层和所述至少一个下采样层，所述下采样层设置在所述多个第一卷积层之间，所述第二子网具有多个第二卷积层和所述至少一个上采样层，所述上采样层设置在所述多个第二卷积层之间，所述第一卷积层和所述第二卷积层对称设置，所述至少一个下采样层和所述至少一个上采样层分别对称地设置。

在本申请的一种实现方式中，在至少一个所述第一卷积层设有用于输出所述第一特征图的第一输出分支，在第二卷积层设有用于输出所述第二特征图的第二输出分支。

在本申请的一种实现方式中，所述第二子网还具有多个第三卷积层，所述第三卷积层的输入包括所述第一输出分支和所述第二输出分支；所述融合特征图预测模块用于从所述多个第三卷积层的输出端分别获取所述融合特征图。

在本申请的一种实现方式中，所述融合特征图中的至少一个点具有与多个物体探测框对应的框融合检测数据以及预测准确信息，所述目标区域框预测模块用于根据至少一个所述融合特征图中的框融合检测数据以及预测准确信息分别获取与所述融合特征图各自对应的目标区域框数据。

在本申请的一种实现方式中，所述目标区域框预测模块用于：分别获取所述融合特征图各自对应的初选目标区域框数据；迭代地执行以下物体区域框回归操作，直到所述迭代满足迭代终止条件为止：通过调整所述融合特征图，从经过调整的融合特征图获取新的初选目标区域框数据；将经过所述迭代得到的所述初选目标区域框数据作为所述待处理的图像中的目标区域框数据。

在本申请的一种实现方式中，所述深度卷积神经网络还包括第三子网，所述第三子网具有多组第四卷积层和多个池化层，所述多组第四卷积层分别与所述第三卷积层对应，所述多个池化层分别与所述多组第四卷积层对应，并且每个所述池化层的输入包括所述经过调整的融合特征图和所述初选目标区域框的数据。

在本申请的一种实现方式中，所述目标区域框预测模块包括：框调整单元，用于通过所述第四卷积层分别对当前的所述融合特征图进行卷积，获取调整融合特征图；区域池化单元，用于根据当前的初选目标区域框数据，通过所述池化层对所述调整融合特征图进行区域池化，获取新的融合特征图；初选框获取单元，用于从所述新的融合特征图获取所述新的初选目标区域框数据。

在本申请的一种实现方式中，所述第三子网还具有设置在所述池化层输出端的第五卷积层，所述初选框获取单元用于通过所述第五卷积层对所述新的融合特征图进行规范化卷积，并且从经过规范化卷积的融合特征图获取所述新的初选目标区域框数据。

在本申请的一种实现方式中，所述第一子网和所述第二子网均为认知一样本归一化（Inception-BN）网络结构，所述第三子网为残差网络（ResNet）结构。

根据本申请的第四方面，提供一种神经网络的训练装置，包括：融合特征图检测模块，用于将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络，检测获取多个融合特征图，所述深度卷积神经网络包括第一子网和第二子网，所述第一子网具有至少一个下采样层，所述第二子网具有至少一个上采样层；所述融合特征图通过第一特征图和第二特征图得到，所述第一特征图从第一子网获取得到，所述第二特征图从第二子网获取得到；目标区域框检测模块，用于根据所述多个融合特征图获取所述样本图像的目标区域框数据；第一差异获取模块，用于根据获取到的所述样本图像的目标区域框数据以及所述目标区域框标注信息确定物体框检测的第一差异数据；第一网络训练模块，用于根据所述第一差异数据调整所述深度卷积神经网络的网络参数。

在本申请的一种实现方式中，所述第二子网设置在所述第一子网的末端，所述第一子网具有多个第一卷积层和所述至少一个下采样层，所述下采样层设置在所述多个第一卷积层之间，所述第二子网具有多个第二卷积层和所述至少一个上采样层，所述上采样层设置在所述多个第二卷积层之间，所述第一卷积层和所述第二卷积层对称设置，所述至少一个下采样层和所述至少一个上采样层分别对称地设置。

在本申请的一种实现方式中，在至少一个所述第一卷积层设有用于输出所述第一特征图的第一输出分支，在所述第二卷积层设有用于输出所述第二特征图的第二输出分支。

在本申请的一种实现方式中，所述第二子网还具有多个第三卷积层，所述第三卷积层的输入包括所述第一输出分支和所述第二输出分支；所述融合特征图检测模块用于从所述多个第三卷积层的输出端分别获取所述融合特征图。

在本申请的一种实现方式中，所述融合特征图中的至少一个点具有与多个物体探测框对应的框融合检测数据以及预测准确信息。

在本申请的一种实现方式中，所述深度卷积神经网络还包括第三子网，所述第三子网具有多组第四卷积层和多个池化层，所述多组第四卷积层分别与所述第三卷积层对应，所述多个池化层分别与所述多组第四卷积层对应，并且每个所述池化层的输入包括所述经过调整的融合特征图和所述初选目标区域框的数据。

在本申请的一种实现方式中，所述装置还包括：框回归迭代训练模块，用于迭代地执行以下目标区域框回归训练操作，直到所述迭代满足迭代终止条件为止：通过所述第四卷积层分别对当前的所述融合特征图进行卷积，获取调整融合特征图；根据当前的初选目标区域框数据，通过所述池化层对所述调整融合特征图进行区域池化，获取新的融合特征图；从所述新的融合特征图获取所述新的初选目标区域框数据；根据未经调整的初选目标区域框数据和新的初选目标区域框数据之间的框回归数据、新的初选目标区域框数据和相应的目标区域框标注信息确定物体框检测的第二差异数据；根据所述第二差异数据调整所述第三子网的网络参数。

在本申请的一种实现方式中，所述第三子网还具有设置在所述池化层输出端的第五卷积层，所述框回归迭代训练模块用于通过所述第五卷积层对所述新的融合特征图进行规范化卷积，并且从经过规范化卷积的融合特征图获取所述新的初选目标区域框数据。

在本申请的一种实现方式中，所述装置还包括：预处理模块，用于在将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络，检测获取多个融合特征图之前，缩放所述样本图像，使得至少一个物体区域框的真值被物体探测框覆盖。

在本申请的一种实现方式中，所述样本图像的目标区域框标注信息包括正样本区域框的标注信息和负样本区域框的标注信息；所述正样本区域框与物体区域框的真值的重叠率不低于第一重叠比率值，所述负样本区域框与物体区域框的真值的重叠率不高于第二重叠比率值，所述第一重叠比率值大于所述第二重叠比率值。

在本申请的一种实现方式中，所述样本图像的目标区域框标注信息还包括中性样本区域框的标注信息，所述中性样本区域框与物体区域框的真值的重叠率在所述第一重叠比率值和所述第二重叠比率值之间。

在本申请的一种实现方式中，在全部所述样本图像当中，标注的正样本区域框的总和在所述正样本区域框、负样本区域框以及中性样本区域框的框总数中的占比不小于预定的第一比值，所述第一比值大于 50%；标注的负样本区域框的总和在框总数中的占比不大于预定的第二比值；标注的中性样本区域框的总和在框总数中的占比不大于预定的第三比值，所述第三比例不大于第一比值和第二比值之和的一半。

在本申请的一种实现方式中，所述第一子网和所述第二子网均为认知一样本归一化网络结构，所述第三子网为残差网络结构。

根据本申请的第五方面，提供了一种电子设备，包括：

处理器和存储器；

所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行本申请任一实施例所述的物体检测方法对应的操作；或者，所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行本申请任一实施例所述的神经网络的训练方法对应的操作。

根据本申请的第六方面，提供了另一种电子设备，包括：

处理器和本申请任一实施例所述的物体检测装置；在处理器运行所述物体检测装置时，本申请任一实

施例所述的物体检测装置中的单元被运行；或者

处理器和本申请任一实施例所述的神经网络的训练装置；在处理器运行所述神经网络的训练装置时，本申请任一实施例所述的神经网络的训练装置中的单元被运行。

根据本申请的第七方面，提供了一种计算机程序，包括计算机可读代码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本申请任一实施例所述的物体检测方法中各步骤的指令；或者

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本申请任一实施例所述的神经网络的训练方法中各步骤的指令。

根据本申请地第八方面，提供了一种计算机可读存储介质，用于存储计算机可读取的指令，所述指令被执行时实现本申请任一实施例所述的物体检测方法中各步骤的操作、或者本申请任一实施例所述的神经网络的训练方法中各步骤的操作。

根据本申请提供的物体检测方案、神经网络的训练技术方案，通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图，其中，从具有至少一个下采样层的第一子网获取多个第一特征图，从具有至少一个上采样层的第二子网获取多个第二特征图，分别由多个第一特征图和多个第二特征图融合得到融合特征图。此后，再根据所述多个融合特征图获取目标区域框数据。由于这些融合特征图较好地表征了图像中高层的语义特征（如：布局、前背景信息）和低层的细节特征（如：小物体信息），因此根据这些融合特征图能够有效地提取到图像中包含的大小物体的目标区域框数据，从而提高物体检测的准确性和鲁棒性。

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是示出根据本申请一实施例的物体检测方法的流程图；

图2是示出根据本申请另一实施例的物体检测方法的流程图；

图3示出根据本申请实施例的深度卷积神经网络的一种示例性结构；

图4是示出根据本申请又一实施例的物体检测方法的流程图；

图5是示出根据本申请一实施例的神经网络的训练方法的流程图；

图6是示出根据本申请另一实施例的神经网络的训练方法的流程图；

图7是示出根据本申请又一实施例的神经网络的训练方法的流程图；

图8是示出根据本申请一实施例的物体检测装置的结构框图；

图9是示出根据本申请另一实施例的物体检测装置的结构框图；

图10是示出根据本申请一实施例的神经网络的训练装置的结构框图；

图11是示出根据本申请另一实施例的神经网络的训练装置的结构框图；

图12是示出根据本申请一实施例的第一电子设备的结构示意图；

图13是示出根据本申请另一实施例的第二电子设备的结构示意图。

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。

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，程序模块可以包括但不限于例程、程序、目标程序、组件、逻辑、数据结构，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。

通过现有技术提供的网络结构执行目标区域框检测，不能够有效地利用从卷积神经网络中的高层得到的特征图中的信息协助处理网络底层的信息，使得从网络获取到的特征数据不具有足够的代表性和鲁棒性，不利于小物体的检测。

下面结合图 1-图 13 对本公开提供的物体检测技术方案进行说明。本公开提供的任一种聚类技术方案可由软件或者硬件或者软硬结合的方式进行示例。例如，本公开提供的聚类技术方案可由某一电子设备实施或者由某一处理器实施，本公开并不限制，所述电子设备可包括但不限于终端或服务器，所述处理器可包括但不限于 CPU 或 GPU。以下不再赘述。

图 1 是示出根据本申请一实施例的物体检测方法的流程图。

参照图 1，本实施例的物体检测方法包括一下步骤：

步骤 S110，通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图。

在一个可选示例中，步骤 S110 可以由处理器调用存储器存储的指令执行或者由被处理器运行的融合特征图预测模块 810 执行。

其中，深度卷积神经网络包括第一子网和第二子网，第一子网具有至少一个下采样层，第二子网具有至少一个上采样层。融合特征图通过第一特征图和第二特征图得到，第一特征图从第一子网获取得到，第二特征图从第二子网获取得到。

本公开上述实施例中待处理的图像是拍摄有一个或多个物体对象的照片或视频帧图像。该图像应满足一定的分辨率要求，至少通过肉眼能够辨别出拍摄到的物体对象。

用于目标区域框检测的深度卷积神经网络中的第一子网通过对待处理的图像进行卷积、池化，可在第一子网不同深度的多个卷积层获取该图像的第一特征图，这些第一特征图表征不同大小程度的区域框的特征。在设置有至少一个下采样层的第一子网中，在较浅的卷积层获得的第一特征图能够较好地表达图像的细节，较浅的卷积层通常指深度卷积神经网络中位置靠前的卷积层，但是难以区分前景和背景；而在较深的卷积层获得的第一特征图能够较好地提取图像中的整体语义特征，较深的卷积层通常指深度卷积神经网络中位置靠后的卷积层，但是将损失图像的细节信息，如小物体信息。

具有至少一个上采样层的第二子网进一步对从第一子网末端获取到的第一特征图执行相反的处理，即反卷积、上采样和池化操作，将从第一子网末端获取到的第一特征图逐步放大，在第二子网不同深度的多个卷积层获取与前述第一特征图相应的第二特征图。由于第二特征图均由经过卷积、下采样的第一特征图进行反卷积和上采样，在此过程中，高层语义特征被逐步反卷积并与低层细节特征结合，可协助识别小物体（小物体的区域框）。

由此，通过第一子网和第二子网执行的图像处理途径形成一个沙漏形的结构，从第一子网的第一卷积层生成的第一特征图通过下采样逐步变小；第一子网末端生成的第一特征图通过第二子网的第二卷积层和上采样层被逐步放大。

在此基础上，将至少一个第一特征图与相应的第二特征图进行融合，得到多个融合特征图，这些融合特征图可较好地表征图像中高层的语义特征和低层的细节特征，以用于识别不同大小的物体区域框。

步骤 S120，根据多个融合特征图获取目标区域框数据。

在一个可选示例中，步骤 S120 可以由处理器调用存储器存储的指令执行或者由被处理器运行的目标区

域框预测模块 820 执行。

具体地，可从至少一个融合特征图提取目标区域框数据，再将从至少一个融合特征图提取的目标区域框数据整合，作为从待处理的图像检测到的目标区域框数据。

根据本申请实施例的物体检测方法，通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图，其中，从具有至少一个下采样层的第一子网获取多个第一特征图，从具有至少一个上采样层的第二子网获取多个第二特征图，分别由多个第一特征图和多个第二特征图融合得到融合特征图。此后，再根据所述多个融合特征图获取目标区域框数据。由于这些融合特征图较好地表征了图像中高层的语义特征（如：布局、前背景信息）和低层的细节特征（如：小物体信息），因此根据这些融合特征图能够有效地提取到图像中包含的大小物体的目标区域框数据，从而提高物体检测的准确性和鲁棒性。

图 2 是示出根据本申请另一实施例的物体检测方法的流程图。

参照图 2，在步骤 S210，通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图。

在一个可选示例中，步骤 S210 可以由处理器调用存储器存储的指令执行或者由被处理器运行的融合特征图预测模块 810 执行。

具体地，在该深度卷积神经网络中，第一子网具有多个第一卷积层和至少一个下采样层，下采样层设置在多个第一卷积层之间；第二子网具有多个第二卷积层和至少一个上采样层，上采样层设置在多个第二卷积层之间。第二子网设置在第一子网的末端，第一卷积层和第二卷积层对称设置，至少一个下采样层和至少一个上采样层分别对称地设置。

可在第一子网中不同深度的多个第一卷积层获取该图像的多个第一特征图，在第二子网中与前述多个第一卷积层对称设置的第二卷积层获取该图像的多个第二特征图。

可选地，在至少一个第一卷积层设有用于输出第一特征图的第一输出分支，在第二卷积层设有用于输出第二特征图的第二输出分支。

根据本申请的一种可选实施方式，第二子网还具有多个第三卷积层，所述第三卷积层的输入包括所述第一输出分支和所述第二输出分支。相应地，从所述多个第三卷积层的输出端分别获取所述融合特征图。

可使用任何具有上述结构的深度卷积神经网络。可选地，将第一子网和第二子网均构建为在物体检测中性能较佳的认知一样本归一化（Inception-BN）网络结构。Inception-BN 网络结构擅长于从图像中提取不同的结构/模式(pattern)，适合执行第一子网和第二子网的任务功能。

图 3 示出根据本公开实施例的深度卷积神经网络的一种示例性结构。

参照图 3，根据本实施例的深度卷积神经网络包括第一子网 SN1 和第二子网 SN2。其中，第一子网 SN1 具有多个第一卷积层 C1 和设置在多个第一卷积层 C1 之间的至少一个下采样层 P1，第二子网 SN2 具有多个第二卷积层 C2 和设置在多个第二卷积层 C2 之间的至少一个上采样层 P2，其中，下采样层 P1 和上采样层 P2 对称地设置，多个第一卷积层 C1 和多个第二卷积层 C2 也对称地设置。此外，至少一个第一卷积层 C1 设置有第一输出分支 F1，至少一个第二卷积层 C2 设置有第二输出分支 F2。第二子网 SN2 还设有多个第三卷积层 C3，自多个第三卷积层 C3 输出融合特征图。

根据本申请的一种可实施方式，融合特征图中的至少一个点具有与多个物体探测框对应的框融合检测数据以及预测准确信息。也就是说，在第一卷积层和第二卷积层分别设有用于进行物体区域框探测的物体探测框的信息，如：卷积参数或特征参数。在不同深度的第一卷积层和第二卷积层中设置的物体探测框的信息分别与两个或两个以上物体探测框集合各自对应，这两个或两个以上物体探测框集合分别包括不同探测框大小范围的物体探测框，以用于在该深度卷积神经网络的不同深度获取不同大小的物体区域框的特征数据。

融合特征图中的至少一个点的框融合检测数据可包括但不限于例如与物体探测框集合中的物体探测框相应的坐标数据、位置及大小数据，该预测准确信息可以是该框融合检测数据的置信度数据，如：预测准确概率。例如，融合特征图中的每个点可具有 1 个、3 个、6 个或 9 个与物体探测框相应的坐标数据以及该坐标数据的置信度数据。

相应地，根据本申请的一种可选实施方式，在步骤 210 之后执行步骤 S220。

步骤 220，根据至少一个融合特征图中的框融合检测数据以及预测准确信息分别获取与融合特征图各自

对应的目标区域框数据。

在一个可选示例中，步骤 S220 可以由处理器调用存储器存储的指令执行或者由被处理器运行。

具体地，可根据融合特征图中至少一个点的框融合检测数据的预测准确信息来获取目标区域框数据。例如，如果某个点的某个框坐标数据的置信度大于预定的阈值（如：60%、70%），则可将该框坐标数据对应的区域框确定为目标区域框数据之一。

根据本申请的另一种可选实施方式，在执行步骤 S210 后，执行步骤 S230-S240。

步骤 S230，分别获取融合特征图各自对应的初选目标区域框数据。

在一个可选示例中，步骤 S230 可以由处理器调用存储器存储的指令执行或者由被处理器运行。

可例如，执行与前述步骤 S220 或 S120 类似的处理，获取初选目标区域框数据，即，将前述步骤 S220 或 S120 获取到的目标区域框数据作为步骤 S230 中的初选目标区域框数据，以进行进一步的调整、修正处理，提高物体区域框检测的准确性。

在步骤 S240，迭代地执行以下物体区域框回归操作，直到迭代满足迭代终止条件为止，通过调整融合特征图，从经过调整的融合特征图获取新的初选目标区域框数据。

在一个可选示例中，步骤 S240 可以由处理器调用存储器存储的指令执行或者由被处理器运行。

也就是说，通过调整各个融合特征图来分别调整其中的初选目标区域框数据，再从经过调整的融合特征图分别获取新的初选目标区域框数据，从而对初选目标区域框进行回归（物体区域框回归操作），来获取更为准确的新的初选目标区域框数据。

在该步骤，迭代地执行这样的物体区域框回归操作，直到满足迭代终止条件为止，以最终获得更为精确的初选目标区域框数据。可根据需要设置该迭代终止条件，如：预定的迭代次数、新的初选目标区域框数据与未经过调整的初选目标区域框数据之间的调整值（即框回归）小于预定的框回归值。

在完成步骤 S240 的迭代之后，将经过迭代得到的初选目标区域框数据作为待处理的图像中的目标区域框数据。

根据本申请另一实施例的物体检测方法，通过具有对称结构的用于目标区域框检测的深度卷积神经网络，从第一子网的多个第一卷积层获取逐步经过卷积、下采样的待处理的图像的多个第一特征图，再从第二子网的对称的多个第二卷积层获取在第一子网的末端获取的第一特征图逐步经过反卷积、上采样的相应多个第二特征图，将多个第一特征图和相应的第二特征图进一步进行卷积，获得较好地表征了图像中高层的语义特征（如：布局、前背景信息）和低层的细节特征（如：小物体信息）的融合特征图，从而能够根据这些融合特征图有效地提取到图像中包含的大小物体的目标区域框数据。

在此基础上，通过调整多个融合特征图来从经过调整的融合特征图获取新的初选目标区域框数据，从而对初选目标区域框数据迭代地进行回归。通过对目标区域框数据进行多次的回归调整，能够更准确地检测到更为精准的包含的大小物体的目标区域框数据，进一步提高物体检测的准确性和鲁棒性。

图 4 是示出根据本申请又一实施例的物体检测方法的流程图。该实施例描述前述步骤 S240 中的一种示例性物体区域框回归操作。

根据该实施例的深度卷积神经网络还包括第三子网，第三子网具有多组第四卷积层和多个池化层，多组第四卷积层分别与第三卷积层对应，多个池化层分别与多组第四卷积层对应，并且每个池化层的输入包括经过调整的融合特征图和初选目标区域框的数据。

也就是说，每组第四卷积层可以包括一个或多个卷积层，每组第四卷积层可连接在前述第三卷积层的输出端，接收融合特征图作为输入。每个池化层设置在对应的第四卷积层的末端，接收经过调整的融合特征图和初选目标区域框数据作为输入。

其中，每组第四卷积层用于对从第三卷积层获取到的融合特征图进行卷积，获得调整融合特征图。在此过程中，对从该融合特征图获取的初选目标区域框数据进行调整。第三子网中的池化层用于对经过第四卷积层卷积获得的调整融合特征图进行区域池化，获取新的融合特征图。从而，可从新的融合特征图获取到新的初选目标区域框数据。

具体地，在每次迭代处理的物体区域框回归操作中，涉及当前迭代开始时的多个融合特征图以及初选目标区域框数据，还涉及当前迭代结束时获得的新的多个融合特征图以及新的初选目标区域框数据。

在步骤 S410，通过第四卷积层分别对当前的融合特征图进行卷积，获取调整融合特征图，从而对当前

的初选目标区域框进行调整,该调整包括对初选目标区域框的位置和/或大小的调整。

在一个可选示例中,该步骤 S410 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框调整单元 821 执行。

在步骤 S420,根据当前的初选目标区域框数据,通过池化层对调整融合特征图进行区域池化,获取新的融合特征图。

在一个可选示例中,该步骤 S420 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区域池化单元 822 执行。

也就是说,将当前的初选目标区域框作为关注区域,对调整融合特征图进行区域池化,获取新的融合特征图。

通过前述根据当前的初选目标区域框数据对调整融合特征图进行区域池化,获得反映对调整的目标区域框的响应程度的新的融合特征图,以便于后续从新的融合特征图获取新的初选目标区域框数据。

在步骤 S430,从新的融合特征图获取新的初选目标区域框数据,从而可完成目标区域框的回归,使得调整的目标区域框更趋近物体区域框的真值 (ground truth)。可通过与步骤 S120 或 S220 类似的处理执行步骤 S430 的处理。

在一个可选示例中,该步骤 S430 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初选框获取单元 823 执行。

根据本申请的一种可选实施方式,第三子网还具有设置在池化层输出端的第五卷积层。相应地,步骤 S430 具体包括:通过第五卷积层对新的融合特征图进行规范化卷积,并且从经过规范化卷积的融合特征图获取所述新的初选目标区域框数据。

可使用任何具有上述结构的卷积神经网络来构建第三子网。可选地,将第三子网构建为在新近开发的物体检测技术中性能较佳的残差网络 (ResNet) 结构结构,来执行区域池化和规范化卷积。

根据本申请上述实施例的物体检测方法,在前述各个实施例的基础上,通过对至少一个融合特征图进一步进行卷积,来对该融合特征图中包含的初选目标区域框数据进行调整,再经过区域池化来获得新的融合特征图,并从新的融合特征图获取新的初选目标区域框数据,从而对预测得到的初选目标区域框数据进行调整、回归,有助于提高物体检测的准确性和鲁棒性。

图 5 是示出根据本申请一实施例的神经网络的训练方法的流程图。

参照图 5,在步骤 S510,将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络,检测获取多个融合特征图。

在一个可选示例中,该步骤 S510 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融合特征图检测模块 1010 执行。

如前所述,深度卷积神经网络包括第一子网和第二子网,第一子网具有至少一个下采样层,第二子网具有至少一个上采样层;融合特征图通过第一特征图和第二特征图得到,第一特征图从第一子网获取得到,第二特征图从第二子网获取得到。

通过使用用于深度卷积神经网络,可从含有目标区域框标注信息的样本图像检测获取到多个融合特征图。

通常对多个样本图像执行步骤 S510 的处理,为至少一个样本图像检测获取多个融合特征图。

步骤 S520,根据多个融合特征图获取样本图像的目标区域框数据。

在一个可选示例中,该步骤 S520 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目标区域框检测模块 1020 执行。

步骤 S520 的处理与步骤 S120 的处理类似,在此不予赘述。

步骤 S530,根据获取到的样本图像的目标区域框数据以及目标区域框标注信息确定物体框检测的第一差异数据。

在一个可选示例中,该步骤 S530 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第一差异获取模块 1030 执行。

例如,可根据获取到的所述样本图像的目标区域框数据以及目标区域框标注信息计算损失值或偏差值作为该第一差异数据,作为后续训练深度卷积神经网络的依据。

在步骤 S540, 根据第一差异数据调整深度卷积神经网络的网络参数。

在一个可选示例中, 该步骤 S540 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第一网络训练模块 1040 执行。

例如, 将确定的第一差异数据反传给该深度卷积神经网络, 以调整该深度卷积神经网络的网络参数。

根据本申请提供的神经网络的训练方法, 将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络, 检测获取多个融合特征图; 其中, 从具有至少一个下采样层的第一子网检测获取多个第一特征图, 从具有至少一个上采样层的第二子网检测获取多个第二特征图, 分别由多个第一特征图和多个第二特征图融合得到融合特征图, 再根据多个融合特征图获取目标区域框数据。此后, 根据获取到的目标区域框数据以及目标区域框标注信息确定第一差异数据, 再根据第一差异数据调整深度卷积神经网络的网络参数。由于从训练获得的深度卷积神经网络的这些融合特征图较好地表征了图像中高层的语义特征(如: 布局、前背景信息)和低层的细节特征(如: 小物体信息), 因此根据这些融合特征图能够有效地提取到图像中包含的大小物体的目标区域框数据。训练获得的深度卷积神经网络能够提高物体检测的准确性和鲁棒性。

图 6 是示出根据本申请另一实施例的神经网络的训练方法的流程图。

根据本实施例, 在训练的深度卷积神经网络中, 第二子网设置在第一子网的末端; 第一子网具有多个第一卷积层和至少一个下采样层, 下采样层设置在多个第一卷积层之间; 第二子网具有多个第二卷积层和至少一个上采样层, 上采样层设置在多个第二卷积层之间。第一卷积层和第二卷积层对称设置, 至少一个下采样层和至少一个上采样层分别对称地设置。

在此基础上, 可选地, 在至少一个第一卷积层设有用于输出第一特征图的第一输出分支, 在第二卷积层设有用于输出第二特征图的第二输出分支。

为此, 可选地, 第二子网还具有多个第三卷积层, 第三卷积层的输入包括第一输出分支和第二输出分支。相应地, 第三卷积层用于对来自第一输出分支和第二输出分支的第一特征图和相应的第二特征图进行卷积, 获取相应的融合特征图。

参照图 6, 在步骤 S610, 缩放样本图像, 使得样本图像中的至少一个物体区域框的真值被物体探测框覆盖。如此, 可确保在任何批量的样本图像中具有正样本。

此外, 可选地, 选取足够数量的正样本, 并选取一定数量的负样本, 以使得训练得到的第一子网和第二子网较好地收敛。

在此, 正样本为正样本区域框, 负样本实为负样本区域框。可按照以下标准定义正样本区域框和负样本区域框: 正样本区域框与物体区域框的真值的重叠率不低于第一重叠比率值, 负样本区域框与物体区域框的真值的重叠率不高于第二重叠比率值, 第一重叠比率值大于第二重叠比率值。

相应地, 根据本申请的一种可实施方式, 样本图像的目标区域框标注信息包括正样本区域框的标注信息和负样本区域框的标注信息。

这里, 可根据设计需要设置第一重叠比率值, 例如: 将第一重叠比率值设置为 70%-95% 中的任何比率值, 将第二重叠比率值设置为 0%-30% 或 0-25% 范围中的任何比率值。

此外, 还可设置中性样本, 即中性样本区域框。具体地, 可按照以下标准定义中性样本区域框: 中性样本区域框与物体区域框的真值的重叠率在第一重叠比率值和第二重叠比率值之间, 如: 30%-70% 之间、25%-80% 之间。

进一步地, 可例如, 按照以下方式控制正样本、负样本和中性样本的数量: 在全部样本图像当中, 标注的正样本区域框的总和在正样本区域框、负样本区域框以及中性样本区域框的框总数中的占比不小于预定的第一比值, 第一比值大于 50%; 标注的负样本区域框的总和在框总数中的占比不大于预定的第二比值; 标注的中性样本区域框的总和在框总数中的占比不大于预定的第三比值, 第三比例不大于第一比值和第二比值之和的一半。适度地使用中性样本图像有助于更好地区分正样本和负样本, 提高训练的第三子网的鲁棒性。

在步骤 S620, 将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络, 检测获取多个融合特征图。其中, 从多个第三卷积层的输出端分别获取所述融合特征图。

在一个可选示例中, 该步骤 S620 可以由处理器调用存储器存储的相应指令执行。

可选地，融合特征图中的至少一个点的框融合检测数据可包括但不限于例如与物体探测框集合中的物体探测框相应的坐标数据、位置及大小数据，该预测准确信息可以是该框融合检测数据的置信度数据，如：预测准确概率。

可使用任何具有上述结构的深度卷积神经网络。可选地，将第一子网和第二子网均构建为在物体检测中性能较佳的 Inception-BN 网络结构。

可选地，融合特征图中的至少一个点的框融合检测数据可包括但不限于例如与物体探测框集合中的物体探测框相应的坐标数据、位置及大小数据，该预测准确信息可以是该框融合检测数据的置信度数据，如：预测准确概率。

相应地，步骤 S630，根据至少一个所述融合特征图中的框融合检测数据以及预测准确信息分别获取与所述融合特征图各自对应的目标区域框数据。

在一个可选示例中，该步骤 S630 可以由处理器调用存储器存储的相应指令执行。

步骤 S640，根据获取到的所述样本图像的目标区域框数据以及所述目标区域框标注信息确定物体框检测的第一差异数据。

在一个可选示例中，该步骤 S640 可以由处理器调用存储器存储的相应指令执行。

例如，可根据获取到的样本图像的目标区域框数据以及目标区域框标注信息计算损失值或偏差值作为该第一差异数据，作为后续训练深度卷积神经网络的依据。

在步骤 S650，根据第一差异数据调整深度卷积神经网络的网络参数。

在一个可选示例中，该步骤 S650 可以由处理器调用存储器存储的相应指令执行。

步骤 S640-S650 的处理与前述步骤 S530-S540 的处理类似，在此不予赘述。

根据本申请的神经网络的训练方法，将含有目标区域框标注信息的样本图像输入具有对称结构的用于目标区域框检测的深度卷积神经网络，检测获取多个融合特征图；其中，从具有至少一个下采样层的第一子网检测获取多个第一特征图，从具有至少一个上采样层的第二子网检测获取多个第二特征图，分别由多个第一特征图和多个第二特征图融合得到融合特征图，再根据多个融合特征图获取目标区域框数据。此后，根据获取到的目标区域框数据以及目标区域框标注信息确定第一差异数据，再根据第一差异数据调整深度卷积神经网络的网络参数。由于从训练获得的深度卷积神经网络的这些融合特征图较好地表征了图像中高层的语义特征（如：布局、前背景信息）和低层的细节特征（如：小物体信息），因此根据这些融合特征图能够有效地提取到图像中包含的大小物体的目标区域框数据。训练获得的深度卷积神经网络能够提高物体检测的准确性和鲁棒性。

图 7 是示出根据本申请又一实施例的神经网络的训练方法的流程图。

如前所述，根据上述实施例训练的该深度卷积神经网络还包括第三子网，第三子网具有多组第四卷积层和多个池化层，多组第四卷积层分别与第三卷积层对应，多个池化层分别与多组第四卷积层对应，并且每个池化层的输入包括经过调整的融合特征图和初选目标区域框的数据。

也就是说，每组第四卷积层可以包括一个或多个卷积层，每组第四卷积层可连接在前述第三卷积层的输出端，接收融合特征图作为输入。每个池化层设置在对应的第四卷积层的末端，接收经过调整的融合特征图和所述初选目标区域框数据作为输入。

在该实施例中，主要描述该深度卷积神经网络中的第三子网的训练。可先通过上述任一实施例的方法训练好第一子网和第二子网，再使用自第一子网和第二子网训练过程中获得的融合特征图，根据该实施例的方法来训练第三子网。

参照图 7，在步骤 S710，获取从含有目标区域框标注信息的样本图像获取的多个融合特征图。

如前步骤 S510 或 S610 所述，从样本图像获取该多个融合特征图。

在一个可选示例中，该步骤 S710 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融合特征图检测模块 1010 执行。在步骤 S720，迭代地执行目标区域框回归训练操作，直到迭代满足迭代终止条件为止。

在一个可选示例中，该步骤 S720 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框回归迭代训练模块 1050 执行。

具体地，步骤 S720 包括步骤 S721-S726。

在步骤 S721, 通过第四卷积层分别对当前的融合特征图进行卷积, 获取调整融合特征图, 从而达到对当前的初选目标区域框进行调整的目的。

在步骤 S722, 根据当前的初选目标区域框数据, 通过池化层对调整融合特征图进行区域池化, 获取新的融合特征图。新的融合特征图包含对初选目标区域框进行调整以及反映对调整的目标区域框的响应程度。

在步骤 S723, 从新的融合特征图获取新的初选目标区域框数据。

步骤 S721-S723 的处理与前述步骤 S410-S430 的处理类似, 在此不予赘述。

根据本申请的一种可选实施方式, 第三子网还具有设置在所述池化层输出端的第五卷积层。相应地, 步骤 S723 具体包括: 通过第五卷积层对新的融合特征图进行规范化卷积, 并且从经过规范化卷积的融合特征图获取新的初选目标区域框数据。

在步骤 S724, 根据未经过调整的初选目标区域框数据和新的初选目标区域框数据之间的框回归数据、新的初选目标区域框数据和相应的目标区域框标注信息确定物体框检测的第二差异数据。

在一个可选示例中, 该步骤 S724 可以由处理器调用存储器存储的相应指令执行。

例如, 可通过新的初选目标区域框数据和相应的目标区域框标注信息确定检测偏移, 并且根据检测偏移和框回归数据 (即框移动/调整数据) 来计算损失值作为第二差异数据。通过综合两个偏移参数 (检测偏移和框回归数据) 作为物体框检测的第二差异数据, 能够提高训练的第三子网的准确性。

在步骤 S725, 根据第二差异数据调整第三子网的网络参数。

在一个可选示例中, 该步骤 S725 可以由处理器调用存储器存储的相应指令执行。

例如, 将确定的第二差异数据反传给第三子网, 以调整第三子网的网络参数。

在步骤 S726, 确定是否满足迭代终止条件。

在一个可选示例中, 该步骤 S726 可以由处理器调用存储器存储的相应指令执行。

如果在步骤 S726, 确定前述的迭代满足迭代终止条件 (如: 达到预定的迭代次数), 则结束对第三子网的训练; 如果在步骤 S726, 确定前述的迭代不满足迭代终止条件 (如: 达到预定的迭代次数), 则返回执行步骤 S721, 继续进行前述对第三子网的训练, 直到确定满足迭代终止条件为止。

现有的用于物体区域框回归的神经网络的训练仅针对一次目标区域框回归执行迭代 (如迭代次数 N) 的训练; 而根据本申请提供的训练方法, 对目标区域框执行多次回归 (如回归次数 M), 每次回归涉及多次迭代 (如迭代次数 N) 的训练, 即涉及 $M \times N$ 次迭代训练。由此训练得到的第三子网在进行物体区域框的定位检测上更为准确。

可使用任何具有上述结构的卷积神经网络来构建第三子网。可选地, 将第三子网构建为在新近开发的物体检测技术中性能较佳的 ResNet 结构, 来执行区域池化和规范化卷积。

根据本申请提供的神经网络的训练方法, 在前述各实施例的基础上, 训练得到的深度卷积神经网络通过对样本图像的每个融合特征图进一步进行卷积, 来对该融合特征图中包含的初选目标区域框数据进行调整, 再经过区域池化来获得新的融合特征图, 并从新的融合特征图获取新的初选目标区域框数据, 从而对得到的初选目标区域框数据进行调整、回归, 能够进一步提高物体检测的准确性和鲁棒性。

图 8 是示出根据本申请一实施例的物体检测装置的结构框图。

参照图 8, 本实施例的物体检测装置包括融合特征图预测模块 810 和目标区域框预测模块 820。

融合特征图预测模块 810 用于通过用于目标区域框检测的深度卷积神经网络, 从待处理的图像预测获取多个融合特征图; 其中, 深度卷积神经网络包括第一子网和第二子网, 第一子网具有至少一个下采样层, 第二子网具有至少一个上采样层; 融合特征图通过第一特征图和第二特征图得到, 第一特征图从第一子网获取得到, 第二特征图从第二子网获取得到。

目标区域框预测模块 820 用于根据融合特征图预测模块 810 获取的多个融合特征图获取目标区域框数据。

本实施例的物体检测装置用于实现前述方法实施例中相应的物体检测方法, 并具有相应的方法实施例的有益效果, 在此不再赘述。

图 9 是示出根据本申请另一实施例的物体检测装置的结构框图。

根据本实施例, 在用于目标区域框检测的深度卷积神经网络中, 第二子网设置在第一子网的末端, 第一子网具有多个第一卷积层和至少一个下采样层, 下采样层设置在多个第一卷积层之间, 第二子网具有多

个第二卷积层和所述至少一个上采样层，上采样层设置在多个第二卷积层之间，第一卷积层和第二卷积层对称设置，至少一个下采样层和至少一个上采样层分别对称地设置。

根据一种可选的实施方式，在至少一个第一卷积层设有用于输出第一特征图的第一输出分支，在第二卷积层设有用于输出第二特征图的第二输出分支。

根据一种可选的实施方式，第二子网还具有多个第三卷积层，第三卷积层的输入包括第一输出分支和所述第二输出分支。相应地，融合特征图预测模块 810 用于从多个第三卷积层的输出端分别获取融合特征图。

可选地，融合特征图中的至少一个点具有与多个物体探测框对应的框融合检测数据以及预测准确信息。相应地，目标区域框预测模块 820 用于根据至少一个融合特征图中的框融合检测数据以及预测准确信息分别获取与融合特征图各自对应的目标区域框数据。

可选地，目标区域框预测模块 820 用于分别获取融合特征图各自对应的初选目标区域框数据；迭代地执行以下物体区域框回归操作，直到迭代满足迭代终止条件为止：通过调整融合特征图，从经过调整的融合特征图获取新的初选目标区域框数据；将经过迭代得到的初选目标区域框数据作为待处理的图像中的目标区域框数据。

可选地，深度卷积神经网络还包括第三子网，第三子网具有多组第四卷积层和多个池化层，多组第四卷积层分别与第三卷积层对应，多个池化层分别与多组第四卷积层对应，并且每个池化层的输入包括经过调整的融合特征图和初选目标区域框的数据。

可选地，目标区域框预测模块 820 包括：

框调整单元 821，用于通过第四卷积层分别对当前的所述融合特征图进行卷积，获取调整融合特征图；

区域池化单元 822，用于根据当前的初选目标区域框数据，通过池化层对调整融合特征图进行区域池化，获取新的融合特征图；

初选框获取单元 823，用于从新的融合特征图获取新的初选目标区域框数据。

可选地，第三子网还具有设置在池化层输出端的第五卷积层；相应地，初选框获取单元 823 用于通过第五卷积层对新的融合特征图进行规范化卷积，并且从经过规范化卷积的融合特征图获取新的初选目标区域框数据。

可选地，第一子网和第二子网均为认知一样本归一化（Inception-BN）网络结构，第三子网为残差网络（ResNet）结构。

本实施例的物体检测装置用于实现前述方法实施例中相应的物体检测方法，并具有相应的方法实施例的有益效果，在此不再赘述。

图 10 是示出根据本申请一实施例的神经网络的训练装置的结构框图。

参照图 10，本实施例的神经网络的训练装置还包括融合特征图检测模块 1010、目标区域框检测模块 1020、第一差异获取模块 1030 和第一网络训练模块 1040。

融合特征图检测模块 1010 用于将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络，检测获取多个融合特征图，深度卷积神经网络包括第一子网和第二子网，第一子网具有至少一个下采样层，第二子网具有至少一个上采样层；融合特征图通过第一特征图和第二特征图得到，第一特征图从第一子网获取得到，第二特征图从第二子网获取得到。

目标区域框检测模块 1020 用于根据多个融合特征图获取样本图像的目标区域框数据。

第一差异获取模块 1030 用于根据获取到的样本图像的目标区域框数据以及目标区域框标注信息确定物体框检测的第一差异数据。

第一网络训练模块 1040 用于根据第一差异数据调整深度卷积神经网络的网络参数。

本实施例的神经网络的训练装置用于实现前述方法实施例中相应的神经网络的训练方法，并具有相应的方法实施例的有益效果，在此不再赘述。

图 11 是示出根据本申请另一实施例的神经网络的训练装置的结构框图。

根据本实施例，在用于目标区域框检测的深度卷积神经网络中，第二子网设置在第一子网的末端，第一子网具有多个第一卷积层和至少一个下采样层，下采样层设置在多个第一卷积层之间，第二子网具有多个第二卷积层和至少一个上采样层，上采样层设置在多个第二卷积层之间，第一卷积层和第二卷积层对称

设置，至少一个下采样层和至少一个上采样层分别对称地设置。

根据一种可选的实施方式，在至少一个第一卷积层设有用于输出第一特征图的第一输出分支，在第二卷积层设有用于输出第二特征图的第二输出分支。

根据一种可选的实施方式，第二子网还具有多个第三卷积层，第三卷积层的输入包括第一输出分支和第二输出分支。相应地，融合特征图检测模块 1010 用于从多个第三卷积层的输出端分别获取融合特征图。

可选地，融合特征图中的至少一个点具有与多个物体探测框对应的框融合检测数据以及预测准确信息。

可选地，深度卷积神经网络还包括第三子网，第三子网具有多组第四卷积层和多个池化层，多组第四卷积层分别与第三卷积层对应，多个池化层分别与多组第四卷积层对应，并且每个池化层的输入包括经过调整的融合特征图和初选目标区域框的数据。

可选地，上述装置还包括：框回归迭代训练模块 1050，用于迭代地执行以下目标区域框回归训练操作，直到迭代满足迭代终止条件为止：通过第四卷积层分别对当前的融合特征图进行卷积，获取调整融合特征图；根据当前的初选目标区域框数据，通过池化层对调整融合特征图进行区域池化，获取新的融合特征图；从新的融合特征图获取新的初选目标区域框数据；根据未经过调整的初选目标区域框数据和新的初选目标区域框数据之间的框回归数据、新的初选目标区域框数据和相应的目标区域框标注信息确定物体框检测的第二差异数据；根据第二差异数据调整第三子网的网络参数。

可选地，第三子网还具有设置在池化层输出端的第五卷积层；相应地，框回归迭代训练模块 1050 用于通过第五卷积层对新的融合特征图进行规范化卷积，并且从经过规范化卷积的融合特征图获取所述新的初选目标区域框数据。

可选地，上述装置还包括：预处理模块 1060，用于在迭代地执行目标区域框回归训练操作之前，缩放样本图像，使得至少一个物体区域框的真值被物体探测框覆盖。

可选地，样本图像的目标区域框标注信息包括正样本区域框的标注信息和负样本区域框的标注信息；正样本区域框与物体区域框的真值的重叠率不低于第一重叠比率值，负样本区域框与物体区域框的真值的重叠率不高于第二重叠比率值，第一重叠比率值大于第二重叠比率值。

可选地，样本图像的目标区域框标注信息还包括中性样本区域框的标注信息，中性样本区域框与物体区域框的真值的重叠率在第一重叠比率值和第二重叠比率值之间。

可选地，，在全部样本图像当中，标注的正样本区域框的总和在正样本区域框、负样本区域框以及中性样本区域框的框总数中的占比不小于预定的第一比值，该第一比值大于 50%；标注的负样本区域框的总和在框总数中的占比不大于预定的第二比值；标注的中性样本区域框的总和在框总数中的占比不大于预定的第三比值，第三比例不大于第一比值和第二比值之和的一半。

可选地，第一子网和第二子网均为认知一样本归一化网络结构，第三子网为残差网络结构。

本实施例的神经网络的训练装置用于实现前述方法实施例中相应的神经网络的训练方法，并具有相应的方法实施例的有益效果，在此不再赘述。

另外，本申请实施例还提供了一种电子设备，包括：处理器和存储器；

所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行本申请上述任一实施例所述的物体检测方法对应的操作；或者，

所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行本申请上述任一实施例所述的神经网络的训练方法对应的操作。

另外，本申请实施例还提供了另一种电子设备，包括：

处理器和本申请上述任一实施例所述的物体检测装置；在处理器运行所述物体检测装置时，本申请上述任一实施例所述的物体检测装置中的单元被运行；或者

处理器和本申请上述任一实施例所述的神经网络的训练装置；在处理器运行所述神经网络的训练装置时，本申请上述任一实施例所述的神经网络的训练装置中的单元被运行。

图 12 是示出根据本申请一个实施例的第一电子设备的结构示意图。

本申请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。下面参考图 12，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第一电子设备 1200 的结构示意图。

如图 12 所示，第一电子设备 1200 包括但不限于一个或多个第一处理器、第一通信元件，所述一个或

多个第一处理器例如：一个或多个第一中央处理单元（CPU）1201，和/或一个或多个第一图像处理器（GPU）1213，第一处理器可以根据存储在一只读存储器（ROM）1202 中的可执行指令或者从第一存储部分 1208 加载到第一随机访问存储器（RAM）1203 中的可执行指令而执行各种适当的动作和处理。第一通信元件包括第一通信组件 1212 和第一通信接口 1209。其中，第一通信组件 1212 可包括但不限于网卡，所述网卡可包括但不限于 IB（Infiniband）网卡，第一通信接口 1209 包括诸如 LAN 卡、调制解调器的网络接口卡的通信接口，第一通信接口 1209 经由诸如因特网的网络执行通信处理。

第一处理器可与一只读存储器 1202 和/或第一随机访问存储器 1203 中通信以执行可执行指令，通过第一总线 1204 与第一通信组件 1212 相连、并经第一通信组件 1212 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，通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图；其中，所述深度卷积神经网络包括第一子网和第二子网，所述第一子网具有至少一个下采样层，所述第二子网具有至少一个上采样层；所述融合特征图通过第一特征图和第二特征图得到，所述第一特征图从第一子网获取得到，所述第二特征图从第二子网获取得到；根据所述多个融合特征图获取目标区域框数据。

此外，在第一 RAM 1203 中，还可存储有装置操作所需的各种程序和数据。第一 CPU 1201、第一 ROM 1202 以及第一 RAM 1203 通过第一总线 1204 彼此相连。在有第一 RAM 1203 的情况下，第一 ROM 1202 为可选模块。第一 RAM 1203 存储可执行指令，或在运行时向第一 ROM 1202 中写入可执行指令，可执行指令使第一处理器 1201 执行上述通信方法对应的操作。第一输入/输出（I/O）接口 1205 也连接至第一总线 1204。第一通信组件 1212 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。

以下部件连接至第一 I/O 接口 1205：包括键盘、鼠标的第一输入部分 1206；包括但不限于诸如阴极射线管（CRT）、液晶显示器（LCD）以及扬声器的第一输出部分 1207；包括但不限于硬盘的第一存储部分 1208；以及包括但不限于诸如 LAN 卡、调制解调器的网络接口卡的第一通信接口 1209。第一驱动器 1210 也需要连接至第一 I/O 接口 1205。第一可拆卸介质 1211，诸如磁盘、光盘、磁光盘、半导体存储器，根据需要安装在第一驱动器 1210 上，以便于从其上读出的计算机程序根据需要被安装入第一存储部分 1208。

需要说明的是，如图 12 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 12 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实现方式，例如 GPU 和 CPU 可分离设置或者可将 GPU 集成在 CPU 上，第一通信组件 1212 可分离设置，也可集成设置在 CPU 或 GPU 上。这些可替换的实施方式均落入本申请的保护范围。

特别地，根据本申请实施例，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，本申请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，用于通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图的可执行代码；其中，所述深度卷积神经网络包括第一子网和第二子网，所述第一子网具有至少一个下采样层，所述第二子网具有至少一个上采样层；所述融合特征图通过第一特征图和第二特征图得到，所述第一特征图从第一子网获取得到，所述第二特征图从第二子网获取得到；用于根据所述多个融合特征图获取目标区域框数据的可执行代码。在这样的实施例中，该计算机程序可以通过通信元件从网络上被下载和安装，和/或从第一可拆卸介质 1211 被安装。在该计算机程序被第一中央处理单元（CPU）1201 执行时，执行本申请的方法中限定的上述功能。

本申请该实施例提供的电子设备，通过用于目标区域框检测的深度卷积神经网络，从待处理的图像预测获取多个融合特征图，其中，从具有至少一个下采样层的第一子网获取多个第一特征图，从具有至少一个上采样层的第二子网获取多个第二特征图，分别由多个第一特征图和多个第二特征图融合得到融合特征图。此后，再根据所述多个融合特征图获取目标区域框数据。由于这些融合特征图较好地表征了图像中高层的语义特征（如：布局、前背景信息）和低层的细节特征（如：小物体信息），因此根据这些融合特征图能够有效地提取到图像中包含的大小物体的目标区域框数据，从而提高物体检测的准确性和鲁棒性。

图 13 是示出根据本申请另一实施例的第二电子设备的结构示意图。

本申请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。下面参考图 13，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第二电子设备 1300 的结构示意图。

如图 13 所示,第二电子设备 1300 包括但不限于一个或多个第二处理器、第二通信元件,所述一个或多个第二处理器例如:一个或多个第二中央处理单元(CPU)1301,和/或 一个或多个第二图像处理器(GPU)1313,第二处理器可以根据存储在第二只读存储器(ROM)1302 中的可执行指令或者从第二存储部分 1308 加载到第二随机访问存储器(RAM)1303 中的可执行指令而执行各种适当的动作和处理。第二通信元件包括第二通信组件 1312 和第二通信接口 1309。其中,第二通信组件 1312 可包括但不限于网卡,所述网卡可包括但不限于 IB(Infiniband)网卡,第二通信接口 1309 包括诸如 LAN 卡、调制解调器的网络接口卡的通信接口,第二通信接口 1309 经由诸如因特网的网络执行通信处理。

第二处理器可与第二只读存储器 1302 和/或第二随机访问存储器 1303 中通信以执行可执行指令,通过第二总线 1304 与第二通信组件 1312 相连、并经第二通信组件 1312 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,将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络,检测获取多个融合特征图,所述深度卷积神经网络包括第一子网和第二子网,所述第一子网具有至少一个下采样层,所述第二子网具有至少一个上采样层;所述融合特征图通过第一特征图和第二特征图得到,所述第一特征图从第一子网获取得到,所述第二特征图从第二子网获取得到;根据所述多个融合特征图获取所述样本图像的目标区域框数据;根据获取到的所述样本图像的目标区域框数据以及所述目标区域框标注信息确定物体框检测的第一差异数据;根据所述第一差异数据调整所述深度卷积神经网络的网络参数。

此外,在第二 RAM 1303 中,还可存储有装置操作所需的各种程序和数据。第二 CPU1301、第二 ROM1302 以及第二 RAM1303 通过第二总线 1304 彼此相连。在有第二 RAM1303 的情况下,第二 ROM1302 为可选模块。第二 RAM1303 存储可执行指令,或在运行时向第二 ROM1302 中写入可执行指令,可执行指令使第二处理器 1301 执行上述通信方法对应的操作。第二输入/输出(I/O)接口 1305 也连接至第二总线 1304。第二通信组件 1312 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。

以下部件连接至第二 I/O 接口 1305:包括但不限于键盘、鼠标的第二输入部分 1306;包括但不限于诸如阴极射线管(CRT)、液晶显示器(LCD)以及扬声器的第二输出部分 1307;包括但不限于硬盘的第二存储部分 1308;以及包括诸如 LAN 卡、调制解调器的网络接口卡的第二通信接口 1309。第二驱动器 1310 也需要连接至第二 I/O 接口 1305。第二可拆卸介质 1311,诸如磁盘、光盘、磁光盘、半导体存储器,根据需要安装在第二驱动器 1310 上,以便于从其上读出的计算机程序根据需要被安装入第二存储部分 1308。

需要说明的是,如图 13 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 13 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实现方式,例如 GPU 和 CPU 可分离设置或者可将 GPU 集成在 CPU 上,第二通信组件 1312 可分离设置,也可集成设置在 CPU 或 GPU 上。这些可替换的实施方式均落入本申请的保护范围。

特别地,根据本申请实施例,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,本申请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,用于将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络,检测获取多个融合特征图的可执行代码,所述深度卷积神经网络包括第一子网和第二子网,所述第一子网具有至少一个下采样层,所述第二子网具有至少一个上采样层;所述融合特征图通过第一特征图和第二特征图得到,所述第一特征图从第一子网获取得到,所述第二特征图从第二子网获取得到;用于根据所述多个融合特征图获取所述样本图像的目标区域框数据的可执行代码;用于根据获取到的所述样本图像的目标区域框数据以及所述目标区域框标注信息确定物体框检测的第一差异数据的可执行代码;用于根据所述第一差异数据调整所述深度卷积神经网络的网络参数的可执行代码。在这样的实施例中,该计算机程序可以通过通信元件从网络上被下载和安装,和/或从第二可拆卸介质 1311 被安装。在该计算机程序被第二中央处理单元(CPU)1301 执行时,执行本申请实施例的方法中限定的上述功能。

本申请该实施例提供的电子设备,将含有目标区域框标注信息的样本图像输入用于目标区域框检测的深度卷积神经网络,检测获取多个融合特征图;其中,从具有至少一个下采样层的第一子网检测获取多个第一特征图,从具有至少一个上采样层的第二子网检测获取多个第二特征图,分别由多个第一特征图和多个第二特征图融合得到融合特征图,再根据所述多个融合特征图获取目标区域框数据。此后,根据获取到

的目标区域框数据以及所述目标区域框标注信息确定第一差异数据，再根据所述第一差异数据调整所述深度卷积神经网络的网络参数。由于从训练获得的深度卷积神经网络的这些融合特征图较好地表征了图像中高层的语义特征（如：布局、前背景信息）和低层的细节特征（如：小物体信息），因此根据这些融合特征图能够有效地提取到图像中包含的大小物体的目标区域框数据。训练获得的深度卷积神经网络能够提高物体检测的准确性和鲁棒性。

另外，本申请实施例还提供了一种计算机程序，包括计算机可读代码，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本申请任一实施例所述的物体检测方法中各步骤的指令；或者

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本申请任一实施例所述的神经网络的训练方法中各步骤的指令。

另外，本申请实施例还提供了一种计算机可读存储介质，用于存储计算机可读取的指令，所述指令被执行时实现本申请任一实施例所述的物体检测方法中各步骤的操作、或者本申请任一实施例所述的神经网络的训练方法中各步骤的操作。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。