



# (12)发明专利申请

(10)申请公布号 CN 107169421 A

(43)申请公布日 2017.09.15

(21)申请号 201710259285.4

(22)申请日 2017.04.20

(71)申请人 华南理工大学

地址 510640 广东省广州市天河区五山路  
381号

(72)发明人 林耀荣 陈康

(74)专利代理机构 广州市华学知识产权代理有  
限公司 44245

代理人 罗观祥

(51)Int.Cl.

G06K 9/00(2006.01)

G06K 9/62(2006.01)

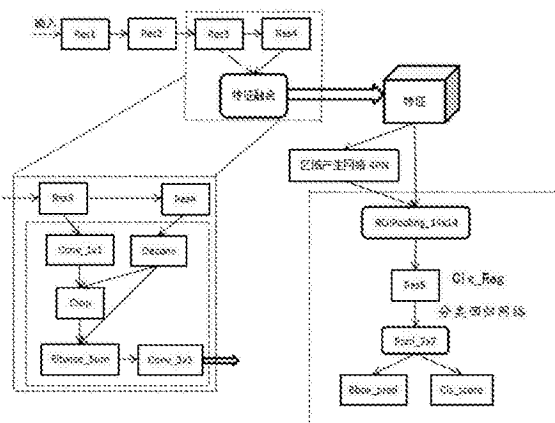
权利要求书3页 说明书7页 附图2页

## (54)发明名称

一种基于深度卷积神经网络的汽车驾驶场景目标检测方法

## (57)摘要

本发明公开了一种基于深度卷积神经网络的汽车驾驶场景目标检测方法,该检测方法包含如下优化技术:首先采用聚类算法对训练数据的目标宽高进行聚类,采用聚类中心优化锚点设置;其次采用分类性能更强的残差网络作为特征提取网络,在此基础上设计一个融合特征网络,利于小目标的检测;其次采用在线困难样本挖掘算法以及均衡正负样本候选区域来优化网络的训练;最后采用训练得到的模型处理图像,并对输出进行后处理优化,实现汽车驾驶场景下的目标检测功能。本发明基于CAFFE深度学习框架以及Faster RCNN算法,通过优化锚点设置、设计融合特征网络、优化训练过程以及对结果后处理,得到了一种高性能的基于深度卷积网络的汽车驾驶场景目标检测方法。



1. 一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于, 所述检测方法包括下列步骤:

S1、通过安装在汽车上的摄像机采集训练图像数据, 并对所述训练图像数据进行预处理;

S2、利用聚类算法对训练集样本的目标宽高进行聚类, 利用聚类中心优化锚点设置;

S3、以CAFFE框架为基础, 基于分类性能更好的残差网络, 采用多层融合特征构建深度卷积神经网络模型;

S4、网络初始化, 利用在ImageNet上预训练过的模型参数对相应网络层进行初始化, 其他新添加的层采用零均值, 标准差为 $\sigma$ 的高斯分布随机初始化;

S5、网络训练, 将图像以及真值信息输入上述构造的网络, 利用聚类得到的锚点设置, 在网络中进行前向计算得到预测值, 计算边界框的预测值与真值之间的Smooth L1损失以及目标类别的预测值与真值之间的SoftMax损失, 再进行反向传播利用随机梯度下降法更新网络权重;

S6、通过汽车上的摄像机采集车辆前方的图像, 输入网络进行检测, 即可输出目标的类别及坐标, 实现目标检测功能。

2. 根据权利要求1所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于,

利用K-Means算法对训练图像数据集中目标的宽高进行聚类, 根据聚类结果定义锚点的宽高, 为了使聚类中心框更好地匹配真值框, 使聚类中心与真值框的交并比IOU最大化, 聚类距离度量定义为:

$$D(gt, centroid) = 1 - IOU(gt, centroid),$$

其中, IOU定义为两个坐标框的交并比:

$$IOU(gt, centroid) = \frac{gt \cap centroid}{gt \cup centroid},$$

其中gt代表真值框, centroid代表聚类中心框,  $gt \cap centroid$ 代表真值框与聚类中心重叠的区域,  $gt \cup centroid$ 表示真值与聚类中心的区域合集, 由此得到的聚类中心定义锚点的宽高。

3. 根据权利要求1所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于,

所述锚点用于候选区域生成网络RPN产生候选区域, 在网络训练阶段, 利用锚点与真值计算候选区域坐标的参考值作为网络回归的真值, 计算如下:

$$x_p^* = (x^* - x_a) / w_a, \quad y_p^* = (y^* - y_a) / h_a, \quad w_p^* = \log(w^* / w_a), \quad h_p^* = \log(h^* / h_a)$$

其中 $x^*$ 、 $x_a$ 分别代表真值以及锚点的中心坐标,  $x_p^*$ 代表计算得到的网络回归真值的中心坐标,  $x$ 、 $y$ 指边界框的中心坐标,  $w$ 、 $h$ 指边界框的宽和高;

在进行目标检测时, 利用锚点与网络回归层的预测值计算候选区域的预测值, 计算如下:

$$x = x_p * w_a + x_a, \quad y = y_p * h_a + y_a, \quad w = w_a * \exp(w_p), \quad h = h_a * \exp(h_p)$$

其中,  $x_p$ 代表网络回归层的预测值,  $x$ 代表由网络预测值与锚点计算得到的候选区域的

预测值。

4. 根据权利要求1所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于, 所述步骤S3过程如下:

以残差网络为特征提取网络, 将残差网络分为五个单元: Res1、Res2、Res3、Res4、Res5, 将前四个单元作为共享的特征提取网络, 通过在第三个残差单元Res3设置侧连接, 将输出调整为与Res4输出相同的通道, 并对Res4进行上采样使输出的宽高加倍, 然后将两个单元的输出相加, 实现多层特征融合。

5. 根据权利要求4所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于,

所述侧连接采用卷积层Conv\_1x1, 该卷积层的卷积核大小为1x1, 步长为1, 填充为0, 输出通道为1024;

所述上采样采用反卷积层Deconv, 该上采样使Res4输出宽高加倍;

采用Crop层将Res4输出维度调整为与Res3输出相同的维度;

采用Eltwise\_Sum将两个单元的输出相加, 实现融合。

6. 根据权利要求4所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于,

候选区域生成网络RPN由全卷积网络组成, 分类回归网络采用Res5单元作为网络的一部分。

7. 根据权利要求1所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于,

在网络训练阶段中, 利用在线困难样本挖掘算法以及正负样本均衡优化训练过程, 其中, 所述在线困难样本挖掘算法利用分类回归网络Cls\_Reg的拷贝Cls\_Reg\_ReadOnly先对所有候选区域进行前向传播, 选取损失最大的候选区域作为困难样本, 再将困难样本输入Cls\_Reg进行反向传播更新权重, 所述正负样本候选区域均衡指控制负正样本候选区域数量比例小于 $\mu$ , 防止负样本区域过多而正样本区域过少导致分类器性能下降。

8. 根据权利要求1所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于, 所述检测方法还包括下列步骤:

对网络输出的检测边界框预测值进行后处理优化, 所述后处理优化是对输出基于边界框置信度投票优化, 具体过程包括:

首先对检测结果进行非极大值抑制, 得到局部区域置信度最高的检测框 ( $score_i, bbox_i$ );

然后利用邻域中与其交并比IOU大于0.5的检测框 ( $score_j, bbox_j$ ), 对输出的bbox\_i进行加权优化, 计算如下:

$$bbox_i = \frac{\sum_j score_j \cdot bbox_j}{\sum_j score_j}$$

其中,  $score$ 代表网络预测的置信度,  $bbox$ 代表网络输出的目标边界框位置坐标。

9. 根据权利要求1所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法, 其特征在于,

所述训练图像数据的预处理包括对输入图像进行镜像翻转、减去均值、尺度调整。

10. 根据权利要求1所述的一种基于深度卷积神经网络的汽车驾驶场景目标检测方法，其特征在于，

所述训练图像数据需要标注目标，包括：车、行人、骑自行车的人，其中，需要标注的信息包括：目标的类别以及包围该目标的边界框坐标。

## 一种基于深度卷积神经网络的汽车驾驶场景目标检测方法

### 技术领域

[0001] 本发明涉及图像处理和计算机视觉技术领域,具体涉及一种基于深度卷积神经网络的汽车驾驶场景目标检测方法。

### 背景技术

[0002] 近年来,交通问题日益突出,各类智能车概念被提出,车辆驾驶辅助系统乃至无人车均获得了空前的发展。汽车驾驶场景中的目标检测系统是其中关键的组成部分。因此,汽车驾驶场景中的目标检测成为当前研究热点。在目标检测任务中,潜在目标往往受到姿态多变、光照变化、尺度不一甚至部分遮挡等不利因素的影响,导致目标检测本身就是很有挑战性的任务。而在汽车驾驶场景中,交通状况复杂多变,各类目标物存在相互遮挡,光照变化影响更加复杂,这些不利因素都进一步制约着汽车驾驶场景目标检测系统的性能。

[0003] 在目标检测中,如何提取高效的特征是提高系统性能的关键。传统的特征提取技术,如Haar特征以及梯度方向直方图特征HOG,这些人工设计的特征描述子缺乏高层语义信息,制约着目标检测精度的提高。经典的目标检测框架主要是基于滑动窗口的模板匹配,如可变形部件模型DPM(Deformable Part based Model),该算法采用梯度方向直方图特征,利用隐支持向量机训练多个模型,通过复杂的模型训练与匹配达到检测的目的。由于采用滑动窗口匹配技术,以及为了使模型能够匹配不同尺度的目标而构建图像特征金字塔,导致模型匹配计算量很大。另外,由于HOG难以提取高效的特征,制约着精度的提升。因此,基于人工设计的特征描述子以及传统模板匹配的目标检测算法在精度以及效率方面均有较大瓶颈。

[0004] 深度卷积神经网络在上世纪九十年代被提出,近几年得益于高性能运算单元的出现,在图像处理领域获得了成功。通过构建高性能的卷积神经网络,可以提取表述性更强、语义信息更丰富的深度卷积特征。基于深度卷积神经网络的目标检测算法是目前发展的方向。目前,基于深度卷积网络的目标检测通常利用分类以及回归达到检测的目的,如RCNN算法将检测分为候选区域提取以及进一步分类回归两个阶段,YOLO算法直接利用回归完成目标检测。RCNN算法及其后续算法Fast RCNN、Faster RCNN在通用目标检测领域取得了较好的结果,常用的特征提取网络包括ZF网络或VGG16网络等。YOLO算法时间效率较高但检测小目标能力较弱。由于在汽车驾驶场景下的目标检测存在如上所述的种种挑战,基于深度卷积神经网络的目标检测算法在汽车驾驶场景下的性能还有很大的提升空间。

### 发明内容

[0005] 本发明的目的是为了解决现有技术中的上述缺陷,提供一种基于深度卷积神经网络的汽车驾驶场景目标检测方法,以便进一步提升在汽车驾驶场景下的目标检测性能。通过设计性能更佳的网络、优化训练过程、挖掘先验信息以及对结果进行后处理,本发明能够对汽车驾驶场景下的各类目标进行有效检测。

[0006] 本发明的目的可以通过采取如下技术方案达到:

[0007] 一种基于深度卷积神经网络的汽车驾驶场景目标检测方法,所述检测方法包括下列步骤:

[0008] S1、通过安装在汽车上的摄像机采集训练图像数据,并对所述训练图像数据进行预处理;

[0009] S2、利用聚类算法对训练集样本的目标宽高进行聚类,利用聚类中心优化锚点设置;

[0010] S3、以CAFFE框架为基础,基于分类性能更好的残差网络,采用多层融合特征构建深度卷积神经网络模型;

[0011] S4、网络初始化,利用在ImageNet上预训练过的模型参数对相应网络层进行初始化,其他新添加的层采用零均值,标准差为 $\sigma$ 的高斯分布随机初始化;

[0012] S5、网络训练,将图像以及真值信息输入上述构造的网络,利用聚类得到的锚点,在网络中进行前向计算得到预测值,计算边界框的预测值与真值之间的Smooth L1损失以及目标类别的预测值与真值之间的SoftMax损失,再进行反向传播利用随机梯度下降法更新网络权重;

[0013] S6、通过汽车上的摄像机采集车辆前方的图像,输入网络进行检测,即可输出目标的类别及坐标,实现目标检测功能。

[0014] 进一步地,利用K-Means算法对训练图像数据集中目标的宽高进行聚类,根据聚类结果定义锚点的宽高,为了使聚类中心框更好地匹配真值框,使聚类中心与真值框的交并比IOU最大化,聚类距离度量定义为:

[0015]  $D(gt, centroid) = 1 - IOU(gt, centroid)$ ,

[0016] 其中,IOU定义为两个坐标框的交并比:

[0017]  $IOU(gt, centroid) = \frac{gt \cap centroid}{gt \cup centroid}$ ,

[0018] 其中gt代表真值框,centroid代表聚类中心框, $gt \cap centroid$ 代表真值框与聚类中心重叠的区域, $gt \cup centroid$ 表示真值与聚类中心的区域合集,由此得到的聚类中心定义锚点的宽高。

[0019] 进一步地,所述锚点用于候选区域生成网络RPN产生候选区域,在网络训练阶段时,利用锚点与真值计算候选区域坐标的参考值作为网络回归的真值,计算如下:

[0020]  $x_p^* = (x^* - x_a) / w_a$ ,  $y_p^* = (y^* - y_a) / h_a$ ,  $w_p^* = \log(w^* / w_a)$ ,  $h_p^* = \log(h^* / h_a)$

[0021] 其中 $x^*$ 、 $x_a$ 分别代表真值以及锚点的中心坐标, $x_p^*$ 代表计算得到的网络回归真值的中心坐标, $x$ 、 $y$ 指边界框的中心坐标, $w$ 、 $h$ 指边界框的宽和高;

[0022] 在进行目标检测时,利用锚点与网络回归层的预测值计算候选区域的预测值,计算如下:

[0023]  $x = x_p * w_a + x_a$ ,  $y = y_p * h_a + y_a$ ,  $w = w_a * \exp(w_p)$ ,  $h = h_a * \exp(h_p)$

[0024] 其中, $x_p$ 代表网络回归层的预测值, $x$ 代表由网络预测值与锚点计算得到的候选区域的预测值。

[0025] 进一步地,所述步骤S3过程如下:

[0026] 以残差网络为特征提取网络,将残差网络分为五个单元:Res1、Res2、Res3、Res4、

Res5,将前四个单元作为共享的特征提取网络,通过在第三个残差单元Res3设置侧连接,将输出调整为与Res4输出相同的通道,并对Res4进行上采样使输出的宽高加倍,然后将两个单元的输出相加,实现多层特征融合。

[0027] 进一步地,所述侧连接采用卷积层Conv\_1x1,该卷积层的卷积核大小为1x1,步长为1,填充为0,输出通道为1024;

[0028] 所述上采样采用反卷积层Deconv,该上采样使Res4输出宽高加倍;

[0029] 采用Crop层将Res4输出维度调整为与Res3输出相同的维度;

[0030] 采用Eltwise\_Sum将两个单元的输出相加,实现融合。

[0031] 进一步地,候选区域生成网络RPN由全卷积网络组成,分类回归网络采用Res5单元作为网络的一部分。

[0032] 进一步地,在网络训练阶段中,利用在线困难样本挖掘算法以及正负样本均衡优化训练过程,其中,所述在线困难样本挖掘算法利用分类回归网络Cls\_Reg的拷贝Cls\_Reg Readonly先对所有候选区域进行前向传播,选取损失最大的候选区域作为困难样本,再将困难样本输入Cls\_Reg进行反向传播更新权重,所述正负样本候选区域均衡指控制负正样本候选区域数量比例小于 $\mu$ ,防止负样本区域过多而正样本区域过少导致分类器性能下降。

[0033] 进一步地,所述检测方法还包括下列步骤:

[0034] 对网络输出的检测边界框预测值进行后处理优化,所述后处理优化是对输出基于边界框置信度投票优化,具体过程包括:

[0035] 首先对检测结果进行非极大值抑制,得到局部区域置信度最高的检测框( $score_i, bbox_i$ );

[0036] 然后利用邻域中与其交并比IOU大于0.5的检测框( $score_j, bbox_j$ ),对输出的 $bbox_i$ 进行加权优化,计算如下:

$$bbox_i = \frac{\sum_j score_j \cdot bbox_j}{\sum_j score_j}$$

[0038] 其中, $score$ 代表网络预测的置信度, $bbox$ 代表网络输出的目标边界框位置坐标。

[0039] 进一步地,所述训练图像数据的预处理包括对输入图像进行镜像翻转、减去均值、尺度调整。

[0040] 进一步地,所述训练图像数据需要标注目标,包括:车、行人、骑自行车的人,其中,需要标注的信息包括:目标的类别以及包围该目标的边界框坐标。

[0041] 本发明相对于现有技术具有如下的优点及效果:

[0042] 1、对于特定的汽车驾驶场景,利用K-Means算法优化锚点设置,达到了匹配场景中多种尺度及比例的目标的目的,同时有效减少网络参数,提升检测效率的同时又提高检测精度;

[0043] 2、对于图像特征的提取,设计了融合多层特征的深度神经网络,有利于汽车驾驶场景众多小目标的检测;

[0044] 3、优化训练过程,利用在线困难样本挖掘算法以及正负样本均衡来优化训练过程,提高模型的特征提取能力以及分类检测能力;

[0045] 4、对结果进行后处理,使输出更加准确。

## 附图说明

- [0046] 图1是本发明所提出的网络结构；  
 [0047] 图2是本发明所提出的结合在线困难样本挖掘算法的网络结构；  
 [0048] 图3(a)是本发明实施例中检测结果示例一；  
 [0049] 图3(b)是本发明实施例中检测结果示例二；  
 [0050] 图3(c)是本发明实施例中检测结果示例三。

## 具体实施方式

[0051] 为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0052] 实施例

[0053] 本实施例公开了一种基于深度卷积神经网络的汽车驾驶场景目标检测方法，具体步骤如下：

[0054] S1、训练数据采集，通过安装在汽车上的摄像机采集训练图像数据，并对所述训练图像数据进行预处理。

[0055] 训练图像数据由安装在汽车上的摄像机采集得到，训练图像预处理包括对输入图像进行镜像翻转、减去均值、尺度调整等操作，训练集中的图片需要标注车、行人、骑自行车的人三类目标。其中需要标注的信息包括目标的类别以及包围该目标的边界框坐标。

[0056] S2、利用聚类算法对训练集样本的目标宽高进行聚类，利用聚类中心优化锚点设置；

[0057] 利用K-Means算法对训练图像数据集中目标的宽高进行聚类，根据聚类结果定义锚点的宽高。为了使聚类中心框更好地匹配真值框，使聚类中心与真值框的交并比IOU最大化，聚类距离度量定义为：

[0058]  $D(gt, centroid) = 1 - IOU(gt, centroid)$ ，

[0059] 其中，IOU定义为两个坐标框的交并比：

[0060]  $IOU(gt, centroid) = \frac{gt \cap centroid}{gt \cup centroid}$ ，

[0061] 其中gt代表真值框，centroid代表聚类中心框， $gt \cap centroid$ 代表真值框与聚类中心重叠的区域， $gt \cup centroid$ 表示真值与聚类中心的区域合集。由此得到的聚类中心定义锚点的宽高。锚点是指目标的参考边界框。

[0062] 表1.利用聚类算法得到的锚点



[0063]

锚点	1	2	3	4	5	6	7	8
尺度	370*185	223*11	125*5	182*7	83*3	21*56	43*29	70*163
比例	2:1	1.9:1	2.5:1	2.4:1	2.4:1	0.4:1	1.5:1	0.4:1
锚点	9	10	11	12	13	14	15	
尺度	126*88	24*22	56*44	83*62	39*9	136*185	260*1	
比例	1.4:1	1.1:1	1.3:1	1.3:1	0.4:1	0.7:1	1.5:1	

[0064] 针对涉及到的训练数据,设置15个聚类中心,得到锚点的尺度比例如表1所示。其中,尺度指区域面积 $w*h$ ,比例指 $w:h$ 。由表1中数据可以看出,锚点呈现多种尺度以及多种比例。由于相同比例的目标可能因与摄像机距离的不同而在成像时呈现不同的尺度,因此存在比例相同但尺度不同的锚点。

[0065] 所述锚点用于候选区域生成网络RPN产生候选区域。在网络训练阶段,需要利用锚点与真值计算候选区域坐标的参考值作为网络回归的真值,计算如下:

$$[0066] \quad x_p^* = (x^* - x_a) / w_a, \quad y_p^* = (y^* - y_a) / h_a, \quad w_p^* = \log(w^* / w_a), \quad h_p^* = \log(h^* / h_a)$$

[0067] 其中 $x^*$ 、 $x_a$ 分别代表真值以及锚点的中心坐标, $x_p^*$ 代表计算得到的网络回归真值的中心坐标, $x$ 、 $y$ 指边界框的中心坐标, $w$ 、 $h$ 指边界框的宽和高。

[0068] 系统进行目标检测时,利用锚点与网络回归层的预测值计算候选区域的预测值,计算如下:

$$[0069] \quad x = x_p * w_a + x_a, y = y_p * h_a + y_a, w = w_a * \exp(w_p), h = h_a * \exp(h_p)$$

[0070] 其中, $x_p$ 代表网络回归层的预测值, $x$ 代表由网络预测值与锚点计算得到的候选区域的预测值。

[0071] S3、以CAFFE框架为基础,基于分类性能更好的残差网络,采用多层融合特征构建深度卷积神经网络模型;

[0072] 以CAFFE框架为基础,基于分类性能更好的残差网络,设计一个融合多层特征的深度卷积神经网络,按图1所示网络结构构建上述技术方案中所述的深度卷积神经网络模型,利用该网络处理输入图像,预测目标的类别以及坐标。

[0073] 具体的,以残差网络ResNet101、ResNet50或者ResNet152为特征提取网络,将残差网络分为五个单元:Res1、Res2、Res3、Res4、Res5,将前四个单元作为共享的特征提取网络。通过在第三个残差单元Res3设置侧连接,将输出调整为与Res4输出相同的通道,并对Res4进行上采样使输出的宽高加倍,然后将两个单元的输出相加,实现多层特征融合。融合后特征有利于小目标的检测。

[0074] 其中,侧连接指卷积层Conv\_1x1,该卷积层的卷积核大小为1x1,步长为1,填充为0,输出通道为1024;Deconv为反卷积层,作用是对Res4进行上采样,使Res4输出宽高加倍;Crop层将Res4输出维度调整为与Res3相同的维度;Eltwise\_Sum对两个输入进行相加操作,实现融合;Conv\_3x3为卷积层,其卷积核大小为3x3,填充1,步长1,输出通道1024,作用是对融合后的特征进行调整。候选区域生成网络RPN由全卷积网络组成,分类回归网络采用Res5

单元作为网络的一部分。

[0075] S4、网络初始化,利用在ImageNet上预训练过的残差网络模型参数对相应网络层进行初始化,其他新添加的层采用零均值,标准差为 $\sigma$ 的高斯分布随机初始化。本实施例中,标准差 $\sigma$ 取值为0.01,但该取值并不构成对本技术方案的限制。

[0076] S5、网络训练,将图像以及真值信息输入上述构造的网络,利用聚类得到的锚点,在网络中进行前向计算得到预测值,计算边界框的预测值与真值之间的Smooth L1损失以及目标类别的预测值与真值之间的SoftMax损失,再进行反向传播,利用随机梯度下降法SGD更新网络权重。

[0077] 在训练过程中,利用在线困难样本挖掘算法(OHEM-Online Hard Example Mining)以及正负样本均衡优化训练过程。在线困难样本挖掘算法是指在训练过程中,利用分类回归网络Cls\_Reg的拷贝Cls\_Reg\_ReadOnly先对所有候选区域进行前向传播,选取损失最大的候选区域作为困难样本,再将困难样本输入Cls\_Reg进行反向传播更新权重。正负样本候选区域均衡指控制负正样本候选区域数量比例小于 $\mu$ ,防止负样本区域过多而正样本区域过少导致分类器性能下降。本实施例中, $\mu$ 取值为3,但该取值并不构成对本技术方案的限制。

[0078] 结合在线困难样本挖掘算法的网络结构如图2所示,其中虚线表示只有前向传播,实线代表既进行前向传播又进行反向传播;Cls\_Reg\_ReadOnly代表分类回归网络的拷贝,对所有候选区域进行前向传播并在线困难样本挖掘,产生的困难样本输入分类回归网络Cls\_Reg进行反向传播,更新网络权重。

[0079] 训练过程中迭代次数为40000次,初始学习率0.001,每一万次减小为上一个值的十分之一,momentum设置为0.9。

[0080] 网络的回归损失定义为:

$$[0081] \quad L_{reg}(x_p, x_p^*) = \sum_i \text{smooth}_{L_1}(x_p - x_p^*)$$

[0082] 其中,

$$[0083] \quad \text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

[0084] 其中, $x_p$ 指网络回归层的预测值, $x_p^*$ 指网络回归真值。

[0085] S6、网络模型训练好以后,通过汽车上的摄像机采集车辆前方的图像,输入网络进行检测,即可输出目标的类别及坐标,实现目标检测功能。检测结果示例如图3(a)-图3(c)所示,具体实施时,检测结果为彩色图片。

[0086] 在进一步优选的实施方式中,对网络输出的检测边界框预测值进行后处理优化。

[0087] 后处理是指对输出基于边界框置信度投票优化。

[0088] 首先对检测结果进行非极大值抑制,得到局部区域置信度最高的检测框( $score_i, bbox_i$ );

[0089] 然后利用邻域中与其交并比IOU大于0.5的检测框( $score_j, bbox_j$ ),对输出的 $bbox_i$ 进行加权优化,计算如下:

$$[0090] \quad bbox_j = \frac{\sum_j score_j \cdot bbox_j}{\sum_j score_j}$$

[0091] 其中, score代表网络预测的置信度, bbox代表网络输出的目标边界框位置坐标。

[0092] 综上所述, 该检测方法包含如下优化技术: 首先采用聚类算法对训练数据的目标宽高进行聚类, 采用聚类中心优化锚点设置; 其次采用分类性能更强的残差网络作为特征提取网络, 在此基础上设计一个融合特征网络, 利于小目标的检测; 其次采用在线困难样本挖掘算法以及均衡正负样本候选区域来优化网络的训练; 最后采用训练得到的模型处理图像, 并对输出进行后处理优化, 实现汽车驾驶场景下的目标检测功能。本发明基于CAFFE深度学习框架以及Faster RCNN算法, 通过优化锚点设置、设计融合特征网络、优化训练过程以及对结果后处理, 得到了一种高性能的基于深度卷积网络的汽车驾驶场景目标检测方法。

[0093] 上述实施例为本发明较佳的实施方式, 但本发明的实施方式并不受上述实施例的限制, 其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化, 均应为等效的置换方式, 都包含在本发明的保护范围之内。

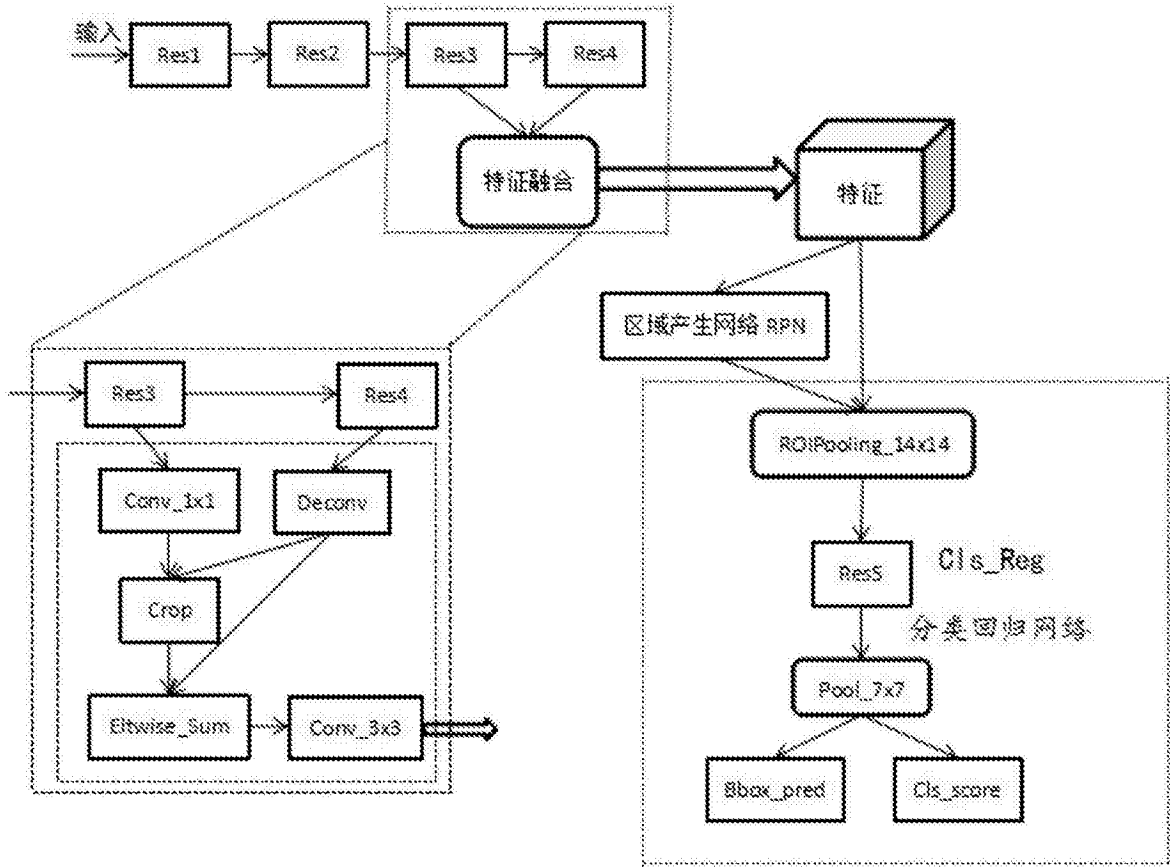


图1

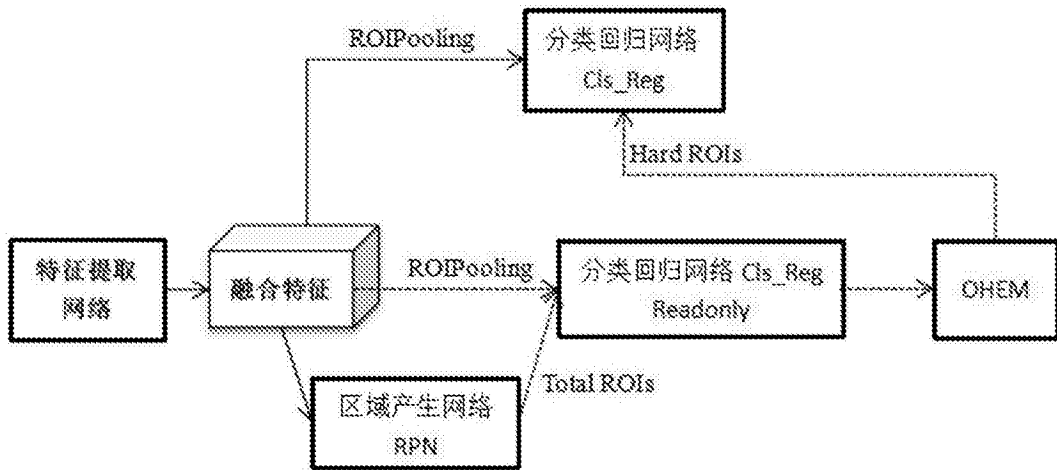


图2

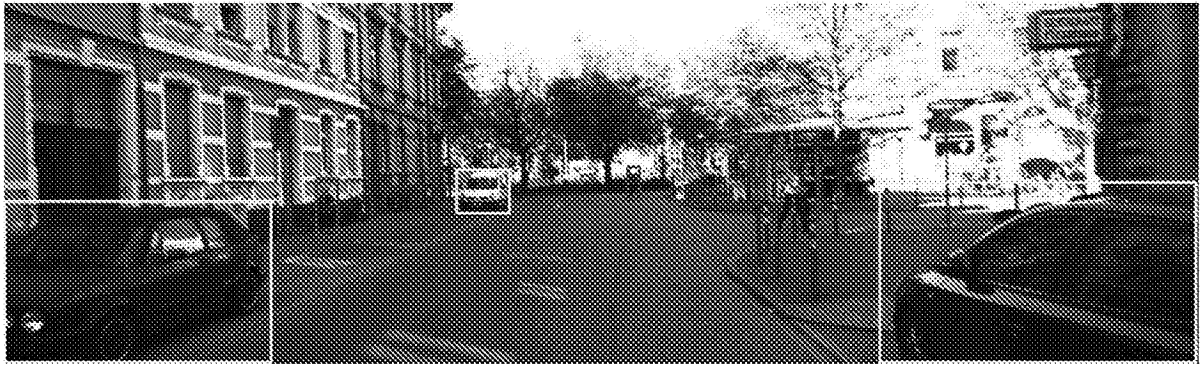


图3 (a)

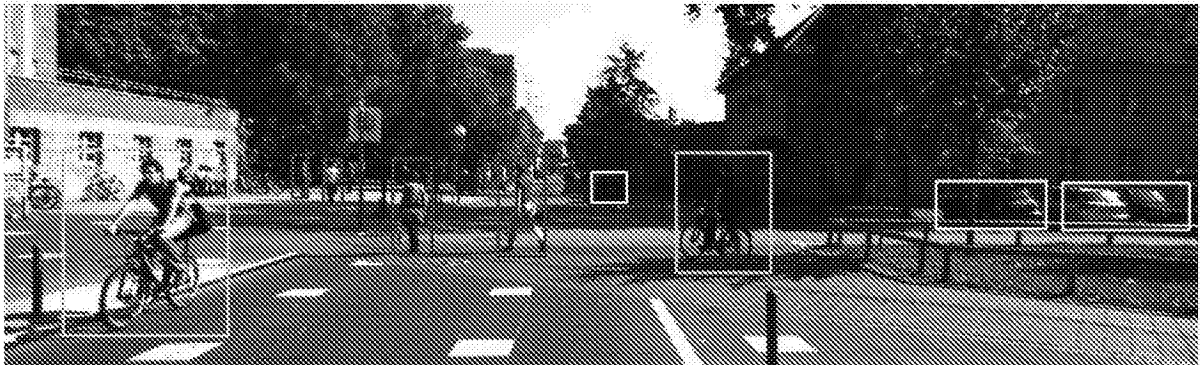


图3 (b)



图3 (c)