



(12)发明专利

(10)授权公告号 CN 106295678 B

(45)授权公告日 2020.03.06

(21)申请号 201610605759.1

G06K 9/00(2006.01)

(22)申请日 2016.07.27

审查员 孙巍巍

(65)同一申请的已公布的文献号

申请公布号 CN 106295678 A

(43)申请公布日 2017.01.04

(73)专利权人 北京旷视科技有限公司

地址 100080 北京市海淀区科学院南路2号
A座313室

专利权人 北京迈格威科技有限公司

(72)发明人 余家辉 印奇

(74)专利代理机构 北京市柳沈律师事务所

11105

代理人 于小宁 胡琪

(51)Int.Cl.

G06K 9/62(2006.01)

权利要求书9页 说明书36页 附图10页

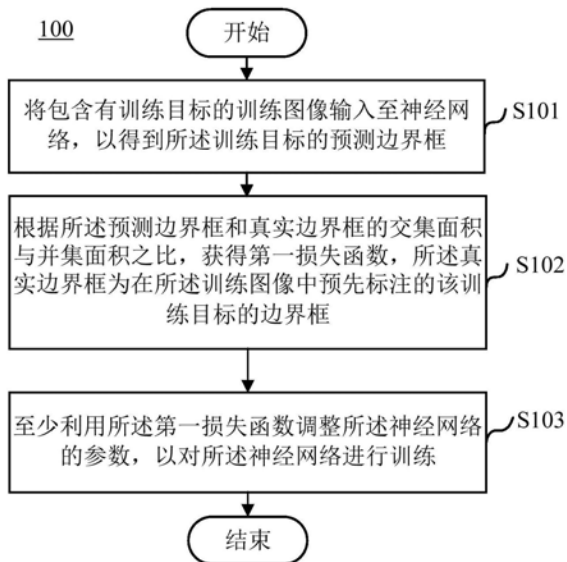
(54)发明名称

神经网络训练与构建方法和装置以及目标检测方法和装置

(57)摘要

本发明的实施例提供了用于目标检测的神经网络训练与构建方法和装置以及基于神经网络的目标检测方法和装置。所述用于目标检测的神经网络训练方法包括：将包含有训练目标的训练图像输入至神经网络，以得到所述训练目标的预测边界框；根据所述预测边界框和真实边界框的交集面积与并集面积之比，获得第一损失函数，所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框；以及至少利用所述第一损失函数调整所述神经网络的参数，以对所述神经网络进行训练。本发明实施例的神经网络训练与构建方法和装置以及目标检测方法和装置利用第一损失函数将目标边界框回归为一个整体单元，显著增加了神经网络的目标检测精度，并且能够通过神经网络的两个分支结构有效提高神经网络的训练和检测效率。

CN 106295678 B



1. 一种用于目标检测的神经网络的训练方法,包括:

将包含有训练目标的训练图像输入至神经网络,以得到所述训练目标的预测边界框;

根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及

至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

所述方法还包括:

选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度,

所述至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:

利用所述第一损失函数和第二损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,以得到训练好的神经网络;

其中,所述神经网络至少包括一个输入层、n个中间层和两个输出层,其中,n是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层,

其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,

所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

2. 如权利要求1所述的训练方法,其中至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:

利用所述第一损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比。

3. 如权利要求1所述的训练方法,其中至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:

确定当所述神经网络的训练轮数达到设定的阈值时参数的值,作为训练好的神经网络的参数的值。

4. 如权利要求1所述的训练方法,其中根据预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,包括:

对于位于所述训练图像中真实边界框内的至少一个像素点,确定第一4维向量和第二4维向量,其中,该第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离,所述第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界

框的上、下、左、右边界的距离；

根据所述第一4维向量,获得所述真实边界框的面积；

根据所述第二4维向量,获得所述预测边界框的面积；

根据所述第一4维向量和所述第二4维向量,获得所述预测边界框和所述真实边界框的交集面积；以及

将获得的所述预测边界框的面积和所述真实边界框的面积的和与所述交集面积的差作为所述并集面积。

5.如权利要求1所述的训练方法,还包括:

对于每个输入的训练图像,还向所述输入层输入与该训练图像相同大小的真实置信度图以及训练边界框图,其中,所述训练边界框图包含所述训练图像中的每个像素点的4维向量,并且,当像素点位于所述训练图像中的真实边界框内时,该像素点对应的4维向量中的元素为该像素点距离所述真实边界框的上、下、左、右边界的距离,所述真实置信度图中的每个像素点表示所述训练图像中对应像素点属于所述训练目标的置信度。

6.如权利要求1所述的训练方法,其中所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

7.一种用于目标检测的神经网络的训练装置,包括:

处理器;

存储器;和

存储在所述存储器中的计算机程序指令,在所述计算机程序指令被所述处理器运行时,使得所述处理器执行以下步骤:

将包含有训练目标的训练图像输入至神经网络,以得到所述训练目标的预测边界框;

根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及

至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

其中,所述处理器还执行以下步骤:

选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度,

所述至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:

利用所述第一损失函数和第二损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,以得到训练好的神经网络;

其中,所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层,

其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输

出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,

所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

8.一种用于目标检测的神经网络的训练装置,包括:

预测部件,配置为将包含有训练目标的训练图像输入至神经网络,以得到所述训练目标的预测边界框;

第一获取部件,配置为根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及

参数调整部件,配置为至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

所述装置还包括:第二获取部件,配置为选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度;

所述参数调整部件利用所述第一损失函数和第二损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,以得到训练好的神经网络;

其中,所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层,

其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,

所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

9.一种用于目标检测的神经网络的构建方法,包括:

构建至少包含一个输入层、 n 个中间层和两个输出层的神经网络, n 是大于等于2的正整数,其中,将第 m 个中间层级联至所述神经网络的第一支路,将第 n 个中间层级联至所述神经网络的第二支路,所述两个输出层中的第一输出层位于所述第一支路,第二输出层位于所述第二支路,并且, m 小于 n ;

将所述神经网络的第一支路的第一输出层配置为在用所述神经网络进行目标检测时输出置信度图的置信度输出层,所述置信度图表示采用所述神经网络检测出的、包含检测

目标的检测图像中的每个像素点属于所述检测目标的置信度;以及

将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该检测目标的边界框;

其中,所述方法还包括:

训练所述神经网络,

其中,所述训练神经网络包括:

将包含有训练目标的训练图像输入至所述神经网络,以得到所述训练目标的预测边界框;

根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,其中,所述预测边界框为采用该神经网络在包含所述训练目标的训练图像中检测到的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及

至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

所述训练神经网络还包括:

选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度,

所述至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:

利用所述第一损失函数和第二损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,以得到训练好的神经网络;其中,当像素点位于所述检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为与所述像素点至所述检测边界框的上、下、左、右边界的距离。

10. 如权利要求9所述的构建方法,其中,所述神经网络的第一支路上还包括:

卷积层,所述卷积层配置为对所述第m个中间层输出的特征图像进行卷积,以获得所述检测图像的第一特征图;

上采样层,所述上采样层配置为对所述第一特征图进行线性插值,将所述第一特征图调整为所述检测图像的大小;以及

剪裁层,所述剪裁层配置为剪裁所述调整后的第一特征图,以使得所述剪裁后的第一特征图与所述检测图像对齐,并将对齐后的第一特征图输出至所述第一支路的第一输出层,以输出所述置信度图。

11. 如权利要求9所述的构建方法,其中,所述神经网络的第二支路上还包括:

卷积层,所述卷积层配置为对所述第n个中间层输出的特征图像进行卷积,以获得所述

检测图像的第二特征图；

上采样层,所述上采样层配置为对所述第二特征图进行线性插值,将所述第二特征图调整为所述检测图像的大小；

剪裁层,所述剪裁层配置为剪裁所述调整后的第二特征图,以使得所述剪裁后的第二特征图与所述检测图像对齐;以及

修正线性单元层,所述修正线性单元层配置为获取对齐后的第二特征图,并将所述第二特征图中的每个像素点的对应向量中的元素调整为不小于0的向量元素后输出至所述第二支路的第二输出层,以输出所述边界框检测图。

12. 如权利要求9所述的构建方法,其中,至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:

利用所述第一损失函数以及反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数对该神经网络进行训练,其中,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度。

13. 一种用于目标检测的神经网络的构建装置,包括:

处理器;

存储器;和

存储在所述存储器中的计算机程序指令,在所述计算机程序指令被所述处理器运行时使得所述处理器执行以下步骤:

构建至少包含一个输入层、 n 个中间层和两个输出层的神经网络, n 是大于等于2的正整数,其中,将第 m 个中间层级联至所述神经网络的第一支路,将第 n 个中间层级联至所述神经网络的第二支路,所述两个输出层中的第一输出层位于所述第一支路,第二输出层位于所述第二支路,并且, m 小于 n ;

将所述神经网络的第一支路的第一输出层配置为在用所述神经网络进行目标检测时输出置信度图的置信度输出层,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度;以及

将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该检测目标的边界框;

其中,所述处理器还执行以下步骤:

训练所述神经网络,

其中,所述训练神经网络包括:

将包含有训练目标的训练图像输入至所述神经网络,以得到所述训练目标的预测边界框;

根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,其中,所述预测边界框为采用该神经网络在包含所述训练目标的训练图像中检测到的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及

至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

其中,所述处理器还执行以下步骤:

选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度,

所述至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:

利用所述第一损失函数和第二损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,以得到训练好的神经网络;

其中,当像素点位于所述检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为与所述像素点至所述检测边界框的上、下、左、右边界的距离。

14. 一种用于目标检测的神经网络的构建装置,包括:

构建部件,配置为构建至少包含一个输入层、n个中间层和两个输出层的神经网络,n是大于等于2的正整数,其中,将第m个中间层级联至所述神经网络的第一支路,将第n个中间层级联至所述神经网络的第二支路,所述两个输出层中的第一输出层位于所述第一支路,第二输出层位于所述第二支路,并且,m小于n;

第一配置部件,配置为将所述神经网络的第一支路的第一输出层配置为在用所述神经网络进行目标检测时输出置信度图的置信度输出层,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度;以及

第二配置部件,配置为将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该检测目标的边界框;

其中,所述构建装置还包括:

训练部件,配置为通过以下方式训练所述神经网络:

将包含有训练目标的训练图像输入至所述神经网络,以得到所述训练目标的预测边界框;

根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,其中,所述预测边界框为采用该神经网络在包含所述训练目标的训练图像中检测到的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及

至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

其中,所述训练部件选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度,

所述至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:利用所述第一损失函数和第二损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,以得到训练好的神经网络;

其中,当像素点位于所述检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为与所述像素点至所述检测边界框的上、下、左、右边界的距离。

15. 一种基于神经网络的目标检测方法,包括:

获取包含检测目标的检测图像;

利用神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含该检测图像中各像素点的对应向量,所述对应向量表示与其对应的像素点与对应检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框;

在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状;

在拟合得到的几何形状内选取预定数量的像素点;

对于每个选取的像素点,确定所述边界框检测图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框;

将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个检测边界框;

其中,所述神经网络是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述第二损失函数反映预测置信度与真实置信度之间的差异,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

其中,当像素点位于对应的检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为与所述像素点至对应的检测边界框的上、下、左、右边界的距离。

16. 如权利要求15所述的目标检测方法,其中将各个所述检测边界框进行合并包括:

利用非最大抑制方法将各个所述检测边界框进行合并。

17. 如权利要求15所述的目标检测方法,其中在所述置信度图中基于各个像素点的置信度通过形状拟合得到至少一个几何形状包括:

确定置信度大于预定阈值的多个像素点;

基于所述多个像素点拟合得到所述至少一个几何形状。

18. 如权利要求15所述的目标检测方法,其中,

所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括用于输出所述置信度图的置信度输出层和用于输出所述边界框检测图的边界框输出层。

19. 如权利要求15所述的目标检测方法,其中所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

20. 一种基于神经网络的目标检测装置,包括:

处理器;

存储器;和

存储在所述存储器中的计算机程序指令,在所述计算机程序指令被所述处理器运行时使得所述处理器执行以下步骤:

获取包含检测目标的检测图像;

利用所述神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含该检测图像中各像素点的对应向量,所述对应向量表示与其对应的像素点与对应检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框;

在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状;

在拟合得到的几何形状内选取预定数量的像素点;

对于每个选取的像素点,确定所述边界框检测图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框;

将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个检测边界框;

其中,所述神经网络是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述第二损失函数反映预测置信度与真实置信度之间的差异,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

其中,所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层,

其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,

所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图

像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

21. 一种基于神经网络的目标检测装置,包括:

图像获取部件,配置为获取包含检测目标的检测图像;

检测部件,配置为利用所述神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含该检测图像中各像素点的对应向量,所述对应向量图中的各对应向量表示与其对应的像素点与对应检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框;

拟合部件,配置为在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状;

选取部件,配置为在拟合得到的几何形状内选取预定数量的像素点;

边界框获取部件,配置为对于每个选取的像素点,确定所述边界框检测图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框;

合并部件,配置为将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个检测边界框;

其中,所述神经网络是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述第二损失函数反映预测置信度与真实置信度之间的差异,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度;

其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值;

其中,所述神经网络至少包括一个输入层、n个中间层和两个输出层,其中,n是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层,

其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,

所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

神经网络训练与构建方法和装置以及目标检测方法和装置

技术领域

[0001] 本公开的实施例涉及图像处理,尤其涉及用于目标检测的神经网络的训练与构建方法和装置、以及基于神经网络的目标检测方法和装置。

背景技术

[0002] 目标检测是计算机视觉领域中一个基础性的研究课题,其在人脸识别、安全监控以及动态追踪等很多方面都有广泛的应用前景。目标检测是指对于任意一幅给定的图像,检测和识别其中特定的目标(例如人脸),并返回目标的位置、大小信息,例如输出包围目标的边界框。目标检测是一个复杂的具有挑战性的模式检测问题,其主要的难点有两方面,一方面是由于目标的细节变化、遮挡等内在的变化所引起的,另外一方面是由于成像角度、光照影响、成像设备的焦距、成像距离、图像获得的途径不同等外在条件变化所引起。

[0003] 基于深度卷积神经网络(CNN)的目标检测方法是当前较为先进的目标检测方法。现有的基于CNN的目标检测方法一般分为三步:1)利用传统的候选区域(region proposal)方法从图像中提取若干个可能是待检测目标的候选区域;2)将所提取的候选区域输入CNN中用以识别和分类;3)采用边界框回归技术将粗糙的候选目标细化为更加精确的目标边界。然而,当前基于CNN的目标检测方法所获取的目标检测结果仍然存在受目标的内在的变化的影响、目标识别不准确、检测效率低等技术问题。

发明内容

[0004] 考虑到上述问题而提出了本公开。

[0005] 根据本公开的一个方面,提供了一种用于目标检测的神经网络的训练方法,包括:将包含有训练目标的训练图像输入至神经网络,以得到所述训练目标的预测边界框;根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。

[0006] 其中,至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练,包括:利用所述第一损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比。

[0007] 其中,至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:确定当所述神经网络的训练轮数达到设定的阈值时参数的值,作为训练好的神经网络的参数的值。

[0008] 其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值。

[0009] 其中,根据预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,包括:对于位于所述训练图像中真实边界框内的至少一个像素点,确定第一4维向量和第二4维向量,其中,该第一4维向量中的元素分别为该像素点至所述真实边界框的上、

下、左、右边界的距离,所述第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离;根据所述第一4维向量,获得所述真实边界框的面积;根据所述第二4维向量,获得所述预测边界框的面积;根据所述第一4维向量和所述第二4维向量,获得所述预测边界框和所述真实边界框的交集面积;以及将获得的所述预测边界框的面积和所述真实边界框的面积的和与所述交集面积的差作为所述并集面积。

[0010] 其中,训练方法还包括:选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度,所述至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:利用所述第一损失函数和第二损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,以得到训练好的神经网络。

[0011] 其中,所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层;其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

[0012] 其中,训练方法还包括:对于每个输入的训练图像,还向所述输入层输入与该训练图像相同大小的真实置信度图以及训练边界框图,其中,所述训练边界框图包含所述训练图像中的每个像素点的4维向量,并且,当像素点位于所述训练图像中的真实边界框内时,该像素点对应的4维向量中的元素为该像素点距离所述真实边界框的上、下、左、右边界的距离,所述真实置信度图中的每个像素点表示所述训练图像中对应像素点属于所述训练目标的置信度。

[0013] 其中,所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0014] 根据本公开的另一个方面,提供一种用于目标检测的神经网络的训练装置,包括:处理器;存储器;和存储在所述存储器中的计算机程序指令,在所述计算机程序指令被所述处理器运行时,使得所述处理器执行以下步骤:将包含有训练目标的训练图像输入至神经网络,以得到所述训练目标的预测边界框;根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。

[0015] 其中,至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练,包括:利用所述第一损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比。

[0016] 其中,至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进

行训练包括：确定当所述神经网络的训练轮数达到设定的阈值时参数的值，作为训练好的神经网络的参数的值。

[0017] 其中，所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值。

[0018] 其中，根据预测边界框和真实边界框的交集面积与并集面积之比，获得第一损失函数，包括：对于位于所述训练图像中真实边界框内的至少一个像素点，确定第一4维向量和第二4维向量，其中，该第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离，所述第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离；根据所述第一4维向量，获得所述真实边界框的面积；根据所述第二4维向量，获得所述预测边界框的面积；根据所述第一4维向量和所述第二4维向量，获得所述预测边界框和所述真实边界框的交集面积；以及将获得的所述预测边界框的面积和所述真实边界框的面积的和与所述交集面积的差作为所述并集面积。

[0019] 其中，所述处理器还执行以下步骤：选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数，所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度，所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度，利用所述第一损失函数和第二损失函数调整所述神经网络的参数，以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比，并最小化所述第二损失函数，以得到训练好的神经网络。

[0020] 其中，所述神经网络至少包括一个输入层、 n 个中间层和两个输出层，其中， n 是大于等于2的正整数，所述两个输出层包括置信度输出层和边界框输出层，其中，在训练阶段，所述输入层配置为接收所述训练图像，所述置信度输出层配置为输出预测置信度图，该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度，所述边界框输出层配置为输出边界框预测图，所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量，其中，当像素点位于所述训练图像中的真实边界框内时，所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

[0021] 其中，所述处理器还执行以下步骤：对于每个输入的训练图像，还向所述输入层输入与该训练图像相同大小的真实置信度图以及训练边界框图，其中，所述训练边界框图包含所述训练图像中的每个像素点的4维向量，并且，当像素点位于所述训练图像中的真实边界框内时，该像素点对应的4维向量中的元素为该像素点距离所述真实边界框的上、下、左、右边界的距离，所述真实置信度图中的每个像素点表示所述训练图像中对应像素点属于所述训练目标的置信度。

[0022] 其中，所述神经网络的第 m 个中间层级联至所述置信度输出层，第 n 个中间层级联至所述边界框输出层，并且， m 小于 n 。

[0023] 根据本公开的又一个方面，提供一种用于目标检测的神经网络的训练装置，包括：预测部件，配置为将包含有训练目标的训练图像输入至神经网络，以得到所述训练目标的预测边界框；第一获取部件，配置为根据所述预测边界框和真实边界框的交集面积与并集面积之比，获得第一损失函数，所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框；以及参数调整部件，配置为至少利用所述第一损失函数调整所述神经网络的

参数,以对所述神经网络进行训练。

[0024] 其中,所述参数调整部件利用所述第一损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比。

[0025] 其中,所述参数调整部件确定当所述神经网络的训练轮数达到设定的阈值时参数的值,作为训练好的神经网络的参数的值。

[0026] 其中,所述第一损失函数为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值。

[0027] 其中,所述第一获取部件对于位于所述训练图像中真实边界框内的至少一个像素点,确定第一4维向量和第二4维向量,其中,该第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离,所述第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离;根据所述第一4维向量,获得所述真实边界框的面积;根据所述第二4维向量,获得所述预测边界框的面积;根据所述第一4维向量和所述第二4维向量,获得所述预测边界框和所述真实边界框的交集面积;以及将获得的所述预测边界框的面积和所述真实边界框的面积的和与所述交集面积的差作为所述并集面积。

[0028] 其中,训练装置还包括:第二获取部件,配置为选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度,所述参数调整部件利用所述第一损失函数和第二损失函数调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,以得到训练好的神经网络。

[0029] 其中,所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层,其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

[0030] 其中,对于每个输入的训练图像,所述预测部件还向所述输入层输入与该训练图像相同大小的真实置信度图以及训练边界框图,其中,所述训练边界框图包含所述训练图像中的每个像素点的4维向量,并且,当像素点位于所述训练图像中的真实边界框内时,该像素点对应的4维向量中的元素为该像素点距离所述真实边界框的上、下、左、右边界的距离,所述真实置信度图中的每个像素点表示所述训练图像中对应像素点属于所述训练目标的置信度。

[0031] 其中,所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0032] 本公开实施例的用于目标检测的神经网络训练方法和装置利用第一损失函数将

目标边界框回归为一个整体单元,利用目标边界框的各边界之间的高度关联性,显著增加了神经网络的目标检测精度,并能够在提高神经网络效率的同时有效适用于不同尺寸的目标的训练和检测。

[0033] 根据本公开的再一个方面,提供一种用于目标检测的神经网络的构建方法,包括:构建至少包含一个输入层、n个中间层和两个输出层的神经网络,n是大于等于2的正整数,其中,将第m个中间层级联至所述神经网络的第一支路,将第n个中间层级联至所述神经网络的第二支路,所述两个输出层中的第一输出层位于所述第一支路,第二输出层位于所述第二支路,并且,m小于n;将所述神经网络的第一支路的第一输出层配置为在用所述神经网络进行目标检测时输出置信度图的置信度输出层,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度;以及将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该检测目标的边界框。

[0034] 其中,当像素点位于所述检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为所述像素点至所述检测边界框的上、下、左、右边界的距离。

[0035] 其中,所述神经网络的第一支路上还包括:卷积层,所述卷积层配置为对所述第m个中间层输出的特征图像进行卷积,以获得第一特征图;上采样层,所述上采样层配置为对所述第一特征图进行线性插值,将所述第一特征图调整为所述检测图像的大小;以及剪裁层,所述剪裁层配置为剪裁所述调整后的第一特征图,以使得所述剪裁后的第一特征图与所述检测图像对齐,并将对齐后的第一特征图输出至所述第一支路的第一输出层,以输出所述置信度图。

[0036] 其中,所述神经网络的第二支路上还包括:卷积层,所述卷积层配置为对所述第n个中间层输出的特征图像进行卷积,以获得所述图像的第二特征图;上采样层,所述上采样层配置为对所述第二特征图进行线性插值,将所述第二特征图调整为所述检测图像的大小;剪裁层,所述剪裁层配置为剪裁所述调整后的第二特征图,以使得所述剪裁后的第二特征图与所述检测图像对齐;以及修正线性单元层,所述修正线性单元层配置为获取对齐后的第二特征图,并将所述第二特征图中的每个像素点的对应向量中的元素调整为不小于0的向量元素后输出至所述第二支路的第二输出层,以输出所述边界框检测图。

[0037] 其中,构建方法还包括:训练所述神经网络,其中,所述训练神经网络包括:将包含有训练目标的训练图像输入至所述神经网络,以得到所述训练目标的预测边界框;根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,其中,所述预测边界框为采用该神经网络在包含所述训练目标的训练图像中检测到的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。

[0038] 其中,至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:利用所述第一损失函数以及反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数对该神经网络进行训练,其中,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置

信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度。

[0039] 根据本公开的又一个方面,提供一种用于目标检测的神经网络的构建装置,包括:处理器;存储器;和存储在所述存储器中的计算机程序指令,在所述计算机程序指令被所述处理器运行时使得所述处理器执行以下步骤:构建至少包含一个输入层、n个中间层和两个输出层的神经网络,n是大于等于2的正整数,其中,将第m个中间层级联至所述神经网络的第一支路,将第n个中间层级联至所述神经网络的第二支路,所述两个输出层中的第一输出层位于所述第一支路,第二输出层位于所述第二支路,并且,m小于n;将所述神经网络的第一支路的第一输出层配置为在用所述神经网络进行目标检测时输出置信度图的置信度输出层,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度;以及将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该检测目标的边界框。

[0040] 其中,当像素点位于所述检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为所述像素点至所述检测边界框的上、下、左、右边界的距离。

[0041] 其中,所述神经网络的第一支路上还包括:卷积层,所述卷积层配置为对所述第m个中间层输出的特征图像进行卷积,以获得所述检测图像的第一特征图;上采样层,所述上采样层配置为对所述第一特征图进行线性插值,将所述第一特征图调整为所述检测图像的大小;以及剪裁层,所述剪裁层配置为剪裁所述调整后的第一特征图,以使得所述剪裁后的第一特征图与所述检测图像对齐,并将对齐后的第一特征图输出至所述第一支路的第一输出层,以输出所述置信度图。

[0042] 其中,所述神经网络的第二支路上还包括:卷积层,所述卷积层配置为对所述第n个中间层输出的特征图像进行卷积,以获得所述检测图像的第二特征图;上采样层,所述上采样层配置为对所述第二特征图进行线性插值,将所述第二特征图调整为所述检测图像的大小;剪裁层,所述剪裁层配置为剪裁所述调整后的第二特征图,以使得所述剪裁后的第二特征图与所述检测图像对齐;以及修正线性单元层,所述修正线性单元层配置为获取对齐后的第二特征图,并将所述第二特征图中的每个像素点的对应向量中的元素调整为不小于0的向量元素后输出至所述第二支路的第二输出层,以输出所述边界框检测图。

[0043] 其中,所述处理器还执行以下步骤:训练所述神经网络,具体包括:将包含有训练目标的训练图像输入至所述神经网络,以得到所述训练目标的预测边界框;根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,其中,所述预测边界框为采用该神经网络在包含所述训练目标的训练图像中检测到的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。

[0044] 其中,至少利用所述第一损失函数调整所述神经网络的参数以对所述神经网络进行训练包括:利用所述第一损失函数以及反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数对该神经网络进行训练,其中,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置

信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度。

[0045] 根据本公开的又一个方面,提供一种用于目标检测的神经网络的构建装置,包括:构建部件,配置为构建至少包含一个输入层、 n 个中间层和两个输出层的神经网络, n 是大于等于2的正整数,其中,将第 m 个中间层级联至所述神经网络的第一支路,将第 n 个中间层级联至所述神经网络的第二支路,所述两个输出层中的第一输出层位于所述第一支路,第二输出层位于所述第二支路,并且, m 小于 n ;第一配置部件,配置为将所述神经网络的第一支路的第一输出层配置为在用所述神经网络进行目标检测时输出置信度图的置信度输出层,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度;第二配置部件,配置为将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该检测目标的边界框。

[0046] 其中,当像素点位于所述检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为所述像素点至所述检测边界框的上、下、左、右边界的距离。

[0047] 其中,所述构建部件在所述神经网络的第一支路上还构建有:卷积层,所述卷积层配置为对所述第 m 个中间层输出的特征图像进行卷积,以获得所述检测图像的第一特征图;上采样层,所述上采样层配置为对所述第一特征图进行线性插值,将所述第一特征图调整为所述检测图像的大小;以及剪裁层,所述剪裁层配置为剪裁所述调整后的第一特征图,以使得所述剪裁后的第一特征图与所述检测图像对齐,并将对齐后的第一特征图输出至所述第一支路的第一输出层,以输出所述置信度图。

[0048] 其中,所述构建部件在所述神经网络的第二支路上还构建有:第二卷积层,所述卷积层配置为对所述第 n 个中间层输出的特征图像进行卷积,以获得所述图像的第二特征图;上采样层,所述上采样层配置为对所述第二特征图进行线性插值,将所述第二特征图调整为所述检测图像的大小;剪裁层,所述剪裁层配置为剪裁所述调整后的第二特征图,以使得所述剪裁后的第二特征图与所述检测图像对齐;以及修正线性单元层,所述修正线性单元层配置为获取对齐后的第二特征图,并将所述第二特征图中的每个像素点的对应向量中的元素调整为不小于0的向量元素后输出至所述第二支路的第二输出层,以输出所述边界框检测图。

[0049] 其中,构建装置还包括:训练部件,配置为通过以下方式训练所述神经网络:将包含有训练目标的训练图像输入至所述神经网络,以得到所述训练目标的预测边界框;根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,其中,所述预测边界框为采用该神经网络在包含所述训练目标的训练图像中检测到的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。

[0050] 其中,所述训练部件利用所述第一损失函数以及反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数对该神经网络进行训练,其中,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置

信度。

[0051] 根据本公开的又一个方面,提供一种用于目标检测的神经网络,包括一个输入层、 n 个中间层和两个输出层, n 是大于等于2的正整数,其中,第 m 个中间层级联至所述神经网络的第一支路,第 n 个中间层级联至所述神经网络的第二支路,所述两个输出层中的第一输出层位于所述第一支路,第二输出层位于所述第二支路,并且, m 小于 n ;所述神经网络的第一支路的第一输出层配置为在用所述神经网络进行目标检测时输出置信度图,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述神经网络的第二支路的第二输出层配置为在利用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该目标的边界框。

[0052] 其中,当像素点位于所述检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为所述像素点至所述检测边界框的上、下、左、右边界的距离。

[0053] 其中,所述神经网络的第一支路上还包括:卷积层,所述卷积层配置为对所述第 m 个中间层输出的特征图像进行卷积,以获得所述检测图像的第一特征图;上采样层,所述上采样层配置为对所述第一特征图进行线性插值,将所述第一特征图调整为所述检测图像的大小;以及剪裁层,所述剪裁层配置为剪裁所述调整后的第一特征图,以使得所述剪裁后的第一特征图与所述检测图像对齐,并将对齐后的第一特征图输出至所述第一支路的第一输出层,以输出所述置信度图。

[0054] 其中,所述神经网络的第二支路上还包括:卷积层,所述卷积层配置为对所述第 n 个中间层输出的特征图像进行卷积,以获得所述检测图像的第二特征图;上采样层,所述上采样层配置为对所述第二特征图进行线性插值,将所述第二特征图调整为所述检测图像的大小;剪裁层,所述剪裁层配置为剪裁所述调整后的第二特征图,以使得所述剪裁后的第二特征图与所述检测图像对齐;以及修正线性单元层,所述修正线性单元层配置为获取对齐后的第二特征图,并将所述第二特征图中的每个像素点的对应向量中的元素调整为不小于0的向量元素后输出至所述第二支路的第二输出层,以输出所述边界框检测图。

[0055] 其中,所述神经网络是利用第一损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框。

[0056] 其中,所述神经网络是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述第二损失函数反映训练图像中各像素点的预测置信度与真实置信度之间的差异,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度。

[0057] 根据本公开实施例中的神经网络构建方法和装置得到的神经网络能够在边界框回归时,使得边界框输出层分支比置信度输出层分支获得更大的感受野,使得边界框输出

层分支采用自下而上的策略从置信度图中提取边界框,从而有效提高了神经网络的训练和检测效率。

[0058] 根据本公开的另一个方面,提供一种基于神经网络的目标检测方法,包括:获取包含检测目标的检测图像;利用神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含该检测图像中各像素点的对应向量,所述对应向量表示与其对应的像素点与对应检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框;在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状;在拟合得到的几何形状内选取预定数量的像素点;对于每个选取的像素点,确定所述边界框检测图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框;将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个检测边界框。

[0059] 其中,当像素点位于对应的检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为与其对应的像素点至对应的检测边界框的上、下、左、右边界的距离。

[0060] 其中,将各个所述检测边界框进行合并包括:利用非最大抑制方法将各个所述检测边界框进行合并。

[0061] 其中,在所述置信度图中基于各个像素点的置信度通过形状拟合得到至少一个几何形状包括:确定置信度大于预定阈值的多个像素点;基于所述多个像素点拟合得到所述至少一个几何形状。

[0062] 其中,所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括用于输出所述置信度图的置信度输出层和用于输出所述边界框检测图的边界框输出层。

[0063] 其中,所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0064] 其中,所述神经网络是利用第一损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框。

[0065] 其中,所述神经网络是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述第二损失函数反映预测置信度与真实置信度之间的差异,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度。

[0066] 根据本公开的再一个方面,提供一种基于神经网络的目标检测装置,包括:处理器;存储器;和存储在所述存储器中的计算机程序指令,在所述计算机程序指令被所述处理器运行时使得所述处理器执行以下步骤:获取包含检测目标的检测图像;利用神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含

该检测图像中各像素点的对应向量,各对应向量表示与其对应的像素点与对应检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框;在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状;在拟合得到的几何形状内选取预定数量的像素点;对于每个选取的像素点,确定所述边界框检测图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框;将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个检测边界框。

[0067] 其中,当像素点位于对应的检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为与其对应的像素点至对应的检测边界框的上、下、左、右边界的距离。

[0068] 其中,将各个所述检测边界框进行合并包括:利用非最大抑制方法将各个所述检测边界框进行合并。

[0069] 其中,在所述置信度图中基于各个像素点的置信度通过形状拟合得到至少一个几何形状包括:确定置信度大于预定阈值的多个像素点;基于所述多个像素点拟合得到所述至少一个几何形状。

[0070] 其中,所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括用于输出所述置信度图的置信度输出层和用于输出所述边界框检测图的边界框输出层。

[0071] 其中,所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0072] 其中,所述神经网络是利用第一损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框。

[0073] 其中,所述神经网络是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述第二损失函数反映预测置信度与真实置信度之间的差异,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度。

[0074] 根据本公开的再一个方面,提供一种基于神经网络的目标检测装置,包括:图像获取部件,配置为获取包含检测目标的检测图像;检测部件,配置为利用神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含该检测图像中各像素点的对应向量,所述对应向量表示与其对应的像素点与对应检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框;拟合部件,配置为在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状;选取部件,配置为在拟合得到的几何形状内选取预定数量

的像素点;边界框获取部件,配置为对于每个选取的像素点,确定所述边界框检测图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框;合并部件,配置为将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个检测边界框。

[0075] 其中,当像素点位于对应的检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为与其对应的像素点至对应的检测边界框的上、下、左、右边界的距离。

[0076] 其中,所述合并部件利用非最大抑制方法将各个所述检测边界框进行合并。

[0077] 其中,所述拟合部件确定置信度大于预定阈值的多个像素点;基于所述多个像素点拟合得到所述至少一个几何形状。

[0078] 其中,所述神经网络至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括用于输出所述置信度图的置信度输出层和用于输出所述边界框检测图的边界框输出层。

[0079] 其中,所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0080] 其中,所述神经网络是利用第一损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框

[0081] 其中,所述神经网络是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述第二损失函数反映预测置信度与真实置信度之间的差异,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述预测置信度为采用该神经网络预测出的所述训练图像中的某一个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度。

[0082] 由此可见,本公开实施例的基于神经网络的目标检测方法和装置能够得到精确和有效的目标注位结果,检测效率高,并且不必改变输入图像的尺寸,适应性更强,适用于不同尺寸的目标,尤其对于小尺寸目标进行检测也能够取得较好的效果。此外由于在本公开实施例中能够针对每个像素点均生成对应检测边界框的对应向量,因此能够针对位于不同检测目标中的像素点获得相应的不同的检测边界框,因此本公开实施例能够同时处理检测图像中的多个检测目标。

[0083] 在根据本公开实施例的技术方案中,采用了能够令神经网络更快、更稳定地收敛的损失函数,从而使得对目标的定位更加精确,目标检测的检测适应性更强并且效率更高;此外,本公开实施例的技术方案中的神经网络结构采用了具有两个分支的神经网络,并且自不同的中间层分别级联至出层,以进一步提高了目标检测的精确度。

附图说明

[0084] 通过结合附图对本公开实施例进行更详细的描述,本公开的上述以及其它目的、特征和优势将变得更加明显。附图用来提供对本公开实施例的进一步理解,并且构成说明

书的一部分,与本公开实施例一起用于解释本公开,并不构成对本公开的限制。在附图中,相同的参考标号通常代表相同部件或步骤。

[0085] 图1示意性地示出了根据本公开实施例的用于目标检测的神经网络的训练方法的流程图;

[0086] 图2示意性地示出了根据本公开实施例的用于目标检测的神经网络的训练方法中真实边界框和预测边界框的示意图;

[0087] 图3示意性地示出了根据本公开实施例的用于目标检测的神经网络的训练装置的示意性框图;

[0088] 图4示意性地示出了根据本公开实施例的用于目标检测的神经网络的训练装置的示例性结构框图;

[0089] 图5示意性地示出了根据本公开实施例的用于目标检测的神经网络的构建方法的流程图;

[0090] 图6示意性地示出了根据本公开实施例的用于目标检测的神经网络的结构示意图;

[0091] 图7示意性地示出了根据本公开实施例的用于目标检测的神经网络的构建装置的示意性框图;

[0092] 图8示意性地示出了根据本公开实施例的用于目标检测的神经网络的构建装置的示意性框图;

[0093] 图9示意性地示出了根据本公开实施例的基于神经网络的目标检测方法的流程图;

[0094] 图10示意性地示出了根据本公开实施例的基于神经网络的目标检测装置的示意性框图;

[0095] 图11示意性地示出了根据本公开实施例的基于神经网络的目标检测装置的示意性框图;

[0096] 图12示出了根据本公开实施例的目标检测方法和装置进行人脸检测的示例;

[0097] 图13(a)、(b)分别示出了利用本公开实施例的损失函数训练的神经网络与利用 L_2 损失函数训练的神经网络进行目标检测的未命中率(miss rate)和查全率(recall rate)的比较;

[0098] 图14示出了针对检测图像的不同像素利用 L_2 损失函数和采用本公开实施例的第一损失函数训练的神经网络进行人脸目标检测的结果比较示意图;

[0099] 图15示出本公开实施例的人脸目标检测方法与其他检测方法的横向比较示意图。

具体实施方式

[0100] 为了使得本公开的目的、技术方案和优点更为明显,下面将参照附图详细描述根据本公开的示例实施例。显然,所描述的实施例仅仅是本公开的一部分实施例,而不是本公开的全部实施例,应理解,本公开不受这里描述的示例实施例的限制。基于本公开中描述的实施例,本领域技术人员在没有付出创造性劳动的情况下所得到的所有其它实施例都应落入本公开的保护范围之内。

[0101] 首先对本公开涉及的基本概念以及根据本公开实施例的用于目标检测的神经网络

络的训练与构建的基本思想进行简单介绍。

[0102] 众所周知,利用神经网络进行目标检测可以分为两个阶段,即训练阶段和检测阶段。所谓训练阶段是指首先需要利用训练目标对神经网络进行训练,以调整神经网络的参数;所谓检测阶段是指随后可以利用训练好的神经网络对包含检测目标的检测图像进行目标检测。

[0103] L_2 损失函数一种常用的目标检测损失函数,其分离地考虑像素点到表示检测到的目标的位置和大小的目标边界框的上下左右边界的距离,并由此对神经网络的各参数进行调整。本发明人认识到,目标边界框中的各条边界并不是分立的多个变量,而是具有关联性的整体,因此针对用于目标检测的神经网络,可以借助损失函数将获得的边界框回归(regress)为一个整体单元,由此可以增强训练得到的神经网络的鲁棒性,进而提高利用该神经网络进行目标检测的检测精度和检测效率。

[0104] 另外,可以将神经网络创建为具有两个分支,其两个输出层分别连接两个不同的中间层。其中,第一输出层所连接的中间层在第二输出层所连接的中间层之前。通过这种设置,能够在神经网络的训练和检测过程中利用第一输出层的输出结果推得第二输出层,并结合第一输出层和第二输出层的预测结果,更精确地进行目标检测。

[0105] 在下文中,为了便于描述,以目标对象是人脸为例进行说明。应该理解的是,本公开的检测的目标对象不局限于人脸。

[0106] 其中,在本公开实施例中,位于边界框“内”的像素点可以包括位于边界框边界上的像素点。

[0107] 下面将参考图1对根据本公开实施例的用于目标检测的神经网络的训练方法进行详细的描述。图1示意性地示出了根据本公开实施例的用于目标检测的神经网络的训练方法的流程图。

[0108] 如图1所示,训练方法100包括步骤S101:将包含有训练目标的训练图像输入至神经网络,以得到所述训练目标的预测边界框。

[0109] 在神经网络的训练过程中,首先可将包含有至少一个训练目标的训练图像输入至神经网络,神经网络可以根据训练目标的相关特征进行特征提取、通过诸如SIFT特征以及HOG特征等本领域中常用的目标检测方法进行检测,以生成训练目标的预测边界框。训练图像中可以预先标注训练目标的真实边界框的位置和尺寸等,例如,以边界框预先标注训练目标。例如,以训练目标是人脸为例,可以通过SIFT特征提取以及HOG特征提取等方法提取人脸特征,再使用SVM分类器、Adaboost等方法进行分类来检测人脸,以得到所述训练人脸的预测边界框。

[0110] 对于所输入的训练图像中的每一个或者某些像素点,都可以利用神经网络预测出一个针对训练目标的预测边界框。为了在获取尽可能多的有效的像素点的预测结果的基础上尽量减小运算量,可选地,可以针对位于训练图像中的真实边界框内的每一个像素点来预测相应的预测边界框。所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框。

[0111] 在步骤S102中,根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框。

[0112] 其中,对于每个输入的、预先标注了训练目标真实边界框位置和尺寸的训练图像,还可以向神经网络输入与该训练图像相同大小的训练边界框图,所述神经网络输出边界框预测图。其中,所述训练边界框图包含所述训练图像中的每个像素点的4维向量,该4维向量表明对应的像素点与真实边界框的位置关系。当像素点位于所述真实边界框内时,该像素点对应的4维向量中的元素分别表明该像素点距离真实边界框的上、下、左、右边界的距离。对于位于所述真实边界框之外的像素点,可以用0向量表示。所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。根据神经网络所接收的同一个训练目标的训练边界框图及其获得的相应的边界框预测图,可以得到位于训练图像中的真实边界框内的每个像素点的第一4维向量和第二4维向量。第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离。第二4维向量中的元素分别为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。根据所述第一4维向量和所述第二4维向量,可以计算出预测边界框和真实边界框的交集面积与并集面积之比,即预测边界框和真实边界框相交部分的面积除以预测边界框和真实边界框并集所占面积的值,并据此获得第一损失函数。

[0113] 在本公开的一个实施例中,当边界框为标示目标的方形边界框时,真实边界框和预测边界框均可以基于对应像素的4维向量来表示。图2示出了根据本公开实施例的用于目标检测的神经网络的训练方法中真实边界框和预测边界框的示意图。如图2所示,对于位于训练图像中真实边界框内的任一个像素点,可以基于训练边界框图中的对应像素点的向量和边界框预测图中的对应像素点的向量,分别确定第一4维向量和第二4维向量。第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,真实边界框位于图2左下角,第一4维向量可以记为 $\tilde{x} = (\tilde{x}_t, \tilde{x}_b, \tilde{x}_l, \tilde{x}_r)$;第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,与该像素点对应的预测边界框位于图2右上角,第二4维向量可以记为 $x = (x_t, x_b, x_l, x_r)$ 。此时,在经过符号处理以保证各向量中的每一个分量均不小于0之后,可根据第一4维向量获得如等式(1)所示的真实边界框的面积:

$$[0114] \quad \tilde{X} = (\tilde{x}_t + \tilde{x}_b) \times (\tilde{x}_l + \tilde{x}_r) \quad \dots(1)$$

[0115] 根据第二4维向量获得如等式(2)所示的对应于该像素点的预测边界框的面积:

$$[0116] \quad X = (x_t + x_b) \times (x_l + x_r) \quad \dots(2)$$

[0117] 根据第一4维向量和第二4维向量获得如等式(3)所示的对应于所述预测边界框和所述真实边界框的交集面积I:

$$[0118] \quad I_h = \min(x_t, \tilde{x}_t) + \min(x_b, \tilde{x}_b),$$

$$[0119] \quad I_w = \min(x_l, \tilde{x}_l) + \min(x_r, \tilde{x}_r),$$

$$[0120] \quad I = I_h \times I_w \quad \dots(3)$$

[0121] 可以将获得的所述预测边界框的面积和所述真实边界框的面积的和与所述交集面积的差作为所述并集面积,由此确定第一损失函数。例如,在图2的实施例中,并集面积可以通过等式(4)来表示:

$$[0122] \quad U = X + \tilde{X} - I \quad \dots(4)$$

[0123] 经过上述运算,可以求得交集面积和并集面积的比值 I/U ,并由此确定第一损失函数。例如,在本公开的一个优选实施例中,第一损失函数可以表示为所述预测边界框和所述真实边界框的交集面积与并集面积之比的自然对数的负值,即, $L=-\ln(I/U)$ 。当然,这仅仅是一个示例,第一损失函数也可以是该比值 I/U 的其他适当函数,例如对数函数、双曲函数等。

[0124] 以上预测边界框和真实边界框的表示方法仅仅是当边界框为方形边界框时的一种示例表示方式,而并非是对本公开的限制,本领域技术人员也可以采用其他的表示方式。比如,当边界框为方形边界框时,还可以利用边界框的中心点坐标位置、长度和宽度的参数进行表示,同样可以计算出预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。此外,当边界框不为方形边界框时同理。例如,边界框为圆形边界框时,可以利用圆心坐标和半径来表示,并利用相似的方法计算预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。

[0125] 在步骤S103中,至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。

[0126] 在此步骤中,可以利用步骤S102中所计算的第一损失函数对神经网络进行训练,调整神经网络的参数以尽量最大化预测边界框和真实边界框交集面积与并集面积之比,即令 I/U 尽可能大,并将此时的神经网络的参数的值作为训练好的神经网络参数的值。具体的,在一个实施例中,当与 I/U 的比值相关的第一损失函数表示为 $L=-\ln(I/U)$ 时,可以调整神经网络的参数以令第一损失函数尽可能小,并将第一损失函数收敛时的神经网络的各项参数的值作为训练好的神经网络的各项参数的值,由此得到训练好的神经网络。在另一个实施例中,还可以对神经网络设定训练轮数的阈值,并确定当所述神经网络的训练轮数达到设定的阈值时各项参数的值,作为训练好的神经网络的各项参数的值。

[0127] 需要说明的是,当对于多个像素点均计算了其预测边界框和真实边界框的交集面积和并集面积的比时,在该步骤中可以利用所述第一损失函数调整所述神经网络的参数,从而使得各像素点对应的预测边界框和真实边界框的交并面积比均为最大化;或者使得各像素点对应的预测边界框和真实边界框的交并面积比之和为最大化,由此得到训练好的神经网络的各个参数。

[0128] 能够理解,在如上根据输入的训练图像预测所述预测边界框并计算损失函数之后,需要根据所计算的损失函数进行后向传导反馈以训练神经网络。该后向传导反馈是神经网络训练中的常规处理,而并非是本发明的关键所在。此处,仅仅是为了说明的完整,对一种可能的后向传导反馈进行简单的描述。

[0129] 例如,针对本公开实施例中边界框为方形边界框,且第一损失函数表示为 $L=-\ln(I/U)$ 时,可以如下计算后向传导的梯度:

$$[0130] \quad \frac{\partial L}{\partial x} = \frac{I(\nabla_x X - \nabla_x I) - U \nabla_x I}{U^2 \times \frac{I}{U}} = \frac{I(\nabla_x X - \nabla_x I) - U \nabla_x I}{UI} \quad \dots(5)$$

$$= \frac{1}{U} \nabla_x X - \frac{U+I}{UI} \nabla_x I$$

[0131] 其中,

$$\frac{\partial X}{\partial x_t(\text{or } \partial x_b)} = x_t + x_r,$$

$$\frac{\partial X}{\partial x_t(\text{or } \partial x_r)} = x_t + x_b, \quad \dots(6)$$

[0132]

$$\frac{\partial I}{\partial x_t(\text{or } \partial x_b)} = \begin{cases} I_w, & \text{if } x_t < \tilde{x}_t(\text{or } x_b < \tilde{x}_b) \\ 0, & \text{otherwise,} \end{cases}$$

$$\frac{\partial I}{\partial x_t(\text{or } \partial x_r)} = \begin{cases} I_w, & \text{if } x_t < \tilde{x}_t(\text{or } x_r < \tilde{x}_r) \\ 0, & \text{otherwise.} \end{cases}$$

[0133] 其中, $\nabla_x X$ 与预测边界框的面积有关,是第一损失函数偏导中的正值, $\nabla_x I$ 与相交部分有关,是第一损失函数偏导中的负值。由此可见,为了使 I/U 尽可能大,也即第一损失函数 $L = -\ln(I/U)$ 尽可能小,其偏导值倾向于使得相交部分(交集)尽量大而并集部分尽量小。而当相交部分(交集)的面积正好等于并集面积时,即为本实施例中第一损失函数所训练的神经网络的最佳情况。

[0134] 可选的,在利用第一损失函数训练神经网络的基础上,还可以选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数来共同训练此神经网络,以得到更好的训练效果。所述预测置信度为采用该神经网络预测出的所述训练图像中的一个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的该像素点属于所述训练目标的置信度。其中,针对图像中的每个像素点,真实置信度可以表示为1(或正值)和0(或负值),分别用以表示该像素点属于所述训练目标和不属于所述训练目标。在神经网络的训练过程中,可以利用所述第一损失函数和第二损失函数共同来调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,从而得到训练好的神经网络。在一个实施例中,可以基于所述第一损失函数和所述第二损失函数的加权和调整所述神经网络的参数,以训练所述神经网络。其中,在利用第二损失函数训练神经网络的过程中,一方面,可以对于训练图像中的每个像素点,或者至少一个像素点计算第二损失函数,并使所述像素点的预测置信度与真实置信度之间的差异最小化;另一方面,可以利用第二损失函数训练神经网络以使得比如对于各个像素点均使得预测置信度与真实置信度之间的差异都最小,或者使得各个像素点的预测置信度与真实置信度之间的差异之和最小化,由此得到训练好的神经网络。

的各个参数。

[0135] 可选的,该实施例所训练的神经网络可以包含两个输出层。具体的,所述神经网络可以至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层,其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

[0136] 其中,为了获得训练图像的预测置信度和边界框预测信息,可以对于每个输入的训练图像,还向所述输入层输入与该训练图像相同大小的真实置信度图以及训练边界框图,其中,所述训练边界框图包含所述训练图像中的每个像素点的4维向量,并且,当像素点位于所述训练图像中的真实边界框内时,该像素点对应的4维向量中的元素为该像素点距离所述真实边界框的上、下、左、右边界的距离,所述真实置信度图中的每个像素点表示训练图像中对应像素点属于所述训练目标的置信度。

[0137] 可选的,所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0138] 在一个实施例中,在对上述神经网络进行训练的过程中,可以根据所述训练边界框图和所述神经网络的边界框输出层输出的边界框预测图,计算所述第一损失函数;并且,可以根据所述真实置信度图以及所述神经网络的置信度输出层输出的预测置信度图,计算所述第二损失函数。根据所述第一损失函数和所述第二损失函数的加权和调整上述神经网络的参数,以实现与所述神经网络的训练。

[0139] 需要说明的是,以上对根据本公开实施例的训练方法所应用的神经网络包含两个输出层的情形进行了说明,应当理解,这仅仅是一个示例,而并非是对本公开的限制。根据本公开实施例的神经网络的训练方法对于所应用的神经网络没有限制,而是可以适用于任何适当的神经网络,例如具有单个输出层的常见神经网络等。

[0140] 由此可见,本公开实施例利用第一损失函数将目标边界框回归为一个整体单元,有效避免了 L_2 损失函数将边界框四边作为四个独立变量考虑导致预测精度不高的问题,显著增加了神经网络的目标检测精度。此外,根据 L_2 损失函数,为了平衡不同尺寸的边界框,需要将输入图像调整为特定的尺寸,因此会使得训练和检测效率低下,并且无法适应不同尺寸的目标,而本公开实施例的上述方法计算的是预测边界框和真实边界框的交并面积比,与输入图像的具体大小并无关系,因此本公开实施例无需对输入图像的尺寸进行调整,能够在提高神经网络效率的同时有效适用于不同尺寸的训练和检测目标。

[0141] 下面,将参考图3来描述根据本公开实施例的用于目标检测的神经网络的训练装置。图3示出了可用于实现根据本公开实施例的神经网络的训练装置的计算设备的框图。

[0142] 如图3所示,计算设备300包括一个或多个处理器302、存储器304、图像采集装置306和输出装置308,这些组件通过总线系统310和/或其它形式的连接机构(未示出)互连。应当注意,图3所示的计算设备300的组件和结构只是示例性的,而非限制性的,根据需要,

计算设备300也可以具有其他组件和结构。

[0143] 处理器302可以是中央处理单元 (CPU) 或者具有数据处理能力和/或指令执行能力的其它形式的处理单元,并且可以控制计算设备300中的其它组件以执行期望的功能。

[0144] 存储器304可以包括一个或多个计算机程序产品,所述计算机程序产品可以包括各种形式的计算机可读存储介质,例如易失性存储器和/或非易失性存储器。所述易失性存储器例如可以包括随机存取存储器 (RAM) 和/或高速缓冲存储器 (cache) 等。所述非易失性存储器例如可以包括只读存储器 (ROM)、硬盘、闪存等。在所述计算机可读存储介质上可以存储一个或多个计算机程序指令,处理器302可以运行所述程序指令,以实现以下步骤:将包含有训练目标的训练图像输入至神经网络,以得到所述训练目标的预测边界框;根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。在所述计算机可读存储介质中还可以存储各种应用程序和各种数据,例如输入的训练图像、第一损失函数、预测边界框的面积、真实边界框的面积、各像素点的对应向量、各像素的预测置信度和真实置信度等等。

[0145] 图像采集装置306可以用于采集带有训练目标的训练图像,并且将所拍摄的训练图像存储在存储器304中以供其它组件使用。当然,也可以利用其他图像采集设备采集所述训练图像,并且将采集的训练图像发送给计算设备300。在这种情况下,可以省略图像采集装置306。

[0146] 输出装置308可以向外部(例如用户)输出各种信息,例如图像信息、训练结果,并且可以包括显示器、扬声器、投影仪、网卡等中的一个或多个。

[0147] 下面,将参考图4来描述根据本公开实施例的用于目标检测的神经网络的训练装置。图4示出了根据本公开实施例的用于目标检测的神经网络的训练装置400的示例性结构框图。如图4所示,该训练装置400可以包括预测部件410、第一获取部件420和参数调整部件430,所述各个单元可分别执行上文中结合图1和图2描述的神经网络的训练方法的各个步骤/功能。以下仅对该神经网络的训练装置400的各单元的主要功能进行描述,而省略以上已经描述过的细节内容。

[0148] 预测部件410配置为将包含有训练目标的训练图像输入至神经网络,以得到所述训练目标的预测边界框。预测部件410可以将包含有至少一个训练目标的训练图像输入至神经网络,神经网络可以根据训练目标的相关特征进行特征提取,以生成训练目标的预测边界框。

[0149] 对于所输入的训练图像中的每一个或者某些像素点,都可以利用神经网络预测出一个针对训练目标的预测边界框。为了在获取尽可能多的有效的像素点的预测结果的基础上尽量减小运算量,可选地,可以针对位于真实边界框内的每一个像素点来预测相应的预测边界框。所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框。

[0150] 第一获取部件420配置为根据所述预测边界框和真实边界框交集面积与并集面积之比,获得第一损失函数,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框。

[0151] 其中,对于每个输入的、预先标注了训练目标真实边界框位置和尺寸的训练图像,预测部件410还可以向神经网络输入与该训练图像相同大小的训练边界框图,所述神经网

络输出边界框预测图。其中,所述训练边界框图包含所述训练图像中的每个像素点的4维向量,该4维向量表明对应的像素点与真实边界框的位置关系。当像素点位于所述真实边界框内时,该像素点对应的4维向量中的元素分别表明该像素点距离真实边界框的上、下、左、右边界的距离。对于位于所述真实边界框之外的像素点,可以用0向量表示。所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。第一获取部件420根据神经网络所接收的同一个训练目标的训练边界框图及其获得的相应的边界框预测图,可以得到位于训练图像中的真实边界框内的每个像素点的第一4维向量和第二4维向量。第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离。第二4维向量中的元素分别为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。根据所述第一4维向量和所述第二4维向量,可以计算出预测边界框和真实边界框的交集面积与并集面积之比,即预测边界框和真实边界框相交部分的面积除以预测边界框和真实边界框共同所占面积的值,并据此获得第一损失函数。

[0152] 如上所述,预测部件410对于所输入的训练图像中的每一个像素点,都可以利用神经网络预测出一个针对训练目标的预测边界框。为了在获取尽可能多的有效的像素点的预测结果的基础上尽量减小运算量,优选地,预测部件410可以针对位于训练图像中的真实边界框内的每一个像素点来预测相应的预测边界框。

[0153] 在本公开的一个实施例中,当边界框为标示目标的方形边界框时,第一获取部件420可将真实边界框和预测边界框基于对应像素的4维向量来表示。以图2为例,对于位于训练图像中真实边界框内的任一个像素点(该像素点可以包括位于真实边界框边界上的像素点),可以基于训练边界框图中的对应像素点的向量和边界框预测图中的对应像素点的向量,分别确定第一4维向量和第二4维向量。第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,真实边界框位于图2左下角,第一4维向量可以记为 $\tilde{x} = (\tilde{x}_t, \tilde{x}_b, \tilde{x}_l, \tilde{x}_r)$;第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,与该像素点对应的预测边界框位于图2右上角,第二4维向量可以记为 $x = (x_t, x_b, x_l, x_r)$ 。如前所述,第一获取部件420可以求得交集面积和并集面积的比值 I/U 并获得第一损失函数。在本公开的一个优选实施例中,第一获取部件420可将第一损失函数表示为 $L = -\ln(I/U)$ 。

[0154] 以上第一获取部件420对预测边界框和真实边界框的表示方法仅仅是当边界框为方形边界框时的一种示例表示方式,而并非是对本公开的限制,本领域技术人员也可以采用其他的表示方式。比如,当边界框为方形边界框时,还可以利用边界框的中心点坐标位置、长度和宽度的参数进行表示,同样可以计算出预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。此外,当边界框不为方形边界框时也同理。例如,边界框为圆形边界框时,可以利用圆心坐标和半径来表示,并利用相似的方法计算预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。

[0155] 最后,参数调整部件430至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。

[0156] 参数调整部件430在利用第一损失函数对神经网络进行训练时,可以调整神经网络

网络的参数以尽量最大化预测边界框和真实边界框的交集面积与并集面积之比,即令 I/U 尽可能大,并将此时的神经网络的参数的值作为训练好的神经网络的参数的值。具体的,在一个实施例中,当与 I/U 的比值相关的第一损失函数表示为 $L=-\ln(I/U)$ 时,参数调整部件430可以调整神经网络的参数使得第一损失函数尽可能小,并将第一损失函数收敛时的神经网络各参数的值作为训练好的神经网络的参数的值,以得到训练好的神经网络。在另一个实施例中,参数调整部件430还可以对神经网络设定训练轮数的阈值,以确定当所述神经网络的训练轮数达到设定的阈值时各参数的值,作为训练好的神经网络的参数的值。

[0157] 在具体实施过程中,神经网络的训练过程不仅包括根据输入的训练图像预测所述预测边界框并计算损失函数的过程,还包括前述的根据所计算的损失函数进行后向梯度传导反馈以训练神经网络的过程。

[0158] 此外,在参数调整部件430利用第一损失函数训练神经网络的基础上,训练装置400还可以包括第二获取部件(未示出),配置为选取反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数来共同训练此神经网络,以得到更好的训练效果。所述预测置信度为采用该神经网络预测出的所述训练图像中的每个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的每个像素点属于所述训练目标的置信度。其中,针对图像中的每个像素点,真实置信度可以表示为1(或正值)和0(或负值),分别用以表示该像素点落在训练目标中和没有落在训练目标中。参数调整部件430在对神经网络的训练过程中,可以利用所述第一损失函数和第二损失函数共同来调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,从而得到训练好的神经网络。在一个实施例中,可以基于所述第一损失函数和所述第二损失函数的加权和来调整所述神经网络的参数,以训练所述神经网络。其中,在利用第二损失函数训练神经网络的过程中,一方面,可以对于训练图像中的每个像素点,或者至少一个像素点计算第二损失函数,并使所述像素点的预测置信度与真实置信度之间的差异最小化;另一方面,可以比如对于各个像素点均使得预测置信度与真实置信度之间的差异都最小,或者也可以使各个像素点的预测置信度与真实置信度之间的差异之和最小化。

[0159] 可选地,所训练的像素级目标检测的神经网络可以包含两个输出层。具体地,所述神经网络可以至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括置信度输出层和边界框输出层,其中,在训练阶段,所述输入层配置为接收所述训练图像,所述置信度输出层配置为输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,所述边界框输出层配置为输出边界框预测图,所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

[0160] 其中,为了获得训练图像的预测置信度和边界框预测信息,预测部件410可以对于每个输入的训练图像,还向所述输入层输入与该训练图像相同大小的真实置信度图以及训练边界框图,其中,所述训练边界框图包含所述训练图像中的每个像素点的4维向量,并且,当像素点位于所述训练图像中的真实边界框内时,该像素点对应的4维向量中的元素为该

像素点距离所述真实边界框的上、下、左、右边界的距离,所述真实置信度图中的每个像素点表示训练图像中对应像素点属于所述训练目标的置信度。并且,在本公开的一个实施例中,所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0161] 在一个实施例中,在对上述神经网络进行训练的过程中,所述第一获取部件420可以根据所述训练边界框图和所述神经网络的边界框输出层输出的边界框预测图,计算所述第一损失函数;并且,所述第二获取部件(未示出)可以根据所述真实置信度图以及所述神经网络的置信度输出层输出的预测置信度图,计算所述第二损失函数。所述参数调整部件430根据所述第一损失函数和所述第二损失函数的加权和调整上述神经网络的参数,以实现对所训练神经网络的训练。

[0162] 需要说明的是,以上对根据本公开实施例的训练装置所训练的神经网络包含两个输出层的情形进行了说明,应当理解,这仅仅是一个示例,而并非是对本公开的限制。根据本公开实施例的神经网络的训练装置对于所应用的神经网络没有限制,而是可以适用于任何适当的神经网络,例如具有单个输出层的常见神经网络等。

[0163] 由此可见,本公开实施例的神经网络训练装置利用第一损失函数将目标边界框回归为一个整体单元,有效避免了现有技术中将边界框四边作为四个独立变量考虑导致预测精度不高的问题,显著增加了神经网络的目标检测精度。此外,由于现有技术的方法中为了平衡不同尺寸的边界框,需要将输入图像调整为特定的尺寸,因此会使得训练和检测效率低下,并且无法适应不同尺寸的目标,而本公开实施例的上述训练装置能够在提高神经网络效率的同时能够有效适用于不同尺寸的训练和检测目标。

[0164] 下面将参考图5对根据本公开实施例的用于目标检测的神经网络的构建方法进行详细的描述。图5示意性地示出了根据本公开实施例的用于目标检测的神经网络的构建方法的流程图。

[0165] 如图5所示,在步骤S501,构建至少包含一个输入层、 n 个中间层和两个输出层的神经网络, n 是大于等于2的正整数,其中,将第 m 个中间层级联至所述神经网络的第一支路,将第 n 个中间层级联至所述神经网络的第二支路,所述两个输出层中的第一输出层位于所述第一支路,第二输出层位于所述第二支路,并且, m 小于 n 。

[0166] 在本步骤中,可以将神经网络设置为包含两个分支的像素级目标检测神经网络,其两个分支分别对应两个输出层,并分别级联至神经网络的不同中间层。

[0167] 在步骤S502中,将所述神经网络的第一支路的第一输出层配置为在用所述神经网络进行目标检测时输出置信度图的置信度输出层,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度;步骤S503将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该检测目标的边界框。

[0168] 可选的,当像素点位于所述检测边界框内时,所述像素点的对应向量为4维向量,该4维向量中的元素为所述像素点至所述检测边界框的上、下、左、右边界的距离。

[0169] 可选的,所述第一和第二支路中还可以构建一个或多个中间层。在一个示例中,连

接置信度输出层的分支(即,级联第 m 个中间层的第一支路)上还可以包括卷积层,配置为对第 m 个中间层输出的特征图像进行卷积,以获得所述检测图像的第一特征图;上采样层,配置为对所述第一特征图进行线性插值,将所述第一特征图调整为所述检测图像的大小;以及剪裁层,配置为剪裁所述调整后的第一特征图,以使得所述剪裁后的第一特征图与所述剪裁图像对齐,并将对齐后的第一特征图输出至置信度输出层,以输出所述置信度图。

[0170] 在一个实施例中,所述 m 个中间层包括至少一个卷积层(例如,可以为神经网络模型中的多层),用于接收所述训练图像,对所述训练图像进行特征提取,生成表示检测图像中每个像素点属于对应检测目标的置信度的特征图A。将所述特征图A输入至所述第一支路上的卷积层(例如,卷积核为 $512 \times 3 \times 3 \times 1$ 、步长为1的卷积层),进行进一步的特征提取,得到所述第一特征图A1,通过所述第一支路上的上采样层和剪裁层对所述第一特征图A1的进一步处理,从而得到所述期望的置信度图,所述置信度图中的每个像素点表示采用所述神经网络检测出的、包含检测目标的检测图像中的对应像素点属于所述检测目标的置信度。在另一个示例中,在边界框输出层的分支(即,级联第 n 个中间层的第二支路)上还可以包括卷积层,配置为对所述第 n 个中间层输出的特征图像进行卷积,以获得所述图像的第二特征图;上采样层,配置为对所述第二特征图进行线性插值,将所述第二特征图调整为所述训练图像的大小;剪裁层,配置为剪裁所述调整后的第二特征图,以使得所述调整后的第二特征图与所述训练图像对齐;以及修正线性单元层,配置为获取对齐后的第二特征图,并将所述第二特征图中的每个像素点的对应向量中的元素调整为不小于0的向量元素后输出至边界框输出层,以输出所述边界框检测图。

[0171] 在一个实施例中,所述 n 个中间层包括所述 m 个中间层与至少一个卷积层C1连接形成的中间层(例如,可以为神经网络模型中的多层),用于接收所述训练图像,对所述训练图像进行特征提取,生成表示检测图像中每个像素点属于对应检测目标的检测边界框的特征图B。例如,上述第 n 个中间层可以将获得特征图A输入至所述卷积层C1,所述卷积层C1对所述特征图进行特征提取,以获得特征图B。将所述特征图输入至所述第二支路上的卷积层(例如,卷积核为 $512 \times 3 \times 3 \times 1$ 、步长为1的卷积层),进行进一步的特征提取,得到所述第二特征图B1,通过所述第二支路上的上采样层和剪裁层对所述第一特征图B1的进一步调整,并通过修正线性单元层将调整后的第二特征图B1中的每个像素点的对应向量中的元素调整为不小于0的向量元素,从而得到所述期望的边界框检测图,所述边界框检测图包含该检测图像中的每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系。在一个实施例中,经过所述修正线性单元层的调整,位于所述对应的检测边界框外的像素点的向量被调整为0向量,位于所述对应的检测边界框内的像素点的对应向量为不小于0的向量。

[0172] 进一步地,本公开实施例的神经网络构建方法还可以包括:训练所述神经网络。此处,可以采用本领域中任何适当的训练方法对该神经网络进行训练,如SGD优化方法、RMSProp优化方法、ADAM优化方法等,本公开对此不做限制。

[0173] 作为一个示例,可以利用本公开前一实施例中的第一损失函数对神经网络进行训练。具体,训练步骤可以包括:将包含有训练目标的训练图像输入至所述神经网络,以得到所述训练目标的预测边界框;根据所述预测边界框和真实边界框的交集面积与并集面积之比,获得第一损失函数,其中,所述预测边界框为采用该神经网络在包含所述训练目标的训

练图像中预测到的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框;以及至少利用所述第一损失函数调整所述神经网络的参数,以对所述神经网络进行训练。

[0174] 在一个实施例中,可以根据所述训练边界框图和所述边界框预测图,计算所述第一损失函数。所述神经网络的输入层接收所述训练图像和训练边界框图,所述神经网络第二支路上的边界框输出层输出边界框预测图,因此,可以通过所述神经网络的第二支路,计算所述第一损失函数。

[0175] 在训练过程中,对于所输入的训练图像中的每一个像素点,都可以利用神经网络预测出一个针对训练目标的对应的预测边界框。为了在获取尽可能多的有效的像素点的预测结果的基础上尽量减小运算量,可选地,可以针对位于所述训练图像中的真实边界框内的每一个像素点来预测相应的预测边界框。

[0176] 可选的,当边界框为标示目标的方形边界框时,训练图像中标注的真实边界框和神经网络预测的预测边界框均可以用4维向量来表示。例如,如上文参照图2所描述的,对于位于训练图像中真实边界框内的任一个像素点,可以基于训练边界框图中的对应像素点的向量和边界框预测图中的对应像素点的向量,分别确定第一4维向量和第二4维向量。第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,真实边界框位于图2左下角,第一4维向量可以记为 $\tilde{x} = (\tilde{x}_t, \tilde{x}_b, \tilde{x}_l, \tilde{x}_r)$;第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,与该像素点对应的预测边界框位于图2右上角,第二4维向量可以记为 $x = (x_t, x_b, x_l, x_r)$ 。由此可以求得交集面积和并集面积的比值I/U并获得第一损失函数。在本公开的一个示例中,第一损失函数可以表示为 $L = -\ln(I/U)$ 。

[0177] 以上对预测边界框和真实边界框的表示方法仅仅是当边界框为方形边界框时的一种示例表示方式,而并非是对本公开的限制,本领域技术人员也可以采用其他的表示方式。比如,当边界框为方形边界框时,还可以利用边界框的中心点坐标位置、长度和宽度的参数进行表示,同样可以计算出预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。此外,当边界框不为方形边界框时同理。例如,边界框为圆形边界框时,可以利用圆心坐标和半径来表示,并利用相似的方法计算预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。

[0178] 在利用第一损失函数对神经网络进行训练时,可以调整神经网络的参数以尽量最大化预测边界框和真实边界框的交集面积与并集面积之比,即令I/U尽可能大,并将此时的神经网络的参数的值作为训练好的神经网络参数的值。换句话说,在一个实施例中,当与I/U的比值相关的第一损失函数表示为 $L = -\ln(I/U)$ 时,需要调整神经网络的参数使得第一损失函数尽可能收敛,以得到训练好的神经网络。在另一个实施例中,还可以对神经网络设定训练轮数的阈值,以确定当所述神经网络的训练轮数达到设定的阈值时参数的值,作为训练好的神经网络的参数的值。

[0179] 此外,在利用第一损失函数训练神经网络的基础上,还可以利用反映预测置信度与真实置信度之间的差异的第二损失函数对该神经网络进行训练,以得到更好的训练效果。所述预测置信度为采用该神经网络预测出的所述训练图像中的每个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的每个像素点属于所述

训练目标的置信度。其中,针对图像中的每个像素点,真实置信度可以表示为1(或正值)和0(或负值),分别用以表示该像素点落在训练目标中和没有落在训练目标中。在对神经网络的训练过程中,可以利用所述第一损失函数和第二损失函数共同来调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,从而得到训练好的神经网络。在一个实施例中,最终的神经网络的损失函数取决于两个分支损失函数的加权和。

[0180] 在一个实施例中,可以根据所述真实置信度图以及所述神经网络的置信度输出层输出的预测置信度图,计算所述第二损失函数。所述神经网络的输入层可以接受所述训练图像和真实置信度图,并且所述第一支路上的置信度输出层可以输出预测置信度图,因此,可以根据所述神经网络的第一支路,计算所述第二损失函数。

[0181] 在利用所述构建的神经网络进行目标检测过程中,利用第一输出层输出的置信度图中的至少一个像素点,选取第二层输出层输出的边界框检测图中的对应像素点,根据所选择的边界框检测图中的对应像素点的向量,可以得到所述像素点与对应的检测边界框的上、下、左、右各个边界的关系,从而获得更精确的边界框。因此,通过将神经网络创建为具有两个分支、其两个输出层分别连接两个不同的中间层的这种设置,可以结合第一输出层和第二输出层的预测结果,从而更精确地进行目标检测。

[0182] 例如,图6例示了根据本公开实施例的神经网络的构建方法构建得到的神经网络的示意性结构图以及训练此神经网络的示意性流程。如图6所示,神经网络包括两个支路,其中置信度输出层位于第一支路中,用于输出置信度图;边界框输出层位于第二支路中,用于输出边界框检测图。

[0183] 考虑到在将边界框回归为一个整体时,边界框输出层比置信度输出层需要更大的感受野(receptive field),可选的,在图6所示的根据本公开实施例的神经网络中,将置信度输出层级联至VGG stage-4末端,而边界框输出层级联至VGG stage-5末端,即,置信度输出层所级联的中间层比边界框输出层级联的中间层靠前。

[0184] 以人脸检测为例,在对图6所示神经网络的训练过程中,首先输入包含人脸的图像作为训练图像,随后经过中间层stage1-4之后,将表示预测置信度的特征图经由所述第一支路输出至第一分支的置信度输出层,以输出预测置信度图,而在进一步经过中间层stage-5之后将表示各像素点至与该像素点对应的预测边界框的位置关系的特征图经由所述第二支路输出至第二分支的边界框输出层,以输出所述边界框预测图,所述边界框预测图包含训练图像中的每个像素点的对应向量。其中,预测的置信度图以渐变的灰度图表示不同像素点对应置信度的大小,范围为0-1,对应于像素点落在训练目标中的概率;边界框预测图也同样以四个渐变的灰度图分别表示训练图像中的像素点与对应的预测边界框的位置关系。当像素点位于对应的预测边界框内时,所述像素点的向量中的元素表示所述像素点至所述与该像素点对应的预测边界框的上、下、左、右四个边界的距离的大小。图6的神经网络中的损失函数可以通过神经网络对于训练图像的预测值和输入的训练图像的真实值之间的比较来确定。其中,如置信度输出层和边界框输出层之后分别对应的两个双箭头所示,神经网络通过边界框输出层所输出的表示距边界框距离的四个向量图与双箭头后的输入层输入的真实边界框向量图(即,训练边界框图)进行比较,获得第一损失函数;通过置信度输出层所输出的预测置信度灰度图和双箭头后的输入层输入的真实置信度图进行比

较,获得第二损失函数。其中,真实边界框的向量图(即,训练边界框图)为像素点距真实边界框四边距离的四个渐变灰度图,真实置信度图分别以白色和黑色表示像素点落在训练目标中(1)和没有落在训练目标中(0)的情况。而最终的神经网络的损失函数取决于两个分支损失函数的加权和。由于在训练过程中置信度输出层所级联的中间层比边界框输出层级联的中间层靠前,因此,可以直观地说,边界框能够从置信度图中预测。

[0185] 由此可见,本公开实施例中的神经网络构建方法能够在边界框回归时,使得边界框输出层分支比置信度输出层分支获得更大的感受野,以使边界框输出层分支采用的自下而上的策略而从置信度图中提取边界框,有效提高了神经网络的训练和检测效率。

[0186] 下面,将参考图7来描述根据本公开实施例的用于目标检测的神经网络的构建装置。图7示出了可用于实现根据本公开实施例的神经网络的构建装置的计算设备的框图。

[0187] 如图7所示,计算设备700包括一个或多个处理器702、存储器704、图像采集装置706和输出装置708,这些组件通过总线系统710和/或其它形式的连接机构(未示出)互连。应当注意,图7所示的计算设备700的组件和结构只是示例性的,而非限制性的,根据需要,计算设备700也可以具有其他组件和结构。

[0188] 处理器702可以是中央处理单元(CPU)或者具有数据处理能力和/或指令执行能力的其它形式的处理单元,并且可以控制计算设备700中的其它组件以执行期望的功能。

[0189] 存储器704可以包括一个或多个计算机程序产品,所述计算机程序产品可以包括各种形式的计算机可读存储介质,例如易失性存储器和/或非易失性存储器。所述易失性存储器例如可以包括随机存取存储器(RAM)和/或高速缓冲存储器(cache)等。所述非易失性存储器例如可以包括只读存储器(ROM)、硬盘、闪存等。在所述计算机可读存储介质上可以存储一个或多个计算机程序指令,处理器702可以运行所述程序指令,以实现以下步骤:构建至少包含一个输入层、 n 个中间层和两个输出层的神经网络, n 是大于等于2的正整数,其中将第 m 个中间层级联至所述神经网络的第一支路,将第 n 个中间层级联至所述神经网络的第二支路,并且, m 小于 n ;将所述神经网络的第一支路的第一输出层配置为在利用所述神经网络进行目标检测时输出置信度图的置信度输出层,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度;以及将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该检测目标的边界框。

[0190] 图像采集装置706用于处理器702对所述神经网络进行训练时采集带有训练目标的训练图像,并且将所采集的训练图像存储在存储器704中以供其它组件使用。当然,也可以利用其他图像采集设备拍摄采集所述训练图像,并且将拍摄的训练图像发送给计算设备700。在这种情况下,可以省略图像采集装置706。

[0191] 输出装置708可以向外部(例如用户)输出各种信息,例如图像信息、训练结果,并且可以包括显示器、扬声器、投影仪、网卡等中的一个或多个。

[0192] 下面,将参考图8来描述根据本公开实施例的用于目标检测的神经网络的构建装置。图8示出了根据本公开实施例的用于目标检测的神经网络的构建装置的示意性框图。

[0193] 如图8所示,构建装置800包括构建部件810,配置为构建至少包含一个输入层、 n 个

中间层和两个输出层的神经网络, n 是大于等于2的正整数, 其中将第 m 个中间层级联至所述神经网络的第一支路, 将第 n 个中间层级联至所述神经网络的第二支路, 并且, m 小于 n ; 第一配置部件820, 配置为将所述神经网络的第一支路的第一输出层配置为在利用所述神经网络进行目标检测时输出置信度图的置信度输出层, 所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度; 以及第二配置部件830, 配置为将所述神经网络的第二支路的第二输出层配置为在用所述神经网络进行目标检测时输出边界框检测图的边界框输出层, 所述边界框检测图包含该检测图像中每个像素点的对应向量, 该对应向量表示与其对应的像素点与对应的检测边界框的位置关系, 所述检测边界框为采用该神经网络在所述图像中检测到的该检测目标的边界框。

[0194] 构建部件810可以将神经网络设置为包含两个分支的像素级目标检测神经网络, 其两个分支分别对应两个输出层, 并分别级联至神经网络的不同中间层。

[0195] 其中, 如图6所示, 在图6的神经网络中的连接置信度输出层的分支中, 构建部件810还可以将神经网络配置有多个层结构, 如卷积层, 配置为对所述第 m 个中间层输出的特征图像(例如, 检测图像经过所述stage1-3以及stage4的输出的特征图像)进行卷积, 以获得所述检测图像的第一特征图; 上采样层, 配置为对所述第一特征图进行线性插值, 将所述第一特征图调整为所述检测图像的大小; 以及剪裁层, 配置为剪裁所述调整后的第一特征图, 以使得所述剪裁后的第一特征图与所述检测图像对齐, 并将对齐后的第一特征图输出至置信度输出层, 以输出所述置信度图。

[0196] 在图6所示的神经网络连接边界框输出层的分支中, 构建部件810还可以将神经网络配置为包括卷积层, 对所述第 n 个中间层输出的特征图像(例如, 检测图像经过所述stage1-3、stage4以及stage5的输出的特征图像)进行卷积, 以获得所述检测图像的第二特征图; 上采样层, 配置为对所述第二特征图进行线性插值, 将所述第二特征图调整为所述检测图像的大小; 剪裁层, 配置为剪裁所述调整后的第二特征图, 以使得所述调整后的第二特征图与所述检测图像对齐; 以及修正线性单元层, 配置为获取对齐后的第二特征图, 并将所述第二特征图中的每个像素点的对应向量中的元素调整为不小于0的向量元素后输出至边界框输出层, 以输出所述边界框检测图。其中修正线性单元层能够将边界框输出层所输出的4维向量的符号均修正为不小于0的向量, 以适应于之后第一损失函数的相关计算。经过所述修正线性单元层的调整, 位于所述对应的检测边界框外的像素点的向量被调整为0向量, 位于所述对应的检测边界框内的像素点的对应向量为不小于0的向量。与现有技术相比, 本公开实施例所构建的神经网络采用了尽量少的神经网络层结构, 以进一步提高了神经网络的检测效率。

[0197] 其中, 当像素点位于所述检测目标的边界框内时, 所述对应向量为4维向量, 该4维向量中的元素为与其对应的像素点至所述对应的检测边界框的上、下、左、右边界的距离。

[0198] 进一步地, 本公开实施例构建装置800还可以包括训练部件(未示出), 配置为训练所述神经网络。其训练步骤可以包括: 将包含有训练目标的训练图像输入至所述神经网络, 以得到所述训练目标的预测边界框; 根据所述预测边界框和真实边界框的交集面积与并集面积之比, 获得第一损失函数, 其中, 所述预测边界框为采用该神经网络在包含所述训练目标的训练图像中检测到的该训练目标的边界框, 所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框; 以及至少利用所述第一损失函数调整所述神经网络的参数,

以对所述神经网络进行训练。

[0199] 在一个实施例中,可以由所述构建的神经网络的输入层接收所述训练图像和训练边界框图,所述神经网络第二支路上的边界框输出层输出边界框预测图。所述训练边界框图包含所述训练图像中的每个像素点的4维向量,该4维向量表明对应的像素点与真实边界框的位置关系。当像素点位于所述真实边界框内时,该像素点对应的4维向量中的元素分别表明该像素点距离真实边界框的上、下、左、右边界的距离。对于位于所述真实边界框之外的像素点,可以用0向量表示。所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

[0200] 在训练过程中,对于所输入的训练图像中的每一个像素点,都可以利用神经网络预测出一个针对训练目标的预测边界框。为了在获取尽可能多的有效的像素点的预测结果的基础上尽量减小运算量,优选地,可以针对位于真实边界框内的每一个像素点来预测相应的预测边界框。

[0201] 在本公开的一个实施例中,当边界框为标示目标的方形边界框时,真实边界框和预测边界框均可以用4维向量来表示。以图2为例,对于位于训练图像中真实边界框内的任一个像素点(该像素点可以包括位于真实边界框边界上的像素点),训练部件可以基于训练边界框图中的对应像素点的向量和边界框预测图中的对应像素点的向量,分别确定所述像素点的第一4维向量和第二4维向量。第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,真实边界框位于图2左下角,第一4维向量可以记为 $\tilde{x} = (\tilde{x}_t, \tilde{x}_b, \tilde{x}_l, \tilde{x}_r)$;第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,与该像素点对应的预测边界框位于图2右上角,第二4维向量可以记为 $x = (x_t, x_b, x_l, x_r)$ 。如前所述,同样可以求得交集面积和并集面积的比值 I/U 并获得第一损失函数。在本公开的一个优选实施例中,第一损失函数可以表示为 $L = -\ln(I/U)$ 。

[0202] 以上对预测边界框和真实边界框的表示方法仅仅是当边界框为方形边界框时的一种示例表示方式,而并非是对本公开的限制,本领域技术人员也可以采用其他的表示方式。比如,当边界框为方形边界框时,还可以利用边界框的中心点坐标位置、长度和宽度的参数进行表示,同样可以计算出预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。此外,当边界框不为方形边界框时同理。例如,边界框为圆形边界框时,可以利用圆心坐标和半径来表示,并利用相似的方法计算预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。

[0203] 在训练部件利用第一损失函数对神经网络进行训练时,可以调整神经网络的参数以尽量最大化预测边界框和真实边界框的交集面积与并集面积之比,即令 I/U 尽可能大,并将此时的神经网络的参数的值作为训练好的神经网络参数的值。换句话说,在一个实施例中,当与 I/U 的比值相关的第一损失函数表示为 $L = -\ln(I/U)$ 时,需要调整神经网络的参数使得第一损失函数尽可能收敛,以得到训练好的神经网络。在另一个实施例中,还可以对神经网络设定训练轮数的阈值,以确定当所述神经网络的训练轮数达到设定的阈值时参数的值,作为训练好的神经网络的参数的值。

[0204] 此外,训练部件在利用第一损失函数训练神经网络的基础上,还可以利用反映训

练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数对该神经网络进行训练,以得到更好的训练效果。所述预测置信度为采用该神经网络预测出的所述训练图像中的每个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的每个像素点属于所述训练目标的置信度。其中,针对图像中的每个像素点,真实置信度可以表示为1(或正值)和0(或负值),分别用以表示该像素点落在训练目标中和没有落在训练目标中。在对神经网络的训练过程中,可以利用所述第一损失函数和第二损失函数共同来调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,从而得到训练好的神经网络。最终的神经网络的损失函数取决于两个分支损失函数的加权和。

[0205] 在一个实施例中,所述构建的神经网络的输入层配置为接收所述训练图像和真实置信度图,并且所述第一支路上的置信度输出层可以输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,因此,可以利用所述构建的神经网络的第一支路计算所述第二损失函数。

[0206] 在利用所述构建装置构建的神经网络进行目标检测过程中,利用第一输出层输出的置信度图中的至少一个像素点,选取第二层输出层输出的边界框检测图中的对应像素点,根据所选择的边界框检测图中的对应像素点的向量,可以得到所述像素点与对应的检测边界框的上、下、左、右各个边界的关系,从而获得更精确的边界框。因此,通过将神经网络创建为具有两个分支、其两个输出层分别连接两个不同的中间层的这种设置,可以结合第一输出层和第二输出层的预测结果,从而更精确地进行目标检测。

[0207] 由此可见,本公开实施例中的神经网络构建装置能够在边界框回归时,使得边界框输出层分支比置信度输出层分支获得更大的感受野,以使边界框输出层分支采用的自下而上的策略而从置信度图中提取边界框,有效提高了神经网络的训练和检测效率。

[0208] 本公开实施例还提供一种用于目标检测的神经网络。图6示出了根据本公开实施例的用于目标检测的神经网络一个优选结构。

[0209] 所述神经网络包括一个输入层、 n 个中间层和两个输出层, n 是大于等于2的正整数,其中第 m 个中间层级联至所述神经网络的第一支路,第 n 个中间层级联至所述神经网络的第二支路,并且, m 小于 n ;所述神经网络的第一支路的第一输出层配置为在利用所述神经网络进行目标检测时输出置信度图,所述置信度图表示采用所述神经网络检测出的、包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述神经网络的第二支路的第二输出层配置为在利用所述神经网络进行目标检测时输出边界框检测图的边界框输出层,所述边界框检测图包含该检测图像中每个像素点的对应向量,该对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述检测边界框为采用该神经网络在所述检测图像中检测到的该目标的边界框。

[0210] 其中,在图6的神经网络连接置信度输出层的分支中,还可以包括多个层结构,如卷积层,配置为对第 m 个中间层输出的特征图像(例如,检测图像经过所述stage1-3以及stage4的输出的特征图像)进行卷积,以获得所述检测图像的第一特征图;上采样层,配置为对所述第一特征图进行线性插值,将所述第一特征图调整为所述检测图像的大小;以及剪裁层,配置为剪裁所述调整后的第一特征图,以使得所述剪裁后的第一特征图与所述检测图像对齐,并将对齐后的第一特征图输出至置信度输出层,以输出所述置信度图。

[0211] 在图6所示的神经网络连接边界框输出层的分支中,还可以包括卷积层,对所述第n个中间层输出的特征图像(例如,检测图像经过所述stage1-3、stage4以及stage5的输出特征图像)进行卷积,以获得所述图像的第二特征图;上采样层,配置为对所述第二特征图进行线性插值,将所述第二特征图调整为所述检测图像的大小;剪裁层,配置为剪裁所述调整后的第二特征图,以使得所述剪裁后的第二特征图与所述检测图像对齐;以及修正线性单元层,配置为获取对齐后的第二特征图,并将所述第二特征图中的每个像素点的对应向量中的元素调整为不小于0的向量元素后输出至边界框输出层,以输出所述边界框检测图。其中修正线性单元层能够将边界框输出层所输出的4维向量的符号均修正为不小于0的向量,以适应于之后第一损失函数的相关计算。经过所述修正线性单元层的调整,位于所述对应的检测边界框外的像素点的向量被调整为0向量,位于所述对应的检测边界框内的像素点的对应向量为不小于0的向量。与现有技术相比,本公开实施例所构建的神经网络采用了尽量少的神经网络层结构,以进一步提高了神经网络的检测效率。

[0212] 其中,当像素点位于所述检测目标的边界框内时,所述对应向量为4维向量,该4维向量中的元素为与其对应的像素点至所述对应的检测边界框的上、下、左、右边界的距离。

[0213] 进一步地,本公开实施例的神经网络是利用第一损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的该训练目标的边界框。

[0214] 在一个实施例中,可以由所述神经网络的输入层接收所述训练图像和训练边界框图,所述神经网络第二支路上的边界框输出层输出边界框预测图。所述训练边界框图包含所述训练图像中的每个像素点的4维向量,该4维向量表明对应的像素点与真实边界框的位置关系。当像素点位于所述真实边界框内时,该像素点对应的4维向量中的元素分别表明该像素点距离真实边界框的上、下、左、右边界的距离。对于位于所述真实边界框之外的像素点,可以用0向量表示。所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。

[0215] 在训练过程中,对于所输入的训练图像中的每一个像素点,都可以利用神经网络预测出一个针对训练目标的预测边界框。为了在获取尽可能多的有效的像素点的预测结果的基础上尽量减小运算量,优选地,可以针对位于真实边界框内的每一个像素点来预测相应的预测边界框。

[0216] 在本公开的一个实施例中,当边界框为标示目标的方形边界框时,真实边界框和预测边界框均可以用4维向量来表示。以图2为例,对于位于训练图像中真实边界框内的任一个像素点(该像素点可以包括位于真实边界框边界上的像素点),训练部件可以基于训练边界框图中的对应像素点的向量和边界框预测图中的对应像素点的向量,分别确定所述像素点的第一4维向量和第二4维向量。第一4维向量中的元素分别为该像素点至所述真实边界框的上、下、左、右边界的距离,例如对于如图2所示的情形,真实边界框位于图2左下角,第一4维向量可以记为 $\tilde{x} = (\tilde{x}_t, \tilde{x}_b, \tilde{x}_l, \tilde{x}_r)$;第二4维向量中的元素分别表示该像素点至与该像素点对应的预测边界框(位于图2中的右上角的边界框)的上、下、左、右边界的距离,例如对于如图2所示的情形,与该像素点对应的预测边界框位于图2右上角,第二4维向量可以记

为 $x = (x_t, x_b, x_l, x_r)$ 。如前所述,同样可以求得交集面积和并集面积的比值 I/U 并获得第一损失函数。在本公开的一个优选实施例中,第一损失函数可以表示为 $L = -\ln(I/U)$ 。

[0217] 以上对预测边界框和真实边界框的表示方法仅仅是当边界框为方形边界框时的一种示例表示方式,而并非是对本公开的限制,本领域技术人员也可以采用其他的表示方式。比如,当边界框为方形边界框时,还可以利用边界框的中心点坐标位置、长度和宽度的参数进行表示,同样可以计算出预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。此外,当边界框不为方形边界框时也同理。例如,边界框为圆形边界框时,可以利用圆心坐标和半径来表示,并利用相似的方法计算预测边界框和真实边界框的交集面积和并集面积的比,作为所述第一损失函数。

[0218] 在训练部件利用第一损失函数对神经网络进行训练时,可以调整神经网络的参数以尽量最大化预测边界框和真实边界框的交集面积与并集面积之比,即令 I/U 尽可能大,并将此时的神经网络的参数的值作为训练好的神经网络参数的值。换句话说,在一个实施例中,当与 I/U 的比值相关的第一损失函数表示为 $L = -\ln(I/U)$ 时,需要调整神经网络的参数使得第一损失函数尽可能收敛,以得到训练好的神经网络。在另一个实施例中,还可以对神经网络设定训练轮数的阈值,以确定当所述神经网络的训练轮数达到设定的阈值时参数的值,作为训练好的神经网络的参数的值。

[0219] 此外,在利用第一损失函数训练神经网络的基础上,还可以利用反映训练图像中各像素点的预测置信度与真实置信度之间的差异的第二损失函数对该神经网络进行训练,以得到更好的训练效果。可选地,第二损失函数可以为sigmoid cross-entropy损失函数。所述预测置信度为采用该神经网络预测出的所述训练图像中的每个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的每个像素点属于所述训练目标的置信度。其中,针对图像中的每个像素点,真实置信度可以表示为1(或正值)和0(或负值),分别用以表示该像素点落在训练目标中和没有落在训练目标中。在对神经网络的训练过程中,可以利用所述第一损失函数和第二损失函数共同来调整所述神经网络的参数,以最大化所述预测边界框和所述真实边界框的交集面积与并集面积之比,并最小化所述第二损失函数,从而得到训练好的神经网络。最终的神经网络的损失函数取决于两个分支损失函数的加权和。

[0220] 在一个实施例中,所述神经网络的输入层配置为接收所述训练图像和真实置信度图,并且所述第一支路上的置信度输出层可以输出预测置信度图,该预测置信度图表示采用该神经网络预测出的、包含所述训练目标的训练图像中的每个像素点属于该训练目标的置信度,因此,可以利用所述构建的神经网络的第一支路计算所述第二损失函数。在利用所述构建装置构建的神经网络进行目标检测过程中,利用第一输出层输出的置信度图中的至少一个像素点,选取第二层输出层输出的边界框检测图中的对应像素点,根据所选择的边界框检测图中的对应像素点的向量,可以得到所述像素点与对应的检测边界框的上、下、左、右各个边界的关系,从而获得更精确的边界框。下文中将结合附图对利用所述神经网络进行目标检测的方法和装置进行具体描述。因此,通过将神经网络创建为具有两个分支、其两个输出层分别连接两个不同的中间层的这种设置,可以结合第一输出层和第二输出层的预测结果,从而更精确地进行目标检测。

[0221] 由此可见,本公开实施例中的神经网络能够在边界框回归时,使得边界框输出层

分支置信度输出层分支获得更大的感受野,以使边界框输出层分支采用的自下而上的策略而从置信度图中提取边界框,有效提高了神经网络的训练和检测效率。

[0222] 下面将参考图9对根据本公开实施例的基于神经网络的目标检测方法进行详细的描述。图9示意性地示出了根据本公开实施例的基于神经网络的目标检测方法的流程图。

[0223] 如图9所示,训练方法900包括步骤S901:获取包含检测目标的检测图像。

[0224] 在步骤S902中,利用所述神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含该检测图像中每个像素点的对应向量,所述对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框。

[0225] 其中,当像素点位于所述对应的检测目标的检测边界框中时,所述向量为4维向量,该4维向量中的元素为与其对应的像素点至所述对应的检测边界框的上、下、左、右边界的距离。

[0226] 在步骤S903中,在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状。

[0227] 在该步骤中,可以设定置信度的预定阈值,并据此确定置信度大于预定阈值的多个像素点,随后基于所述多个像素点拟合得到与所述检测目标的形状匹配的几何形状。如检测目标为人脸时,可以根据预定阈值确定的像素点进行椭圆拟合,得到与人脸匹配的椭圆;而当检测目标为车辆或其他物品时,也可以拟合为一个矩形或其他形状。

[0228] 在步骤S904中,在拟合得到的几何形状内选取预定数量的像素点。

[0229] 在该步骤中,可以对拟合得到的几何形状内的每个像素点都进行计算。但为了减少计算量,提高检测效率,可以提取几何形状内的一个或几个像素点并进行后续步骤的计算。如在上述实施例中提到的人脸检测所拟合的椭圆内,可以选择椭圆的中心点,其长/短轴的中点等一个或多个点来计算。

[0230] 在步骤S905中,对于每个选取的像素点,确定所述对应边界框检测图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框。

[0231] 本步骤中,对于置信度图中每个选择的像素点,都可以对应到所输入的检测图像中的某个像素点,并可以据此得到此像素点的在对应边界框检测图中的对应向量,根据对应向量与所述对应的检测边界框的位置关系,例如,当所述像素点位于对应的检测边界框内时,所述像素点的对应向量中的元素为与所述像素点至对应的检测边界框的上、下、左、右边界的距离,得到像素点的对应的检测边界框。

[0232] 在步骤S906中,将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个检测边界框。

[0233] 针对同一个检测目标,在步骤S905中所得到的检测边界框可能不止一个。更明确的说,在步骤S905中,由对应边界框检测图中的各个对应像素点确定的各个检测边界框可能并不是完全重合的。因此,在该步骤中可以对这些检测边界框进行合并,以使得对于同一个检测目标合并得到一个检测边界框。可选的,可以利用非最大抑制方法将各个检测边界框进行合并,以消除多余的边界框。而最终合并得到的一个检测边界框,即代表检测到图像中的一个检测目标,所述检测边界框是一个精确的检测边界框结果。

[0234] 可选的,本公开实施例的目标检测方法所使用的神经网络可以至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括用于输出所述置信度图的置信度输出层和用于输出所述边界框检测图的边界框输出层。

[0235] 其中所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0236] 在一个实施例中,所述神经网络可以由图6所示的神经网络实现。所述神经网络的第 m 个中间层经由卷积层、上采样层以及剪裁层在神经网络的第一支路上级联至所述置信度输出层,通过所述卷积层、上采样层以及剪裁层对所述第 m 个中间层输出的特征图特征提取和调整,经由所述置信度输出层输出所述置信度图。所述神经网络的第 n 个中间层经由卷积层、上采样层、剪裁层以及修正线性单元层在神经网络的第二支路上级联至所述边界框输出层,通过所述卷积层、上采样层、剪裁层以及修正线性单元层对所述第 n 个中间层输出的特征图进行特征提取和调整,经由所述边界框输出层输出所述边界框检测图。经过所述修正线性单元层的调整,位于所述对应的检测边界框外的像素点的向量被调整为0向量,位于所述对应的检测边界框内的像素点的对应向量为不小于0的向量。

[0237] 另一方面,可选的,本公开实施例所使用的神经网络可以是利用第一损失函数训练得到的,所述第一损失函数反映所述预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框。

[0238] 在一个实施例中,可以由所述神经网络的输入层接收所述训练图像和训练边界框图,所述神经网络的边界框输出层输出边界框预测图。所述训练边界框图包含所述训练图像中的每个像素点的4维向量,该4维向量表明对应的像素点与真实边界框的位置关系。当像素点位于所述真实边界框内时,该像素点对应的4维向量中的元素分别表明该像素点距离真实边界框的上、下、左、右边界的距离。对于位于所述真实边界框之外的像素点,可以用0向量表示。所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。可以根据所述训练边界框图和所述边界框预测图,通过所述神经网络的第二支路计算所述第一损失函数。

[0239] 可选地,本公开实施例中所使用的神经网络还可以是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述第二损失函数反映训练图像中各像素点的预测置信度与真实置信度之间的差异,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述预测置信度为采用该神经网络预测出的所述训练图像中的每个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的每个像素点属于所述训练目标的置信度。在对神经网络的训练过程中,可以利用所述第一损失函数和第二损失函数共同来调整所述神经网络的参数,最终的神经网络的损失函数取决于两个分支损失函数的加权和。

[0240] 由此可见,本公开实施例的基于神经网络的目标检测方法能够得到精确和有效的目标定位结果,检测效率高,并且不必改变输入图像的尺寸,适应性更强,适用于不同尺寸

的目标,尤其对于小尺寸目标进行检测也能够取得较好的效果。此外由于在本公开实施例中能够针对每个像素点均生成对应检测边界框的对应向量图,因此能够针对位于不同检测目标中的像素点获得相应的不同的检测边界框,因此本公开实施例能够同时处理检测图像中的多个检测目标

[0241] 下面,将参考图10来描述根据本公开实施例的基于神经网络的目标检测装置。图10示出了根据本公开实施例的基于神经网络的目标检测装置。该目标检测装置可以是配备有图像采集装置的计算机或服务器。

[0242] 如图10所示,训练装置1000包括一个或多个处理器1002、存储器1004、图像采集装置1006和输出装置1008,这些组件通过总线系统1010和/或其它形式的连接机构(未示出)互连。应当注意,图10所示的训练装置1000的组件和结构只是示例性的,而非限制性的,根据需要,训练装置1000也可以具有其他组件和结构。

[0243] 处理器1002可以是中央处理单元(CPU)或者具有数据处理能力和/或指令执行能力的其它形式的处理单元,并且可以控制训练装置1000中的其它组件以执行期望的功能。

[0244] 存储器1004可以包括一个或多个计算机程序产品,所述计算机程序产品可以包括各种形式的计算机可读存储介质,例如易失性存储器和/或非易失性存储器。所述易失性存储器例如可以包括随机存取存储器(RAM)和/或高速缓冲存储器(cache)等。所述非易失性存储器例如可以包括只读存储器(ROM)、硬盘、闪存等。在所述计算机可读存储介质上可以存储一个或多个计算机程序指令,处理器1002可以运行所述程序指令,以实现以下步骤:获取包含检测目标的检测图像;利用所述神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含该检测图像中每个像素点的对应向量,所述对应向量表示与其对应的像素点与检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框;在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状;在拟合得到的几何形状内选取预定数量的像素点;对于每个选取的像素点,确定所述边界框检测图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框;将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个精确的检测边界框。

[0245] 图像采集装置1006用于采集带有训练目标的训练图像,并且将所采集的训练图像存储在存储器1004中以供其它组件使用。当然,也可以利用其他图像采集设备采集所述训练图像,并且将采集的训练图像发送给目标检测装置1000。在这种情况下,可以省略图像采集装置1006。

[0246] 输出装置1008可以向外部(例如用户)输出各种信息,例如图像信息、目标检测结果,并且可以包括显示器、扬声器等中的一个或多个。

[0247] 下面,将参考图11来描述根据本公开实施例的基于神经网络的目标检测装置。图11示出了根据本公开实施例的基于神经网络的目标检测装置的示意性框图。

[0248] 如图11所示,训练装置1100包括图像获取部件1110,配置为获取包含检测目标的检测图像;检测部件1120,配置为利用所述神经网络在所述检测图像中进行目标检测,以输出置信度图和边界框检测图,所述边界框检测图包含该检测图像中每个像素点的对应向

量,所述对应向量表示与其对应的像素点与对应的检测边界框的位置关系,所述置信度图中的每个像素点表示包含检测目标的检测图像中的每个像素点属于所述检测目标的置信度,所述检测边界框为采用该神经网络在所述检测图像中检测到的所述检测目标的边界框;拟合部件1130,配置为在所述置信度图中基于各个像素点的置信度、通过形状拟合得到至少一个几何形状;选取部件1140,配置为在拟合得到的几何形状内选取预定数量的像素点;边界框获取部件1150,配置为对于每个选取的像素点,确定所述对应向量图中的对应像素点,并基于该对应像素点的对应向量得到与该对应像素点对应的检测边界框;合并部件1160,配置为将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个精确的检测边界框。

[0249] 其中,当像素点位于所述对应的检测目标的边界框中时,所述像素点对应的向量为4维向量,该4维向量中的元素为所述像素点至所述对应的检测边界框的上、下、左、右边界的距离。

[0250] 在拟合部件1130通过形状拟合得到几何形状时,可以首先置信度的预定阈值,并据此确定置信度大于预定阈值的多个像素点;随后基于所述多个像素点拟合得到与所述检测目标的形状匹配的几何形状。如检测目标为人脸时,可以根据预定阈值确定的像素点进行椭圆拟合,得到与人脸匹配的椭圆;而当检测目标为车辆或其他物品时,也可以拟合为一个矩形或其他形状。

[0251] 在选取部件1140自拟合得到的几何形状内选取预定数量的像素点时,可以对拟合得到的几何形状内的每个像素点都进行计算。但为了减少计算量,提高检测效率,可以提取几何形状内的一个或几个像素点并进行后续步骤的计算。如在上述实施例中提到的人脸检测所拟合的椭圆内,可以选择椭圆的中心点,其长/短轴的中点等一个或多个点来计算。

[0252] 随后,边界框获取部件1150对于置信度图中每个选择的像素点,都可以对应到所输入的检测图像中的某个像素点,并可以据此得到此像素点的在边界框检测图中的对应向量,根据对应向量得到像素点的检测边界框。

[0253] 其中,针对同一个检测目标,所得到的检测边界框可能不止一个。当对应同一个检测目标的检测边界框为多个时,合并部件1160可以将各个所述检测边界框进行合并,以使得对于同一个检测目标合并得到一个精确的检测边界框。此时,合并部件1160可以利用非最大抑制方法将各个检测边界框进行合并,以消除多余的边界框。当每个检测目标只对应一个检测边界框时,合并部件1160也可以省略。

[0254] 本公开实施例的目标检测装置所使用的神经网络可以至少包括一个输入层、 n 个中间层和两个输出层,其中, n 是大于等于2的正整数,所述两个输出层包括用于输出所述置信度图的置信度输出层和用于输出所述边界框检测图的边界框输出层。

[0255] 其中所述神经网络的第 m 个中间层级联至所述置信度输出层,第 n 个中间层级联至所述边界框输出层,并且, m 小于 n 。

[0256] 在一个实施例中,所述神经网络可以由图6所示的神经网络实现。所述神经网络的第 m 个中间层经由卷积层、上采样层以及剪裁层在神经网络的第一支路上级联至所述置信度输出层,通过所述卷积层、上采样层以及剪裁层对所述第 m 个中间层输出的特征图进行特征提取和调整,经由所述置信度输出层输出所述置信度图。所述神经网络的第 n 个中间层经由卷积层、上采样层、剪裁层以及修正线性单元层在神经网络的第二支路上级联至所述边

界框输出层,通过所述卷积层、上采样层、剪裁层以及修正线性单元层对所述第n个中间层输出的特征图进行特征提取和调整,经由所述边界框输出层输出所述边界框检测图。经过所述修正线性单元层的调整,位于所述对应的检测边界框外的像素点的向量被调整为0向量,位于所述对应的检测边界框内的像素点的对应向量为不小于0的向量。

[0257] 本公开实施例所使用的神经网络可以是利用第一损失函数训练得到的,所述第一损失函数反映所述预测边界框和真实边界框的交集面积与并集面积之比,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框。

[0258] 在一个实施例中,可以由所述神经网络的输入层接收所述训练图像和训练边界框图,所述神经网络的边界框输出层输出边界框预测图。所述训练边界框图包含所述训练图像中的每个像素点的4维向量,该4维向量表明对应的像素点与真实边界框的位置关系。当像素点位于所述真实边界框内时,该像素点对应的4维向量中的元素分别表明该像素点距离真实边界框的上、下、左、右边界的距离对于位于所述真实边界框之外的像素点,可以用0向量表示。所述边界框预测图包含位于所述训练图像中的每个像素点的4维向量,其中,当像素点位于所述训练图像中的真实边界框内时,所述像素点的4维向量中的元素为该像素点至与该像素点对应的预测边界框的上、下、左、右边界的距离。可以根据所述训练边界框图和所述边界框预测图,通过所述神经网络的第二支路计算所述第一损失函数。

[0259] 可选地,本公开实施例中所使用的神经网络还可以是利用第一损失函数和第二损失函数训练得到的,所述第一损失函数反映预测边界框和真实边界框的交集面积与并集面积之比,所述第二损失函数反映训练图像中各像素点的预测置信度与真实置信度之间的差异,所述预测边界框为采用该神经网络在包含训练目标的训练图像中预测出的该训练目标的边界框,所述真实边界框为在所述训练图像中预先标注的训练目标的边界框,所述预测置信度为采用该神经网络预测出的所述训练图像中的每个像素点属于所述训练目标的置信度,所述真实置信度表示在所述训练图像中预先标注的每个像素点属于所述训练目标的置信度。在对神经网络的训练过程中,可以利用所述第一损失函数和第二损失函数共同来调整所述神经网络的参数,最终的神经网络的损失函数取决于两个分支损失函数的加权和。

[0260] 由此可见,本公开实施例的基于神经网络的目标检测装置能够得到精确和有效的目标定位结果,检测效率高,并且不必改变输入图像的尺寸,适应性更强,适用于不同尺寸的目标,尤其对于小尺寸目标进行检测也能够取得较好的效果,并且能够同时处理检测图像中的多个检测目标。

[0261] 图12示出了利用根据本公开实施例的目标检测方法和装置进行人脸检测的示例,可见利用本公开实施例的方法和装置所得到的人脸目标检测边界框定位精确,并对不同尺寸的人脸目标适应性强,针对有部分遮挡的人脸同样能够得到较好的检测效果。

[0262] 图13(a)、(b)分别示出了利用上文中提到的第一损失函数和两个输出层的神经网络结构的目标检测方法,与利用 l_2 损失函数并采用相同神经网络结构的目标检测方法的未命中率(miss rate)和查全率(recall rate)的比较。根据图13(a)可以看出,随着神经网络训练轮数的增加,利用本公开实施例所对应的损失函数训练的神经网络收敛的更快、更稳定,而采用 l_2 损失函数训练的神经网络收敛很慢并且不稳定。而根据图13(b)可以看出,使

用根据本公开实施例的损失函数训练根据本公开实施例的神经网络进行目标检测的工作性能曲线 (ROC) 效果更好。

[0263] 图14示出了在检测图像的大小分别为图像长边尺寸为60像素到960像素不等时, 利用 l_2 损失函数训练的神经网络和采用第一损失函数训练的神经网络进行人脸目标检测的情形。可以看到, 随着图像尺寸的变化, 利用基于 l_2 损失函数训练的神经网络的检测方法对检测边界框的判断能力随之变化, 并对于小尺寸的人脸 (60像素) 基本没有判断能力。而比较而言, 利用基于本公开实施例的损失函数训练的神经网络检测方法对不同大小的人脸检测目标都有很好的判断。

[0264] 图15示出本公开实施例的人脸目标检测方法与其他检测方法的横向比较示意图。可见本公开实施例所采用的目标检测方法性能比现有技术中的多种方法均有明显提高。

[0265] 以上结合具体实施例描述了本公开的基本原理, 但是, 需要指出的是, 在本公开中提及的优点、优势、效果等仅是示例而非限制, 不能认为这些优点、优势、效果等是本公开的各个实施例必须具备的。另外, 上述公开的具体细节仅是为了示例的作用和便于理解的作用, 而非限制, 上述细节并不限制本公开为必须采用上述具体的细节来实现。

[0266] 本公开中涉及的器件、装置、设备、系统的方框图仅作为例示性的例子并且不意图要求或暗示必须按照方框图所示的方式进行连接、布置、配置。如本领域技术人员将认识到的, 可以按任意方式连接、布置、配置这些器件、装置、设备、系统。诸如“包括”、“包含”、“具有”等等的词语是开放性词汇, 指“包括但不限于”, 且可与其互换使用。这里所使用的词汇“或”和“和”指词汇“和/或”, 且可与其互换使用, 除非上下文明确指示不是如此。这里所使用的词汇“诸如”指词组“诸如但不限于”, 且可与其互换使用。

[0267] 另外, 如在此使用的, 在以“至少一个”开始的项的列举中使用的“或”指示分离的列举, 以便例如“A、B或C的至少一个”的列举意味着A或B或C, 或AB或AC或BC, 或ABC (即A和B和C)。此外, 措辞“示例的”不意味着描述的例子是优选的或者比其他例子更好。

[0268] 还需要指出的是, 在本公开的系统和方法中, 各部件或各步骤是可以分解和/或重新组合的。这些分解和/或重新组合应视为本公开的等效方案。

[0269] 可以不脱离由所附权利要求定义的教导的技术而进行对在此所述的技术的各种改变、替换和更改。此外, 本公开的权利要求的范围不限于以上所述的处理、机器、制造、事件的组成、手段、方法和动作的具体方面。可以利用与在此所述的相应方面进行基本相同的功能或者实现基本相同的结果的当前存在的或者稍后要开发的处理、机器、制造、事件的组成、手段、方法或动作。因而, 所附权利要求包括在其范围内的这样的处理、机器、制造、事件的组成、手段、方法或动作。

[0270] 提供所公开的方面的以上描述以使本领域的任何技术人员能够做出或者使用本公开。对这些方面的各种修改对于本领域技术人员而言是非常显而易见的, 并且在此定义的一般原理可以应用于其他方面而不脱离本公开的范围。因此, 本公开不意图被限制到在此示出的方面, 而是按照与在此公开的原理和新颖的特征一致的最宽范围。

[0271] 为了例示和描述的目的已经给出了以上描述。此外, 此描述不意图将本公开的实施例限制到在此公开的形式。尽管以上已经讨论了多个示例方面和实施例, 但是本领域技术人员将认识到其某些变型、修改、改变、添加和子组合。

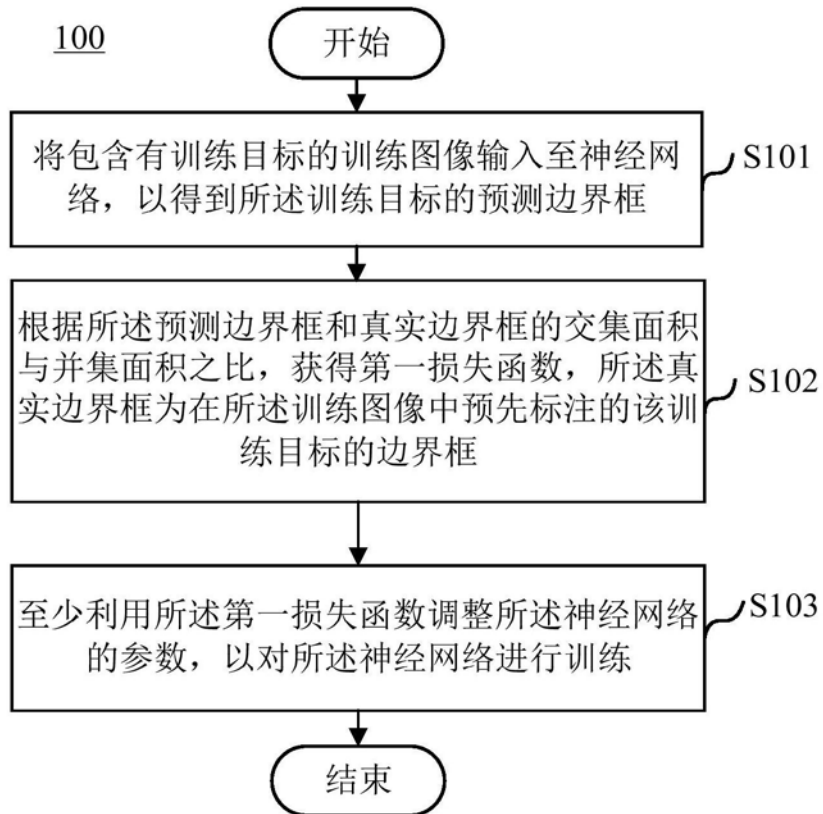


图1

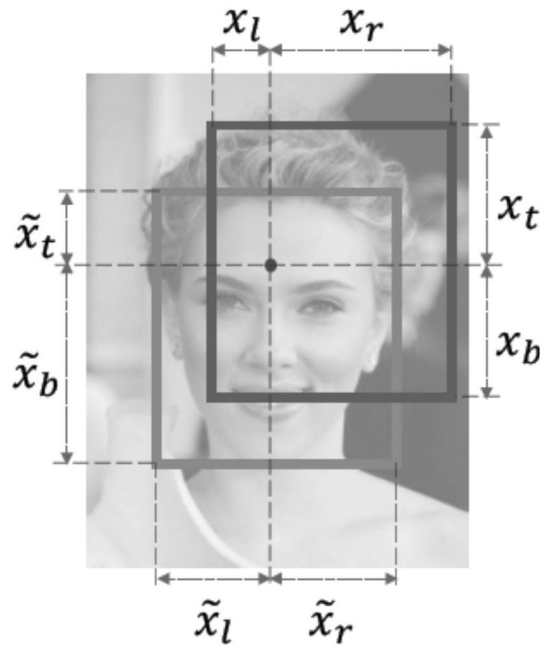


图2

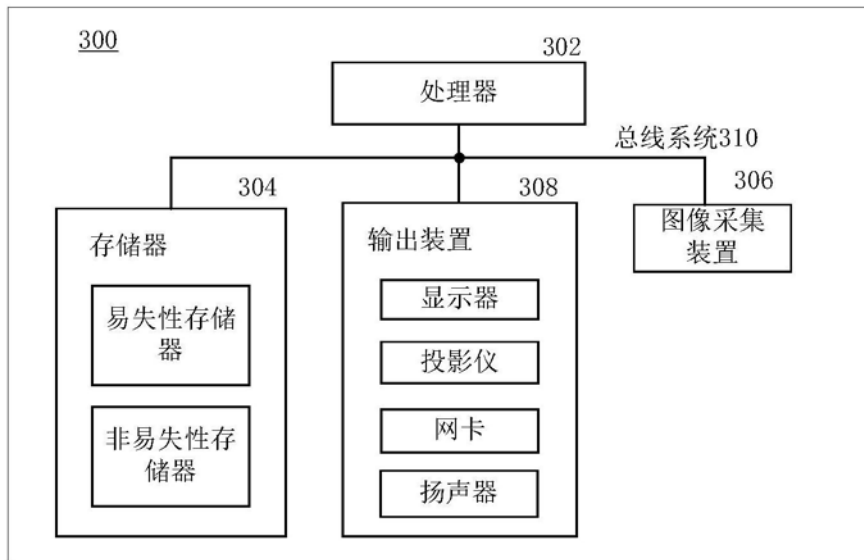


图3

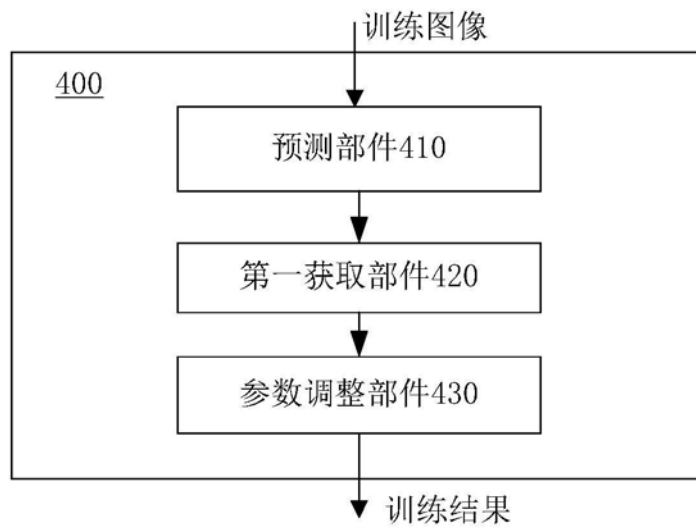


图4

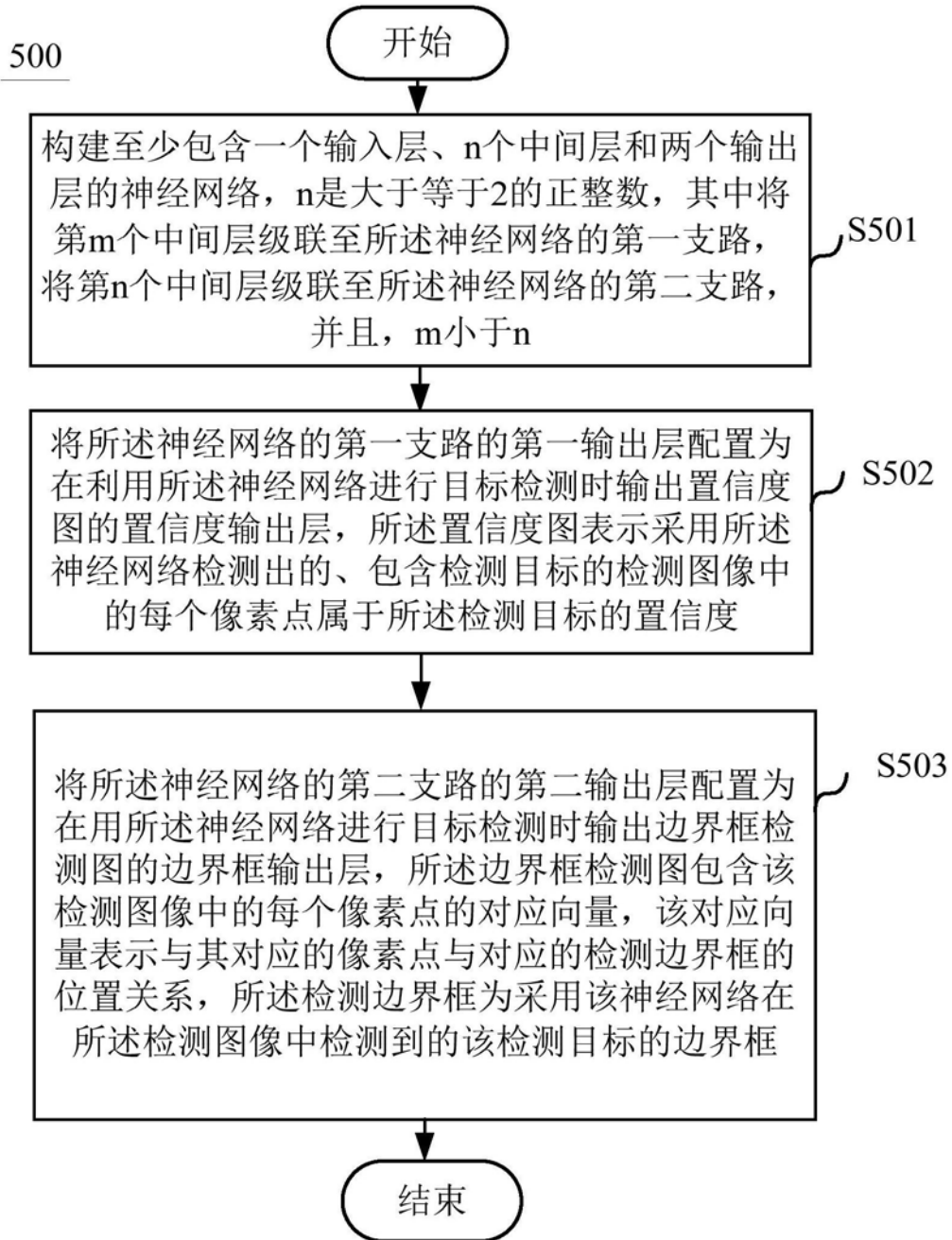


图5

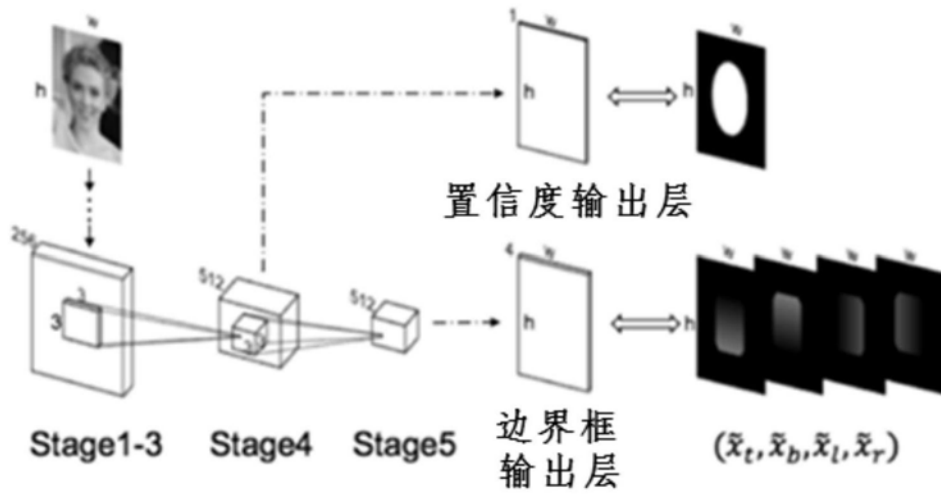


图6

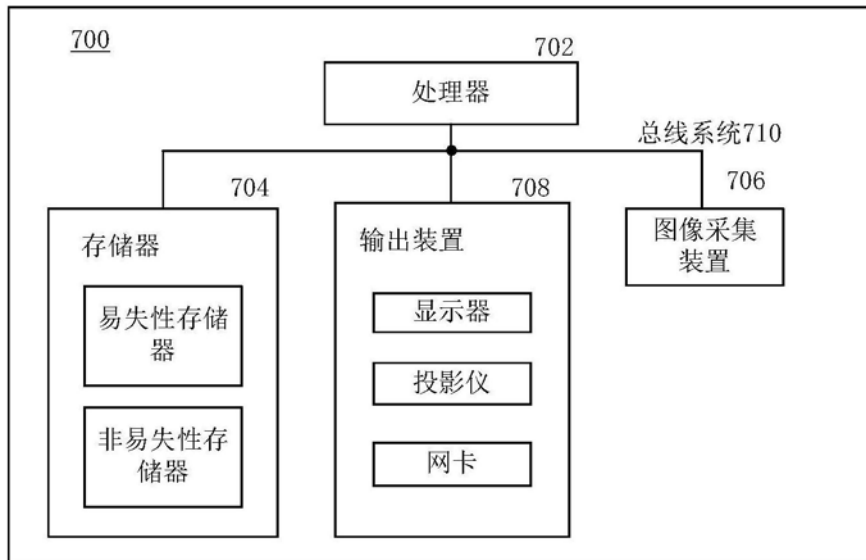


图7

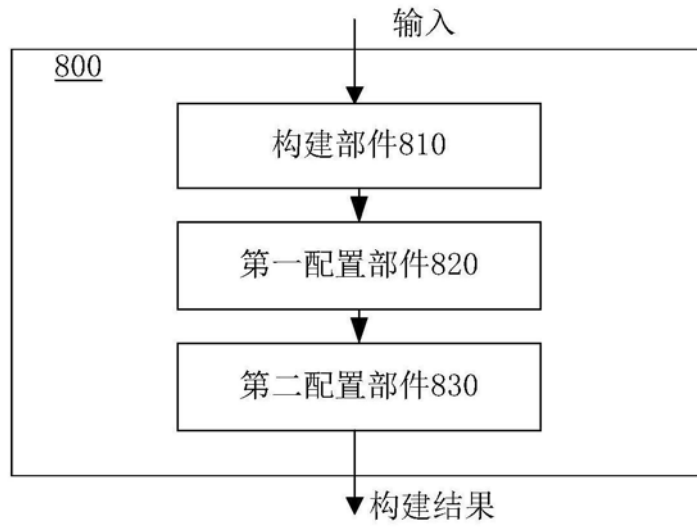


图8

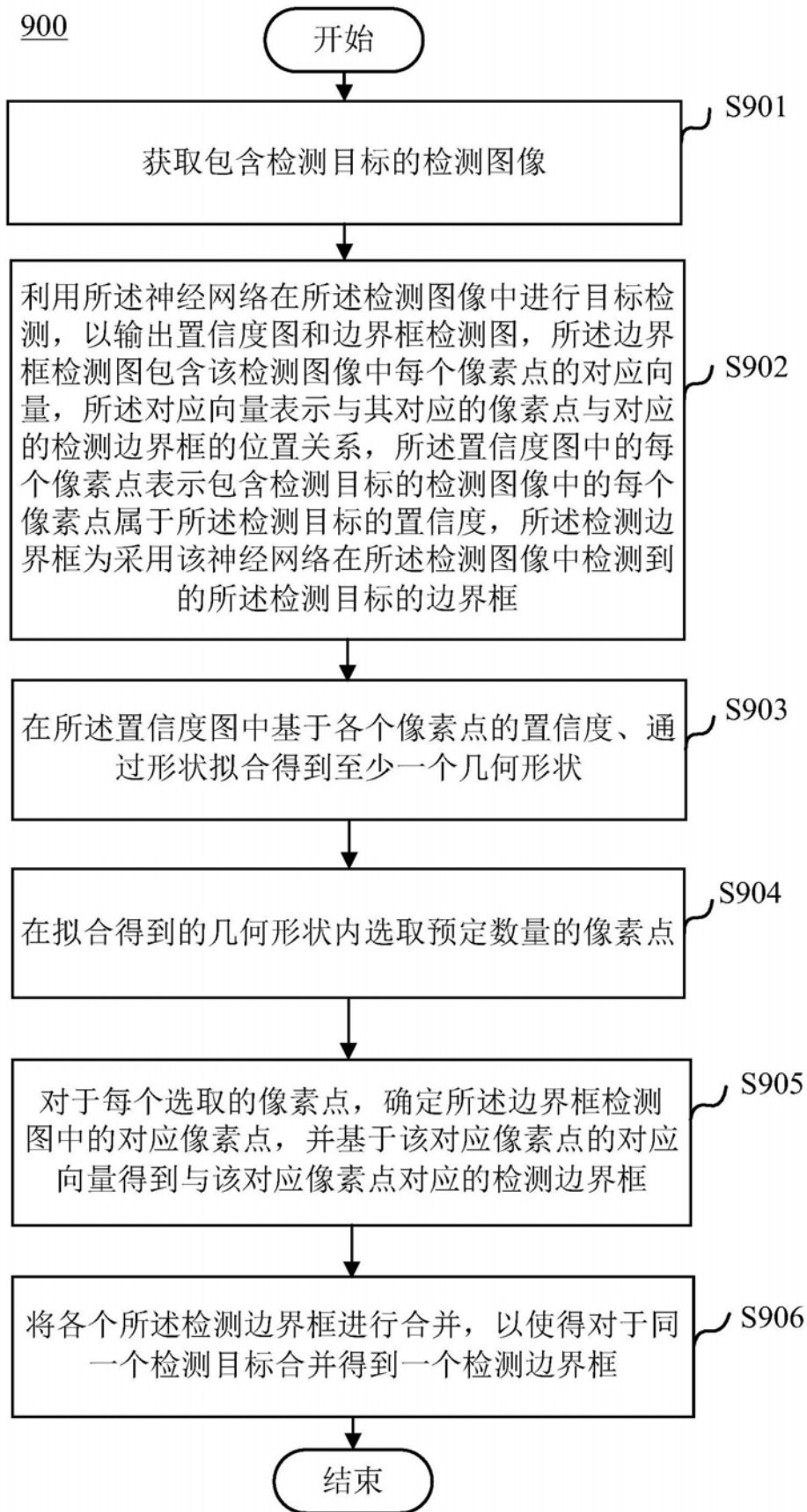


图9

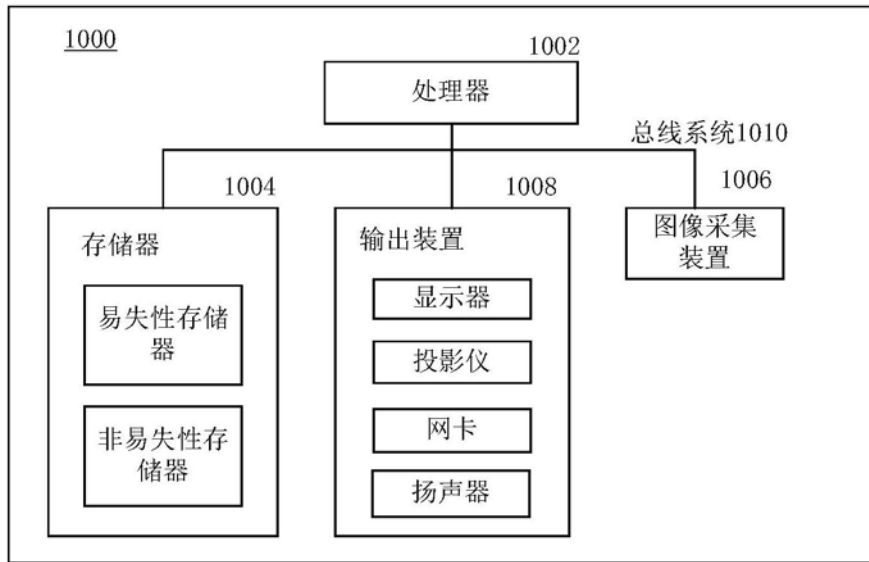


图10

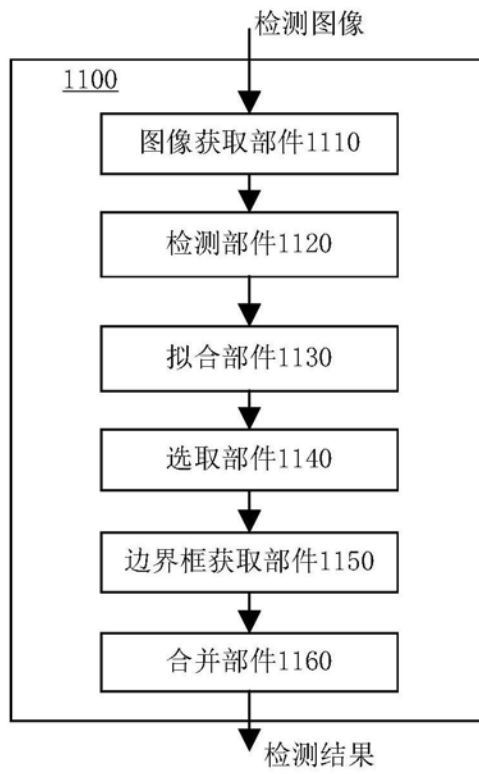


图11



图12

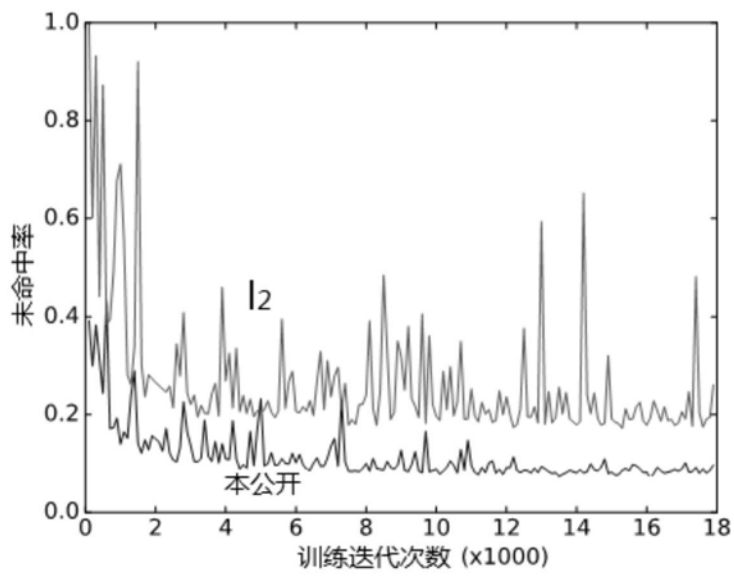


图13(a)

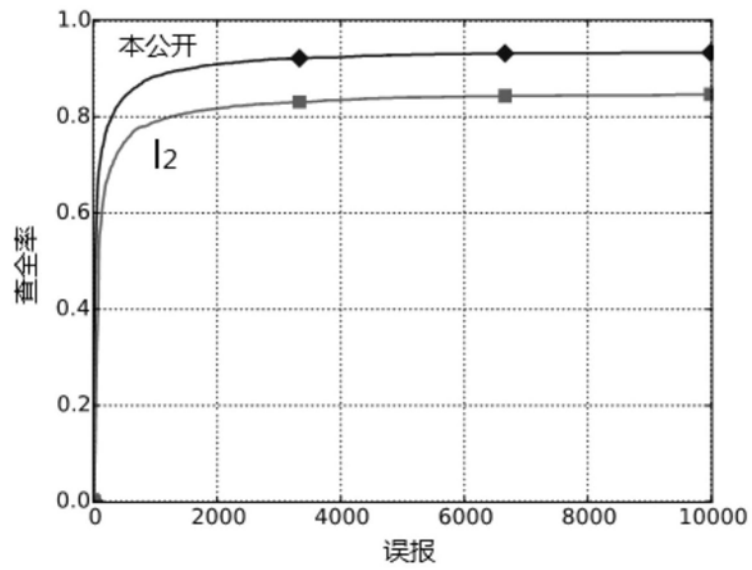


图13 (b)

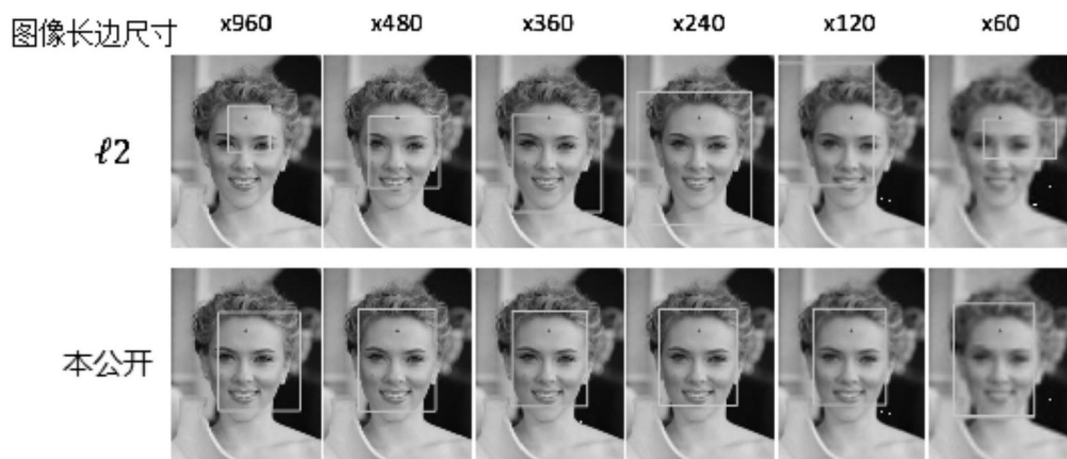


图14

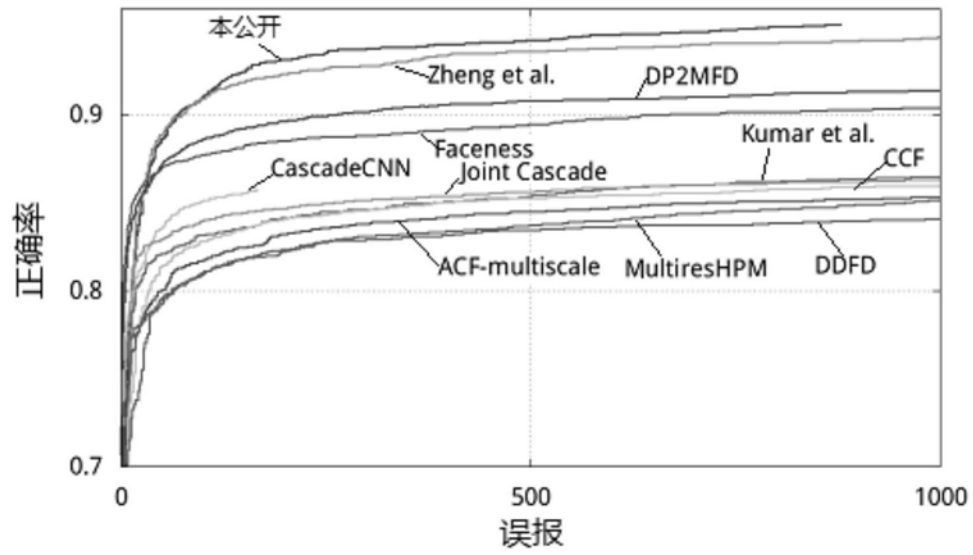


图15