



(12) 发明专利

(10) 授权公告号 CN 108399362 B

(45) 授权公告日 2022.01.07

(21) 申请号 201810069322.X

G06N 3/04 (2006.01)

(22) 申请日 2018.01.24

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 105956608 A, 2016.09.21

申请公布号 CN 108399362 A

CN 105787439 A, 2016.07.20

CN 106934346 A, 2017.07.07

(43) 申请公布日 2018.08.14

审查员 孙建萍

(73) 专利权人 中山大学

地址 510000 广东省广州市海珠区新港西路135号大院

(72) 发明人 林惊 尹森堂 张冬雨 王青

(74) 专利代理机构 广州容大知识产权代理事务所(普通合伙) 44326

代理人 刘新年

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

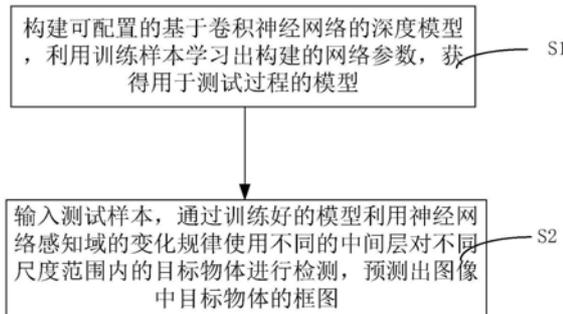
权利要求书1页 说明书8页 附图5页

(54) 发明名称

一种快速行人检测方法及装置

(57) 摘要

本发明公开了一种快速行人检测方法及装置,所述方法包括如下步骤:步骤S1,构建可配置的基于卷积神经网络的深度模型,利用训练样本学习出构建的网络参数,获得用于测试过程的模型;步骤S2,输入测试样本,通过训练好的模型利用神经网络感知域的变化规律使用不同的中间层对不同尺度范围内的目标物体进行检测,预测出图像中目标物体的框图,本发明通过利用神经网络感知域的变化规律,使用不同的中间层对特定尺度范围内的目标物体进行检测,更好的适应了感知域与物体大小的关系,有效提高了检测结果。



1. 一种快速行人检测方法,包括如下步骤:

步骤S1,构建可配置的基于卷积神经网络的深度模型,利用训练样本学习出构建的网络参数,获得用于测试过程的模型;

步骤S2,输入测试样本,通过训练好的模型利用神经网络感知域的变化规律使用不同的中间层对不同尺度范围内的目标物体进行检测,预测出图像中目标物体的框图;

步骤S1进一步包括:

构建可配置的基于卷积神经网络的深度模型;

输入训练样本;

初始化卷积神经网络及其参数,包括网络层中每层连接的权重和偏置;

采用前向传播算法和后向传播算法,利用训练样本学习出构建的网络参数,即用于测试过程的模型;

所述深度模型包括多尺度的目标候选网络与目标检测网络,所述目标候选网络基于卷积神经网络不同层提出特征的差异性,在中间层分别生成对不同尺度目标物体的候选框图;所述目标检测网络在所述目标候选网络输出的候选框图的基础上进行精细化的分类和检测;

所述卷积神经网络由卷积层、降采样层、上采样层堆叠而成,所述卷积层是指对输入的图像或者特征图在二维空间上进行卷积运算,提取层次化特征;所述降采样层使用没有重叠的max-pooling操作,该操作用于提取形状和偏移不变的特征,同时减少特征图大小,提高计算效率;所述上采样层,是指对输入的特征图在二维空间上进行去卷积的操作,用以增大特征图的像素。

2. 如权利要求1所述的一种快速行人检测方法,其特征在于:所述深度模型采用Squeeze VGG-16卷积神经网络作为骨干网络,所述Squeeze VGG-16卷积神经网络采用conv1-1层和紧随其后的12层Fire模块层为特征提取的网络结构。

3. 如权利要求2所述的一种快速行人检测方法,其特征在于:所述目标候选网络在所述Squeeze VGG-16卷积神经网络基础上,根据卷积层特征,在Fire9、Fire12、conv6以及增加的pooling层,产生网络分支,以进行不同尺度检测到物体的候选框的回归。

4. 如权利要求2所述的一种快速行人检测方法,其特征在于:所述目标检测网络在所述目标候选区域的基础上,将目标候选区域预设倍数大小的图片区域作为目标的背景语义信息,将Fire9层的特征图进行一次上采样,作为增强对小物体感知的信息,并将背景语义信息与上采样信息经过感兴趣区域的池化获得固定大小的特征,之后增加一层全连接层,进行类别和最终候选框的回归。

一种快速行人检测方法及装置

技术领域

[0001] 本发明涉及行人检测技术领域,特别是涉及一种基于深度学习的面向嵌入式系统的快速行人检测方法及装置。

背景技术

[0002] 作为计算机视觉中目标检测的一部分,行人检测在现实世界的应用有着重要意义,随着图像采集技术的成熟与存储技术成本的下降,越来越多的摄像机被部署在公共场所,另一方面,随着自动驾驶、智能交通的推行,车载摄像头也产生了海量的视频资源。传统的人工筛选和处理,不仅效率低下,耗费大量人力物力,而且可能引入一些人为因素,导致一些偏差。近年来,深度学习在计算机视觉领域取得前所未有的突破,不仅效率远胜人力,准确度在很多领域也超过人类。因此,有效利用深度学习的方法进行行人检测的课题备受关注。

[0003] 人是视频监控或自动驾驶中最主要的目标之一,而行人检测的首要任务就是识别人体的存在,并提供相应的标注信息。由于在现实世界中捕捉到的图像质量参差不齐,对于小物体、遮挡的物体的检测一直是行人检测的难点,另一方面,车载摄像头也经常捕捉到一些模糊的图像,这样的图像中也存在大量类似行人却不是行人的物体。而具体到嵌入式系统,由于识别能力强的大型神经网络模型通常难以有效率的运行在计算资源有限的嵌入式设备上,而对于嵌入式设备的应用需求又是实时的,因此兼顾检测准确率和效率是面向嵌入式系统的快速行人检测的重中之重。

发明内容

[0004] 为克服上述现有技术存在的不足,本发明之一目的在于提供一种快速行人检测方法及装置,通过利用神经网络感知域的变化规律,使用不同的中间层对特定尺度范围内的目标物体进行检测,更好的适应了感知域与物体大小的关系,有效提高了检测结果。

[0005] 本发明之另一目的在于提供一种快速行人检测方法及装置,通过调整并训练VGG-16的网络得到适应嵌入式系统要求的squeeze VGG-16网络,有效降低了网络模型的参数量并加快了计算效率。

[0006] 本发明之再一目的在于提供一种快速行人检测方法及装置,通过去卷积的方法对特定网络层的特征图进行放大,增强了对小物体的检测,相比于传统图片放大的方法,几乎不增加显存和计算量。

[0007] 本发明之又一目的在于提供一种快速行人检测方法及装置,通过使用目标对象1.5倍大小的区域作为背景语义特征增加到网络中,对于模糊物体和远距离小物体的检测,有着极佳的性能。

[0008] 为达上述及其它目的,本发明提出一种快速行人检测方法,包括如下步骤:

[0009] 步骤S1,构建可配置的基于卷积神经网络的深度模型,利用训练样本学习出构建的网络参数,获得用于测试过程的模型;

[0010] 步骤S2,输入测试样本,通过训练好的模型利用神经网络感知域的变化规律使用不同的中间层对不同尺度范围内的目标物体进行检测,预测出图像中目标物体的框图。

[0011] 优选地,步骤S1进一步包括:

[0012] 构建可配置的基于卷积神经网络的深度模型;

[0013] 输入训练样本;

[0014] 初始化卷积神经网络及其参数,包括网络层中每层连接的权重和偏置;

[0015] 采用前向传播算法和后向传播算法,利用训练样本学习出构建的网络参数,即用于测试过程的模型。

[0016] 优选地,所述该深度模型包括多尺度的目标候选网络与目标检测网络,所述目标候选网络基于卷积神经网络不同层提出特征的差异性,在中间层分别生成对不同尺度目标物体的候选框图;所述目标检测网络在所述目标候选网络输出的候选框图的基础上进行精细化的分类和检测。

[0017] 优选地,所述卷积神经网络由卷积层、降采样层、上采样层堆叠而成。所述卷积层是指对输入的图像或者特征图在二维空间上进行卷积运算,提取层次化特征;所述降采样层使用没有重叠的max-pooling操作,该操作用于提取形状和偏移不变的特征,同时减少特征图大小,提高计算效率;所述上采样层,是指对输入的特征图在二维空间上进行去卷积的操作,用以增大特征图的像素。

[0018] 优选地,所述深度模型采用Squeeze VGG-16卷积神经网络作为骨干网络,所述Squeeze VGG-16卷积神经网络采用conv1-1层和紧随其后的12层Fire模块层为特征提取的网络结构。

[0019] 优选地,所述目标候选网络在所述Squeeze VGG-16卷积神经网络基础上,根据卷积层特征,在Fire9、Fire12、conv6以及增加的pooling层,产生网络分支,以进行不同尺度检测到物体的候选框的回归。

[0020] 优选地,所述目标检测网络在所述目标候选区域的基础上,将目标候选区域预设倍数大小的图片区域作为目标的背景语义信息,将Fire9层的特征图进行一次上采样,作为增强对小物体感知的信息,并将背景语义信息与上采样信息经过感兴趣区域的池化获得固定大小的特征,之后增加一层全连接层,进行类别和最终候选框的回归。

[0021] 优选地,所述训练样本包括RGB图像数据和图像中行人区域的标注信息,实际训练用的图像数据是根据行人所在区域裁剪得到的小的patch。

[0022] 优选地,所述后向传播算法,需先求出正向传播预测的目标框图与图像实际目标框图的损失函数 $\mathcal{L}(W)$,然后求得其对参数W的梯度,采用梯度下降的算法更新W以最小化损失函数 \mathcal{L} 。假定中间层有M个分支可以输出目标候选区域, l^m 表示分支m的损失函数, α_m 表示 l^m 函数的权重, $S = \{S^1, S^2, \dots, S^M\}$ 指相应尺度的目标对象,则损失函数 $\mathcal{L}(W)$ 可定义为:

$$[0023] \quad \mathcal{L}(W) = \sum_{m=1}^M \sum_{i \in S^m} \alpha_m l^m(X_i, Y_i | W)。$$

[0024] 为达到上述目的,本发明还提供一种快速行人检测系统,包括:

[0025] 训练单元,用于构建可配置的基于卷积神经网络的深度模型,利用训练样本学习出构建的网络参数,获得用于测试过程的模型;

[0026] 检测单元,用于输入测试样本,通过训练好的模型利用神经网络感知域的变化规

律使用不同的中间层对不同尺度范围内的目标物体进行检测,预测出图像中目标物体的框图。

[0027] 与现有技术相比,本发明一种快速行人检测方法及其装置借鉴压缩网络的方法,调整并训练VGG-16的网络得到适应嵌入式系统要求的squeeze VGG-16网络,有效降低了网络模型的参数量并加快了计算效率;另一方面,针对传统检测方法中感知域与物体大小不一致的问题,本发明利用神经网络感知域的变化规律(即神经网络层越深,感知域越大,适合检测大一些的目标物体),使用不同的中间层对特定尺度范围内的目标物体进行检测,更好的适应了感知域与物体大小的关系,有效提高了检测结果;另外,为了增强对小物体的检测,本发明使用去卷积的方法对特定网络层的特征图进行放大,相比于传统图片放大的方法,几乎不增加显存和计算量;为了增强对于模糊物体的检测,在该层的特征图上,使用目标对象1.5倍大小的区域作为背景语义特征增加到网络中,对于模糊物体和远距离小物体的检测,有着极佳的性能。

附图说明

- [0028] 图1为本发明一种快速行人检测方法的步骤流程图;
- [0029] 图2为本发明具体实施例中Squeeze VGG-16神经网络结构示意图;
- [0030] 图3为本发明具体实施例中Fire模块的示意图;
- [0031] 图4为本发明具体实施例中目标候选网络的结构示意图;
- [0032] 图5为本发明具体实施例中目标检测网络的结构示意图;
- [0033] 图6为本发明具体实施例中快速行人检测的过程示意图;
- [0034] 图7为本发明一种快速行人检测装置的系统架构图;
- [0035] 图8为本发明具体实施例中训练单元的细部结构图;
- [0036] 图9为本发明具体实施例中检测单元的细部结构图。

具体实施方式

[0037] 以下通过特定的具体实例并结合附图说明本发明的实施方式,本领域技术人员可由本说明书所揭示的内容轻易地了解本发明的其它优点与功效。本发明亦可通过其它不同的具体实例加以施行或应用,本说明书中的各项细节亦可基于不同观点与应用,在不背离本发明的精神下进行各种修饰与变更。

[0038] 图1为本发明一种快速行人检测方法的步骤流程图。如图1所示,本发明一种快速行人检测方法,包括如下步骤:

[0039] 步骤S1,构建可配置的基于卷积神经网络的深度模型,利用训练样本学习出构建的网络参数,获得用于测试过程的模型。在本发明具体实施例中,该深度模型由两个子网路组成:第一个子网络,为多尺度的目标候选网络,用于提取人物特征并给出候选区域,具体地,该目标候选网络基于卷积神经网络不同层提出特征的差异性,在中间层分别生成对不同尺度行人的候选框图;第二个子网络,为目标检测网络,增强检测的效果,其与目标候选网络共享参数,在候选框图的基础上进行精细化的分类和检测。具体地,步骤S1进一步包括:

[0040] 步骤S100,构建可配置的基于卷积神经网络的深度模型。

[0041] 所述卷积神经网络由卷积层、降采样层、上采样层堆叠而成,所述卷积层是指对输入的图像或者特征图在二维空间上进行卷积运算,提取层次化特征;所述的降采样层使用没有重叠的max-pooling操作,该操作用于提取形状和偏移不变的特征,同时减少特征图大小,提高计算效率;所述的上采样层,是指对输入的特征图在二维空间上进行去卷积的操作,用以增大特征图的像素,主要用于目标检测网络,提升检测效果,在本发明具体实施例中,采用Squeeze VGG-16卷积神经网络作为骨干网络,如图2所示,该Squeeze VGG-16卷积神经网络采用conv1-1层和紧随其后的12层Fire模块作为卷积层,用以提取特征;其中的pool1-pool5是降采样层;使用在ImageNet数据集上预先训练好的模型作为初始化。即本发明首先利用ImageNet数据集预先训练Squeeze VGG-16作为网络初始化。

[0042] 图3为本发明具体实施例中Fire模块的结构示意图。如图3所示,Fire模块由两个卷积核大小为 1×1 的卷积层和一个卷积核大小为 3×3 的卷积层组成,目的在于用 1×1 的卷积核代替 3×3 的卷积核,从而使参数量减少9倍,但为了不影响网络的表征能力,不是全部替代,而是一部分是用 1×1 的卷积核,一部分使用 3×3 的卷积核,这样做的另一个好处是减少 3×3 卷积核的输入通道,同时起到降低参数量的效果,具体地,Fire模块先是使用 1×1 的卷积层对输入层进行降维操作,然后参照GoogLeNet结构,使用 1×1 和 3×3 的卷积层提取特征,最后将两部分特征连接起来,这样的方式极大降低了计算量和模型参数。

[0043] 图4为本发明具体实施例中目标候选网络的架构示意图。在本发明具体实施例中,所述目标候选网络在Squeeze VGG-16卷积神经网络基础上,根据卷积层特征,在Fire9、Fire12、conv6以及增加的pooling层共计4层,产生网络分支,分支进行不同尺度检测到物体的候选框的回归。但对于Fire-9层,它比较接近主干网络的低层,相比其他层对梯度的影响会很大,学习过程不稳定,因此多了一个buffer(缓冲)层,如图4中det-conv层所示,buffer层避免检测分支的梯度被直接back-propagated(反向传播)到主干层。

[0044] 本发明利用神经网络感知域的变化规律(即神经网络层越深,感知域越大,适合检测大一些的目标物体),使用不同的中间层对特定尺度范围内的目标物体进行检测,更好的适应了感知域与物体大小的关系,有效提高了检测结果。

[0045] 图5为本发明具体实施例中目标检测网络的架构示意图。所述目标检测网络与目标候选网络共享参数,将目标候选网络的候选框汇总,以增强监测网络对物体与背景的区别能力。在本发明具体实施例中,所述目标检测网络,在目标候选区域的基础上,将目标候选区域1.5倍大小的图片区域作为目标的背景语义信息;将Fire9层的特征图进行一次上采样,作为增强对小物体感知的信息,将背景语义信息与上采样信息经过感兴趣区域的池化(ROI pooling)获得固定大小的特征,之后增加一层全连接层,进行类别和最终候选框的回归,具体地,主干cmn层连接一个proposals的节点,用于汇总目标候选网络所得到的候选框信息;另一方面,针对fire9层的特征图,W和H是输入图片的宽度和高度,立方体1代表物体区域的在特征图的映射,而立方体2代表context区域在特征图上的映射,context区域约为物体区域的1.5倍,同时为了加强对小物体的检测,再对Fire9层进行一次上采样,之后与faster RCNN算法类似,使用感兴趣区域的池化获得固定大小的特征;将Fire9层处理后的特征与proposals汇总的特征连接(concat)到一起,后增加一层全连接层,进行类别和最终候选框的回归,在此不予赘述。

[0046] 步骤S101,输入训练样本。

[0047] 训练过程需要提供图像中参考人物的对应的框,同时为了加速训练,训练过程将含有参考人物的图像从原始图像中裁剪出来,形成一个个patch(图像块),patch相比于原始图像更小,用以训练,有效加速了训练过程。具体地,在本发明中,输入的训练样本包括RGB图像数据和图像中行人区域的标注信息,实际训练用的图像数据是根据行人所在区域裁剪得到的小的patch(图像块)。用数学语言表示,训练样本 $S = \{(X_i, Y_i)\}_{i=1}^N$,其中 X_i 表示训练图片的一个patch;在实际应用中,除了行人这一类别,还有其他类别,例如背景、骑自行车的人、坐着的人等K个类别,因此标注数据 $Y_i = (y_i, b_i)$ 由类别标签 $y_i \in \{0, 1, 2, \dots, K\}$ 和框图坐标点 $b_i = (b_i^x, b_i^y, b_i^w, b_i^h)$ 组成,其中 b_i^x, b_i^y 为框图左上角的起始坐标点, b_i^w, b_i^h 为框图宽度和高度。

[0048] 步骤S102,初始化卷积神经网络及其参数,包括网络层中每层连接的权重和偏置。具体地,本发明利用ImageNet数据集预先训练Squeeze VGG-16卷积神经网络作为网络初始化。

[0049] 步骤S103,采用前向传播算法和后向传播算法,利用训练样本学习出构建的网络参数,即用于测试过程的模型。

[0050] 在本发明中,所述前向传播算法,首先将输入图像的大小归一化为 $3 \times 480 \times 640$,截取 $3 \times 448 \times 448$ 大小的patch和相应的标注信息作为卷积神经网络的输入,经过卷积层、降采样层和矫正线性单元层(ReLU Nonlinearity Layer),在Fire9层,图像特征图大小为 $512 \times 60 \times 80$;在Fire12层,特征图大小为 $512 \times 30 \times 40$,在后面两个分支特征图大小依次是 $512 \times 15 \times 20$ 和 $512 \times 8 \times 10$ 。在不同特征图上,采用卷积的方式得到目标框图的四个坐标点和类别信息,以Fire9层为例,假定只检测行人和背景,则输出为特征大小为 $6 \times 60 \times 80$,其中6包含背景、行人两个类别和候选框图四个坐标点。在目标检测网络中,将各个分支层得到的候选框图在proposals节点进行汇总,同时与Fire9层的背景语义信息和上采样信息经过感兴趣区域的池化操作得到的特征进行叠加,做最后的框图回归和类别回归。

[0051] 在本发明中,所述后向传播算法,需要先求出正向(即前向)传播预测的目标框图与图像实际目标框图的损失函数 $\mathcal{L}(W)$,然后求得其对参数W的梯度,采用梯度下降的算法更新W以最小化损失函数 \mathcal{L} 。假定中间层有M个分支可以输出目标候选区域(M个尺度的感知域可以近似的检测出图像中所有目标物体), l^m 表示分支m的损失函数, α_m 表示 l^m 函数的权重, $S = \{S^1, S^2, \dots, S^M\}$ 指相应尺度的目标对象,则损失函数 $\mathcal{L}(W)$ 可定义为:

$$[0052] \quad \mathcal{L}(W) = \sum_{m=1}^M \sum_{i \in S^m} \alpha_m l^m(X_i, Y_i | W) \quad (1)$$

[0053] 所述损失函数,对于特定的检测层m,只有目标尺度在m所能检测的范围内,才对损失函数有贡献,故将损失函数定义为

$$[0054] \quad l^m(X, Y | W) = L_{cls}(p(X), y) + \lambda[y \geq 1]L_{loc}(b, \hat{b}) \quad (2)$$

[0055] 其中, $p(X) = (p_0(X), \dots, p_K(X))$ 表示目标类别的概率分布; λ 是平衡系数; b 为框图的4个坐标点, \hat{b} 指前向传播得到的坐标点;损失函数中,使用交叉熵损失函数定义类别回归,即

[0056] $L_{cls}(p(X), y) = -\log_y(P(X))$ (3)

[0057] 使用平滑的曼哈顿距离标准(smooth L1 criterion)进行目标框图的回归,定义如下

[0058] $L_{loc}(b, \hat{b}) = \frac{1}{4} \sum_{j \in \{x, y, w, h\}} smooth_{L_1}(b_j, \hat{b}_j)$ (4)

[0059] 步骤S2,通过训练好的模型利用神经网络感知域的变化规律使用不同的中间层对不同尺度范围内的目标物体进行检测,预测出图像中目标物体(如行人)的框图。

[0060] 具体地,步骤S2进一步包括:

[0061] 步骤S200,载入训练好的模型;

[0062] 步骤S201,输入测试样本;

[0063] 步骤S202,利用训练好的模型,通过神经网络感知域的变化规律使用不同的中间层对不同尺度范围内的行人进行检测,预测图像中行人的框图。图6为本发明具体实施例中快速行人检测的过程示意图,即利用模型中的目标候选网络在Squeeze VGG-16卷积神经网络基础上,根据卷积层特征,在fire9、fire12、conv6以及增加的pooling层共计4层产生网络分支,进行不同尺度检测到物体的目标候选区域(中间层a,中间层b,中间层c);然后利用目标检测网络,在目标候选区域的基础上,将目标候选区域1.5倍大小的图片区域作为目标的背景语义信息,将Fire9层的特征图进行一次上采样,作为增强对小物体感知的信息,将背景语义信息与上采样信息经过感兴趣区域的池化获得固定大小的特征,之后增加一层全连接层,进行类别和最终候选框的回归。优选地,于步骤S202中,还使用去卷积的方法对特定网络层的特征图进行放大。

[0064] 本发明提出的行人检测方法,分别借鉴两方面的评价指标:平均查准率mAP和每秒帧数FPS。mAP用于评价最后检测区域与真实目标人物区域的交并比的情况,在不同交并比下查准率的平均值;FPS,主要是效率指标,指每秒可以处理的图片数目。

[0065] 图7为本发明一种快速行人检测装置的系统架构图。如图7所示,本发明一种快速行人检测装置,包括:

[0066] 训练单元70,用于构建可配置的基于卷积神经网络的深度模型,利用训练样本学习出构建的网络参数,获得用于测试过程的模型。在本发明具体实施例中,训练单元70所构建的深度模型由两个子网路组成:第一个子网络,为多尺度的目标候选网络,用于提取人物特征并给出候选区域,具体地,该目标候选网络基于卷积神经网络不同层提出特征的差异性,在中间层分别生成对不同尺度行人的候选框图;第二个子网络,为目标检测网络,增强检测的效果,其与目标候选网络共享参数,在候选框图的基础上进行精细化的分类和检测。具体地,如图8所示,训练单元70进一步包括:

[0067] 模型构建单元701,用于构建可配置的基于卷积神经网络的深度模型。

[0068] 所述卷积神经网络由卷积层、降采样层、上采样层堆叠而成,所述卷积层是指对输入的图像或者特征图在二维空间上进行卷积运算,提取层次化特征;所述的降采样层使用没有重叠的max-pooling操作,该操作用于提取形状和偏移不变的特征,同时减少特征图大小,提高计算效率,所述的上采样层,是指对输入的特征图在二维空间上进行去卷积的操作,用以增大特征图的像素。在本发明具体实施例中,采用Squeeze VGG-16卷积神经网络作为骨干网络。

[0069] 在本发明具体实施例中,所述目标候选网络在Squeeze VGG-16卷积神经网络基础上,根据卷积层特征,在fire9、fire12、conv6以及增加的pooling层共计4层,产生网络分支,分支进行不同尺度检测到物体的候选框的回归。但对于fire-9层,它比较接近主干网络的低层,相比其他层对梯度的影响会很大,学习过程不稳定,因此多了一个buffer(缓冲)层,buffer层避免检测分支的梯度被直接back-propagated(反向传播)到主干层。

[0070] 所述目标检测网络与目标候选网络共享参数,将目标候选网络的候选框汇总,以增强监测网络对物体与背景的区分能力。在本发明具体实施例中,所述目标检测网络,在目标候选区域的基础上,将目标候选区域1.5倍大小的图片区域作为目标的背景语义信息;将Fire9层的特征图进行一次上采样,作为增强对小物体感知的信息,将背景语义信息与上采样信息经过感兴趣区域的池化获得固定大小的特征,之后增加一层全连接层,进行类别和最终候选框的回归,具体地,主干cnn层连接一个proposal的子网,W和H是输入图片的宽度和高度,立方体1代表物体区域的pooling,而立方体2代表context区域的pooling,context区域约为物体区域的1.5倍,同时为了加强对小物体的检测,再对Fire9层进行一次上采样,之后与faster RCNN算法类似,使用感兴趣区域的池化获得固定大小的特征,之后增加一层全连接层,进行类别和最终候选框的回归。

[0071] 训练样本输入单元702,用于输入训练样本。

[0072] 具体地,训练样本 $S = \{(X_i, Y_i)\}_{i=1}^N$,其中 X_i 表示训练图片的一个patch,标注数据 $Y_i = (y_i, b_i)$ 由类别标签 y_i 和框图坐标点 $b_i = (b_i^x, b_i^y, b_i^w, b_i^h)$ 组成。

[0073] 初始化单元703,用于初始化卷积神经网络及其参数,包括网络层中每层连接的权重和偏置。具体地,本发明利用ImageNet数据集预先训练Squeeze VGG-16卷积神经网络作为网络初始化。

[0074] 样本训练单元704,用于采用前向传播算法和后向传播算法,利用训练样本学习出构建的网络参数,即用于测试过程的模型。

[0075] 在本发明中,所述前向传播算法,首先将输入图像的大小归一化为 $3 \times 480 \times 640$,截取 $3 \times 448 \times 448$ 大小的patch和相应的标注信息作为卷积神经网络的输入,经过卷积层、降采样层和矫正线性单元层(ReLU Nonlinearity Layer),在Fire9层,图像特征图大小为 $512 \times 60 \times 80$;在Fire12层,特征图大小为 $512 \times 30 \times 40$,在后面两个分支特征图大小依次是 $512 \times 15 \times 20$ 和 $512 \times 8 \times 10$ 。在不同特征图上,采用卷积的方式得到目标框图的四个坐标点和类别信息,以Fire9层为例,假定只检测行人和背景,则输出为特征大小为 $6 \times 60 \times 80$,其中6包含背景、行人两个类别和候选框图四个坐标点。在目标检测网络中,将各个分支层得到的候选框图在proposals节点进行汇总,同时与Fire9层的背景语义信息和上采样信息经过感兴趣区域的池化操作得到的特征进行叠加,做最后的框图回归和类别回归。

[0076] 所述后向传播算法,需要先求出正向传播预测的目标框图与图像实际目标框图的损失函数 $\mathcal{L}(W)$,然后求得其对参数W的梯度,采用梯度下降的算法更新W以最小化损失函数 \mathcal{L} 。假定中间层有M个分支可以输出目标候选区域(M个尺度的感知域可以近似的检测出图像中所有目标物体), l^m 表示分支m的损失函数, α_m 表示 l^m 函数的权重, $S = \{S^1, S^2, \dots, S^M\}$ 指相应尺度的目标对象,则损失函数 $\mathcal{L}(W)$ 可定义为:

$$[0077] \quad \mathcal{L}(W) = \sum_{m=1}^M \sum_{i \in S^m} \alpha_m l^m(X_i, Y_i | W)$$

[0078] 所述损失函数,对于特定的检测层 m ,只有目标尺度在 m 所能检测的范围内,才对损失函数有贡献,故将损失函数定义为

$$[0079] \quad l^m(X, Y | W) = L_{cls}(p(X), y) + \lambda[y \geq 1]L_{loc}(b, \hat{b})$$

[0080] 其中, $p(X) = (p_0(X), \dots, p_K(X))$ 为目标类别的概率分布。损失函数中,使用交叉熵损失函数定义类别回归,即

$$[0081] \quad L_{cls}(p(X), y) = -\log_y(P(X))$$

[0082] 使用smooth L1 criterion进行目标框图的回归,定义如下

$$[0083] \quad L_{loc}(b, \hat{b}) = \frac{1}{4} \sum_{j \in \{x, y, w, h\}} \text{smooth}_{L_1}(b_j, \hat{b}_j)$$

[0084] 检测单元71,用于输入测试样本,通过训练好的模型利用神经网络感知域的变化规律使用不同的中间层对不同尺度范围内的目标物体(如行人)进行检测,预测出图像中目标物体(如行人)的框图。

[0085] 具体地,如图9所示,检测单元71进一步包括:

[0086] 模型载入单元710,用于载入训练好的模型;

[0087] 测试样本输入单元711,用于输入测试样本;

[0088] 图像预测单元712,用于利用训练好的模型,通过训练好的模型利用神经网络感知域的变化规律使用不同的中间层对不同尺度范围内的行人进行检测,预测图像中行人的框图。具体地,图像预测单元712利用模型中的目标候选网络,在Squeeze VGG-16卷积神经网络基础上,根据卷积层特征,在Fire9、Fire12、conv6以及增加的pooling层共计4层产生网络分支,进行不同尺度检测到物体的目标候选区域;然后利用目标检测网络,在目标候选区域的基础上,将目标候选区域1.5倍大小的图片区域作为目标的背景语义信息,将Fire9层的特征图进行一次上采样,作为增强对小物体感知的信息,将背景语义信息与上采样信息经过感兴趣区域的池化获得固定大小的特征,之后增加一层全连接层,进行类别和最终候选框的回归。

[0089] 综上所述,本发明一种快速行人检测方法及装置借鉴压缩网络的方法,调整并训练VGG-16的网络得到适应嵌入式系统要求的squeeze VGG-16网络,有效降低了网络模型的参数量并加快了计算效率;另一方面,针对传统检测方法中感知域与物体大小不一致的问题,本发明利用神经网络感知域的变化规律(即神经网络层越深,感知域越大,适合检测大一些的目标物体),使用不同的中间层对特定尺度范围内的目标物体进行检测,更好的适应了感知域与物体大小的关系,有效提高了检测结果;另外,为了增强对小物体的检测,本发明使用去卷积的方法对特定网络层的特征图进行放大,相比于传统图片放大的方法,几乎不增加显存和计算量;为了增强对于模糊物体的检测,在该层的特征图上,使用目标对象1.5倍大小的区域作为背景语义特征增加到网络中,对于模糊物体和远距离小物体的检测,有着极佳的性能。

[0090] 上述实施例仅例示性说明本发明的原理及其功效,而非用于限制本发明。任何本领域技术人员均可在不违背本发明的精神及范畴下,对上述实施例进行修饰与改变。因此,本发明的权利保护范围,应如权利要求书所列。

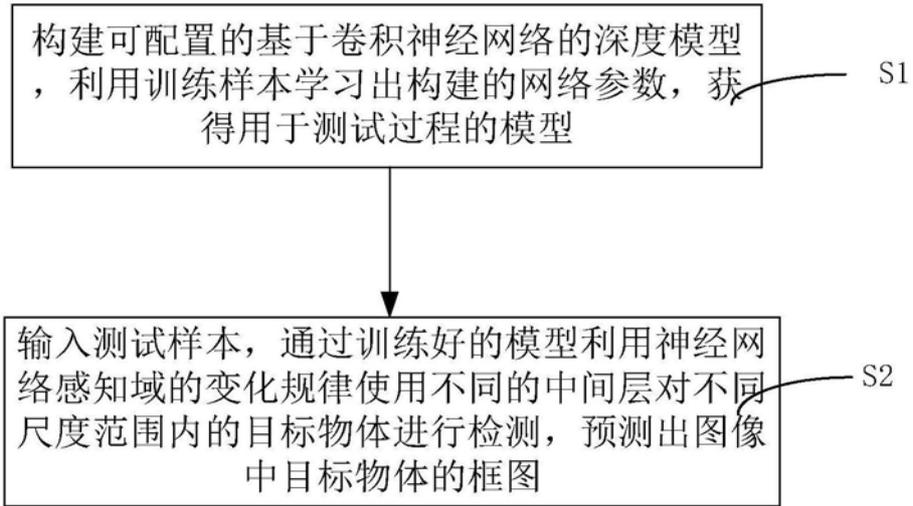


图1

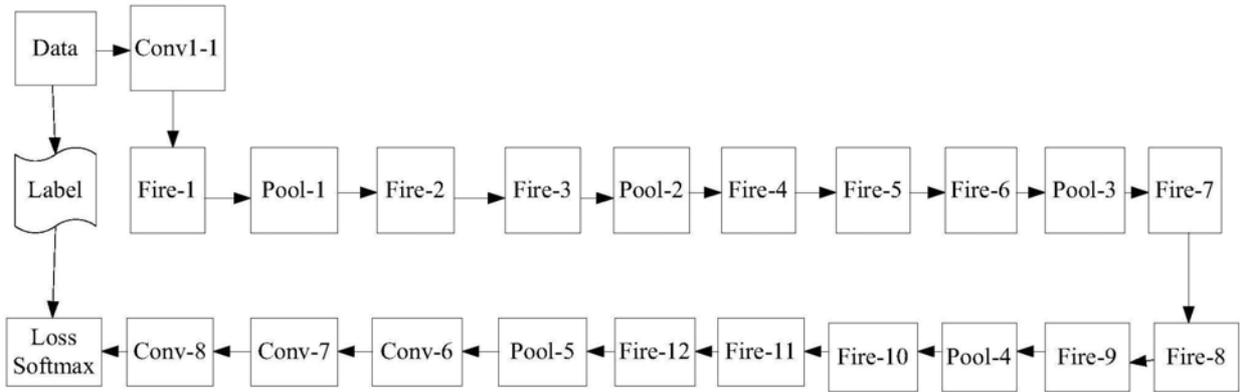


图2

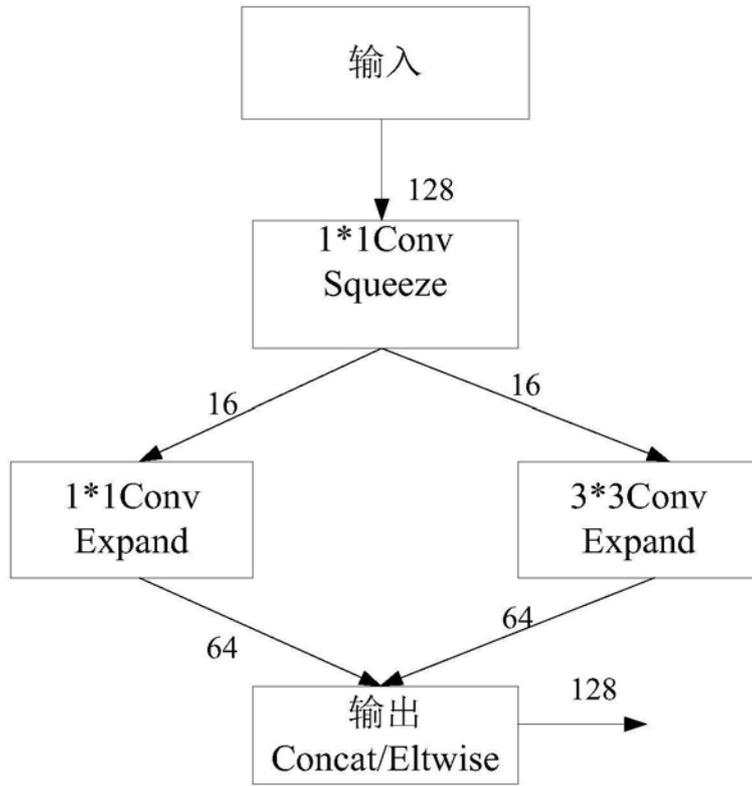


图3

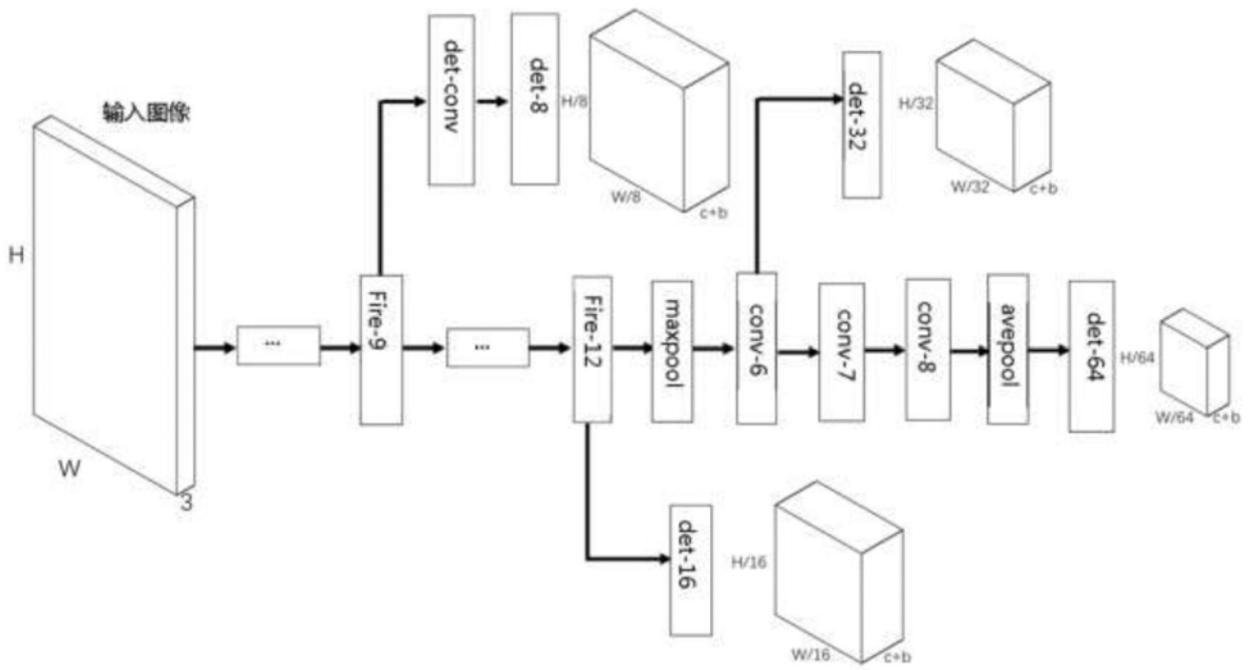


图4

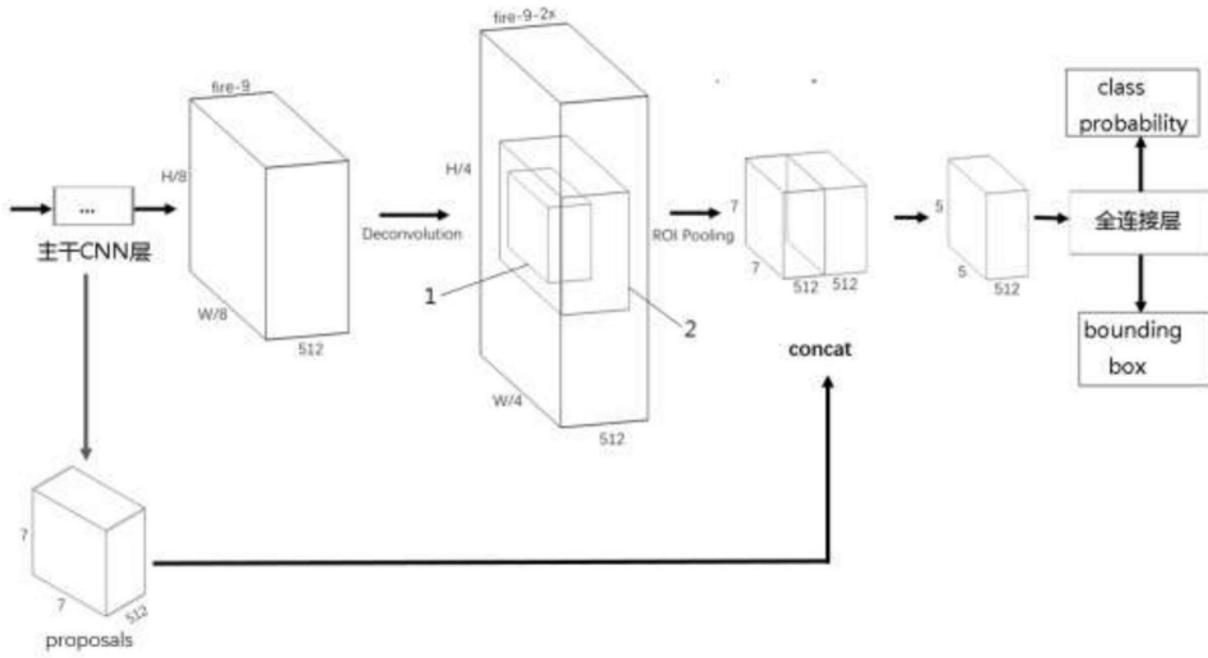


图5

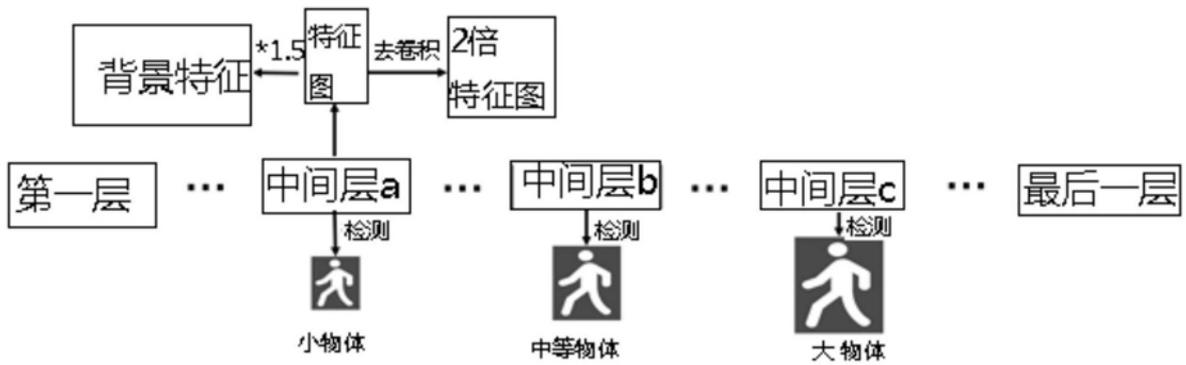


图6

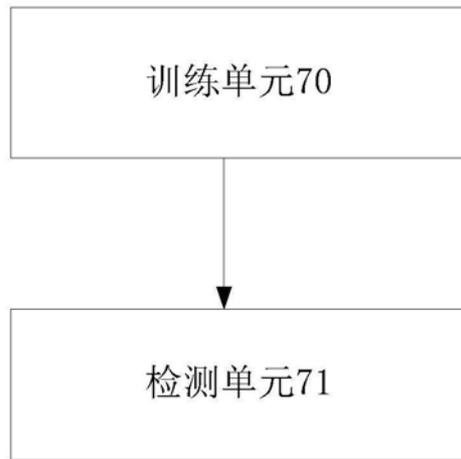


图7

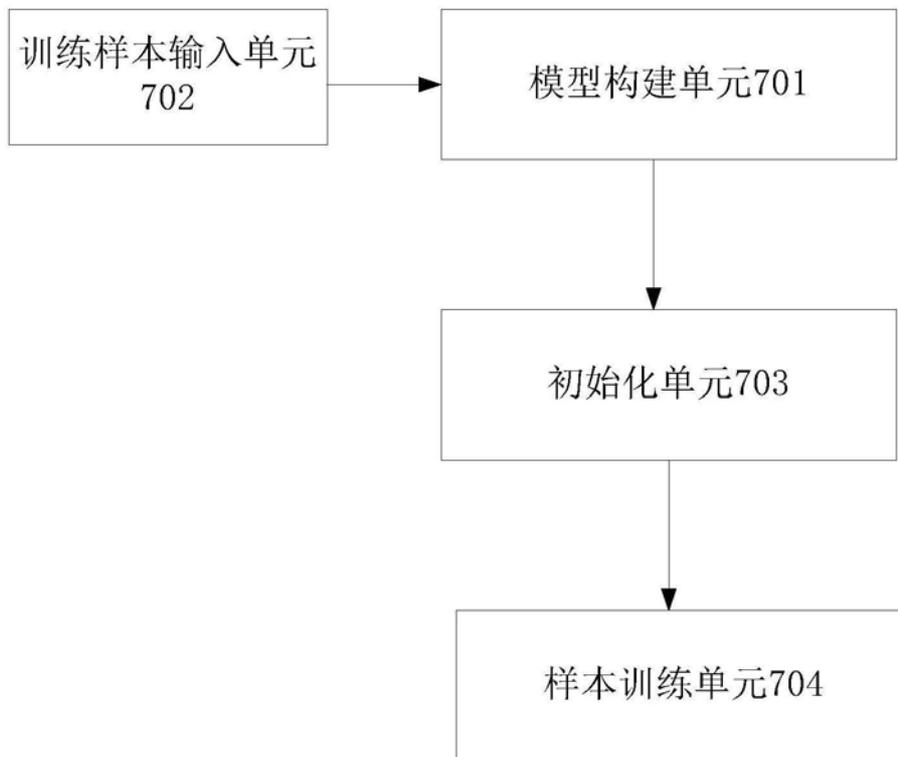


图8

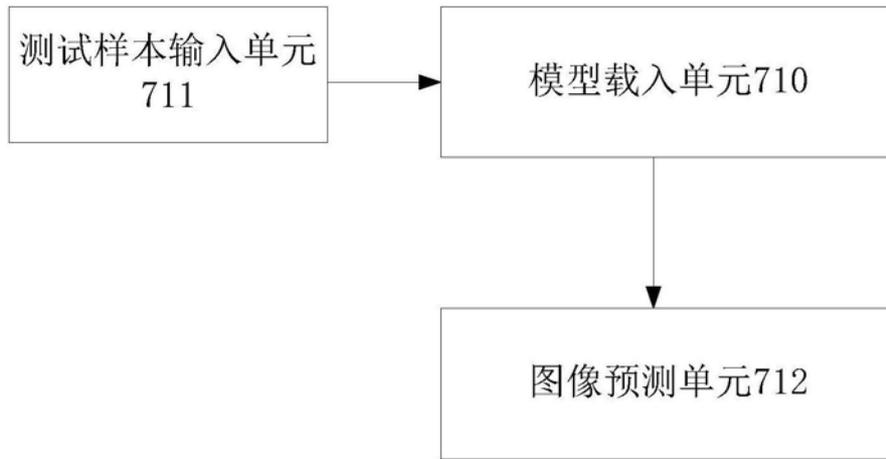


图9