文章编号: 2096-1472(2021)-01-13-04

# 基于深度学习的实时车辆检测研究

# 黄生鹏, 范平清

(上海工程技术大学机械与汽车工程学院,上海 201620)
 №hsp0285@163.com; fanpingqing@163.com



摘 要:针对城市交通复杂场景下车辆检测存在准确率低的问题,提出改进SSD(单发多箱探测器)目标检测算法。 首先基于轻量化的PeleeNet(一种基于密集卷积网络的轻量化网络变体)网络结构改进SSD算法中VGG16(视觉几何群网络) 特征提取网络,在保证提取丰富特征的前提下,有效地减少模型参数,提高模型的实时性,其次设计了多尺度特征融合 模块和底层特征增强模块,提高特征的表达性能;最后根据数据集中目标的大小调整默认框的长宽比例,并在后三个特 征层的每个单元上增加默认框。实验结果表明,改进后的目标检测算法的准确率mAP(平均精度)为79.83%,与原始SSD 相比提高了2.25%,并验证了改进SSD算法的有效性。

关键词:实时性,SSD,默认框,特征提取 中图分类号:TP391 文献标识码:A



# **Research on Real-Time Vehicle Detection based on Deep Learning**

HUANG Shengpeng, FAN Pingqing

(Shanghai University of Engineering Science, School of Mechanical and Automotive Engineering, Shanghai 201620, China) Shanghai 201620, China)

**Abstract:** This paper proposes an improved SSD (Single Shot Multiple Box Detector) target detection algorithm to improve the vehicle detection accuracy in complex urban traffic scenarios. Firstly, VGG16 (Visual Geometry Group Network) feature extraction network algorithm in SSD is improved based on the lightweight PeleeNet (A lightweight network variant based on dense convolution network.) network structure. Under the premise of ensuring the extraction of rich features, it can effectively reduce model parameters and improve the real-time performance of the model. Secondly, a multi-scale feature fusion module and a low-level feature enhancement module are designed to improve the expression performance of features. Finally, length-width ratio of the default frame is adjusted according to the target size in data set, and the default frame is added to each cell of the last three feature layers. Experimental results show that the accuracy of mAP (mean Average Precision) of target detection algorithm is improved to 79.83%, which is 2.25% higher than the original SSD. So the improved SSD algorithm is verified to be effective.

Keywords: real-time; SSD; default frame; feature extraction

# 1 引言(Introduction)

高级辅助驾驶系统(Advanced Driver Assistance Systems, ADAS)是提高行车安全的重要技术手段之一,能否 对行车环境中前方行驶车辆进行有效的实时检测是完成ADAS 功能的首要前提<sup>11</sup>。

目前,视觉车辆检测技术主要分为两类:基于传统图 像特征的目标检测算法和基于深度学习的目标检测算法。 传统的目标检测算法主要采用人工设计的目标物特征,如 HOG(Histogram of Oriented Gradient)<sup>[2]</sup>、Haar(Haarlike features)<sup>[3]</sup>等,将这些特征送入SVM(Support Vector Machines)<sup>[4]</sup>等分类器中进行目标分类检测。基于深度学习 的目标检测算法主要有两个分支:以R-CNN(Region-based Convolutional Neural Networks)<sup>[5]</sup>、Fast R-CNN<sup>[6]</sup>、 Faster-RCNN(Towards Real-Time object Detection with Region Proposal Networks)<sup>[7]</sup>、R-FCN(Object detection via region-based fully convolution networks)<sup>[8]</sup>算法为代

基金项目: 上汽工业科技发展基金会项目 (1910); 新能源汽车振动噪声测试与控制专业技术服务平台(18DZ2295900).

表的基于候选框的两阶段目标检测算法,另一条路线是以 YOLO(You Only Look Once)<sup>[9]</sup>、SSD<sup>[10]</sup>、Retina-Net(Focal Loss for Dense Object Detection)<sup>[11]</sup>为代表的单阶段目标检测 算法。基于深度学习的单阶段目标检测算法是当前应用于车 辆实时性检测的主流方法,尤其是SSD目标检测框架的应用场 景广泛。

本文主要在多尺度目标检测算法SSD的基础上进行改进, 采用一种基于DenseNet(Dense Convolutional Network)<sup>[12]</sup>的轻 量化网络结构变体PeleeNet<sup>[13]</sup>,改进原始SSD目标检测算法。 在保证提取丰富特征的前提下降低计算成本,满足实时性要 求。针对实际工程项目采集的数据集进行目标尺寸分析,设 计合理的区域候选框大小及长宽比例,提高车辆目标的检测 准确率。

## 2 SSD算法(SSD algorithm)

## 2.1 SSD模型结构

SSD目标检测算法的主干网络结构为VGG16,在VGG16 的基础上新增加了不同卷积层来获得多尺度特征图,最终利 用这些特征图进行多尺度的目标检测。SSD网络结构如图1 所示,输入端是一张高为300,宽为300,通道数为R、G、B 的彩色图像;接着将两个全连接层fc6、fc7分别替换为卷积 层Conv6、Conv7;然后再添加四个卷积层。整个网络共获 得六个不同尺度的特征图:Conv4\_3、Conv7、Conv8\_2、 Conv9\_2、Conv10\_2、Conv11\_2。不同尺度的特征图所对应 的感受不一样,各特征图设置先验框的数量也不一样,针对 这六个特征层分别用两个并列的卷积核进行卷积,一个用来 预测边框,另一个用来获得不同类别的置信度。最终的检测 结果由非极大值抑制算法NMS(Non-Maximum Suppression) 输出。





#### 2.2 SSD候选框

SSD采用多尺度特征图预测的方法,就是借鉴Faster R-CNN中的anchor的理念,在特征图的feature map cell上 设置不同尺寸和不同长宽比的候选框。每个特征图上的候选 框按以下公式计算:

$$S_{k} = S_{\min} + \frac{S_{\max} - S_{\min}}{m - 1} (k - 1), \quad k \in [1, m]$$
(1)

式中, $S_k$ 代表第k个特征图默认框大小相对于网络输入的比例,是一个归一化的值。 $S_{min}$ 代表最小的比例, $S_{max}$ 代表最大的比例。k代表m个特征图中的第k个,其中m为特征图的个数。

默认框在不同的特征层有不同的尺寸,而且在同一个特征层又有不同的宽高比。共有 $a_r = \{1, 2, 3, \frac{1}{2}, \frac{1}{3}\}$ 五种不同的宽高比。宽高的计算公式如下:

$$w_k^a = s_k \sqrt{a_r} \tag{2}$$

 $h_k^a = s_k \sqrt{a_r} \tag{3}$ 

式(2)、式(3)中 $w_k^a$ 、 $h_k^a$ 分别表示候选框的宽、高。若宽高比为1 时,候选框另增加一个 $s_k^i = \sqrt{s_k s_{k+1}}$ 。用于检测的特征图上默认 有六个候选框,但实际上Conv4\_3、Conv10\_2和Conv11\_2层 仅设置了四个候选框,它们均舍弃了3:1、1:3的候选框。SSD 各预测层的候选框大小详见表1。从统计数据可看出,高层特 征图的尺寸越来越小,候选框的大小逐渐增大。所以底层特 征图主要用来检测小目标,高层特征图被用来检测大目标。

表1	SSD特征层的候选材	E
----	------------	---

Tab.1 Default boxes for SSD feature layer								
预测层	k	w×h	1:1	1:1(大)	1:2	2:1	1:3	3:1
Conv4_3	4	38 × 38	30 × 30	$42 \times 42$	$21 \times 42$	42 × 21	-	-
Fc7	6	19 × 19	60 × 60	$82 \times 82$	$42 \times 84$	$84 \times 42$	35 × 105	105 × 35
Conv8_2	6	10  imes 10	111 × 111	136 × 136	79 × 158	158  imes 79	65 × 195	195 × 65
Conv9_2	6	5×5	162  imes 162	190 × 190	116 × 232	232 × 116	95 × 285	285  imes 95
Conv10_2	4	3 × 3	213 × 213	243 × 243	154  imes 308	308  imes 154	-	-
Conv11_2	4	1 X 1	$264 \times 264$	$289 \times 289$	191 × 382	382 × 191	_	_

# 3 基于PeleeNet的改进SSD模型(Improved SSD model based on PeleeNet)

原始的SSD框架主要分为三部分: 主干网络、检测网 络、分类网络。主干网络主要用来提取特征,几种常见的特 征提取网络包括ResNet、GoogleNet、VGGNet、Inception-Net,各种改进主干网络的SSD模型由此产生。以VGGNet 为主干结构的SSD模型,在进行目标检测时的帧率难以满足 ADAS功能的实时应用需求。为实现在移动端的目标检测,检 测精度和检测准确率需要同时兼顾,因此特征提取网络结构 不能太复杂,参数量不能太大。本文在原始SSD框架的基础 上,将特征提取网络VGGNet替换为轻量化的网络PeleeNet, 显著减少了网络参数;同时PeleeNet采用了卷积、批归一化、 激活函数的顺序组合,而不是卷积、激活函数、批归一化的 组合,这样可以将批归一化和卷积进行合并运算,进一步提高 检测速度。依然采用SSD模型的多尺度预测方法,并在此基 础上提出一种多尺度特征融合模块(Feature Fusion Module, FFM)与特征增强模块(Feature Enhance Module, FEM), 以提 高底层特征图的语义信息,进一步提高对小目标的检测率。

PeleeNet-SSD算法的模型结构如图2所示。模型的整体 结构分为两个部分:一部分是在前端的PeleeNet网络,主要用 于提取目标特征,另一部分是位于后端的检测网络,主要作 用是对前面产生的特征层进行目标的类别分类和位置回归, 最后通过非极大值抑制算法输出目标框。



Fig.2 SSD framework based on PeleeNet network structure

4

#### 3.1 特征融合模块与特征增强模块

#### 3.1.1 特征融合模块

本文设计了多尺度特征融合模块(FFM),如图3所示。将 含有不同语义信息的特征图进行融合产生新的特征图,并用 其进行后面的预测阶段。

为了同时拥有底层特征图的语义信息与高层特征图抽象 的语义信息,本文采用了3×3的卷积尺度以扩大感受野,并 在其后面增加1×1尺度的卷积以增强非线性特性。其中第三 个分支(branch3)是下一尺度的特征图进行反卷积操作,使其 和上一尺度拥有相同的宽高和通道数。最后通过Concatenate 操作将二者合并产生融合后的特征图,用其进行目标预测可 以很好地检测出小尺度目标。



图3特征融合模块

Fig.3 Feature fusion module

#### 3.1.2 特征增强模块

SSD算法利用底层特征图进行小目标的检测,然而底层 特征图的语义信息相对较少,导致检测效果一般。本文针对 底层特征图语义信息不强的特点,设计了底层特征增强模 块,如图4所示。通过两组类似的卷积组来加深浅层特征图, 同时在每组后面都有1×1尺度的卷积,用来增大感受野和非线 性特性。同时本结构采用的是S×1和1×S的卷积层而不是直 接采用S×S的卷积层,从而可以减少计算时间成本。



Fig.4 Feature enhance module

#### 3.2 候选框的设计

特征提取的好坏决定了目标检测算法的性能,而有效地 特征提取取决于训练数据。标准的SSD目标检测算法中的训练 数据就是候选框区域。为了保证对不同大小和不同纵横比的 目标进行预测,SSD算法在不同的特征图上设置了不同的候 选框。当候选框与真实框的重叠度大于0.5时, 被认定为正样 本:反之,为负样本。候选框的大小和纵横比应该根据真实 的数据集进行设置,这样可以获得较高的匹配度,同时也可 以减少背景无效区域对于特征提取的影响。而且当候选框与 真实框越相近,坐标框的回归也就越容易,如果差异较大, 就需要复杂的非线性模型进行求解。本文所采集的数据分布 如图5所示,图中的彩色直线表示SSD原始的默认框比例,彩 色正方形表示各预测层。从图5可知,原候选框长宽比与数据 集分布偏差过大,因此需要根据实际采集的车辆数据集设计 合理的候选框尺度和比例。



图5 车辆尺寸数据分布

Fig.5 Vehicle size data distribution

由车辆尺寸分布图可知,大部分数据集的长宽比介于1至 3。因此,可以去掉比例为1/3的默认框,同时在后三个预测 层加入3/2的长宽比默认框。这样可以提高模型的采样率,而 且能降低计算量。改进后的候选框尺寸分布如图6所示。



Fig.6 Improved default box distribution

# 实验分析(Experimental analysis) 4.1 数据集制作及训练过程

本文所采用的数据集,来自中国某城市道路交通的行驶 车辆数据。通过搭建单目相机采集系统,获取各个路段的视 频数据,通过对视频数据的剪切处理最终获得训练图片。使用 LabelImg图像标注工具,对图片中的车辆进行人工标注,同时 生成与之对应的xml文档,如图7所示。本数据集包含10 000 张图片用于训练和验证,2000张图片用于测试,数据样本标 签为汽车(Car)、卡车(Truck)、公交车(Bus)、工程车(Shop truck)。数据集中包含街道、环线、高速等路段的交通状况。 为了进一步增强数据,本文对训练图片进行随机缩放、水平 旋转,随机改变图片的色调、饱和度及镜像处理。



图7 数据标注样本

# Fig.7 Data labeling sample

本实验所采用的深度学习框架为caffe,GPU型号为RTX 2080 TI。本文首先在KITTI数据集上对改进的车辆检测算法 进行20轮的迭代,获取预训练权重参数。并在此基础上对改 进的目标检测算法进行微调训练,训练时采用的优化方法为 随机梯度下降算法(SGD),初始学习率为0.01,学习策略采用 multistep, gamma为0.5。当迭代次数为50000次时更新一次 学习率,其中momentum为0.9,权重衰减为0.0005,训练的 batch size设为32, 共迭代200 000次。训练过程中的loss值如 图8所示。



# 4.2 评价方法

本文采用的主要评价指标是平均准确率均值mAP,其 是目标检测算法中衡量精确度的指标。mAP的计算与准确率 (Precision)、召回率(Recall)相关。准确率表示预测为正的样 本中的正样本所占的比例,召回率表示正样本中被正确预测 所占的比例。P-R曲线全面评价了检测器的精度性能,该指 标为平均准确率(Average Precision, AP)。计算公式如下:

$$precision = \frac{TP}{TP + FP}$$
(4)

$$recall = \frac{TF}{TP + FN}$$
(5)

$$AP = \int_{0}^{c} P(R) dR \approx \frac{1}{m} \sum_{i=1}^{m} P(R_i)$$
(6)

$$mAP = \frac{1}{C} \sum_{i=1}^{c} AP(i) \tag{7}$$

式中,*TP*表示被正确检测的数量,*FP*表示被误检的数量,*FN* 表示漏检的数量,*m*表示样本点数量,*C*表示需要检测的类别 数量。

#### 4.3 实验结果与分析

本文将标准SSD和改进后的SSD\_our网络分别在所采集的数据集上进行训练与测试,模型的检测性能详见表2。改进后检测模型的mAP为79.83%,且模型大小仅为原始模型大小的1/5。

表2 模型的mAP值							
Tab.2 The	mAP	value	of	the mod	lel -		

齿刑	mAP	单类AP				
侯堂		Car	Bus	Truck	Shop_truck	
SSD	77.58%	77.89%	77.7%	78.45%	76.30%	
SSD_our	79.83%	80.45%	80.94%	79.41%	78.54%	

实际道路视频测试效果如图9所示。改进后的SSD对高速 行驶的小型汽车检测的准确率更高,这是因为特征提取网络 拥有更少的网络参数,提高了计算速度。同时在分析数据尺 度的基础上,在预测层上设置了合理的长宽比的默认框,进 一步降低了噪声的影响以及提升了特征提取的效果。



图9两种场景车辆目标检测算法测试效果

Fig.9 Two scenarios of vehicle target detection algorithm test results

#### 5 结论(Conclusion)

本文针对车辆目标检测实时性低、准确率低、鲁棒性 差等问题,在SSD目标检测算法的基础上,采用特征学习能 力更强的peleeNet网络进行主干网络的替换。由于轻量化的 peleeNet网络结构设计,使得模型参数更少,满足实时性要 求。并且,重新设计了SSD候选框,使其更加符合真实的目 标框,提高了训练过程中正样本的采样率。实验结果表明, 本文改进后的目标检测算法的准确率达到了79.83%,与标准 SSD相比提高了2.25%。本文的方法在检测速度和检测精度方 面有所提高,后续研究工作将在丰富的数据集上进行实验, 进一步提高模型的泛化能力。

## 参考文献(References)

- Reynolds C W. Flocks, Herds, and Schools: A Distributed Behavioral Model[J]. ACM SIGGRAPH Computer Graphics, 1987, 21(4):25–34.
- [2] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 2005:886-893.
- [3] Viola P, Jones M. Robust Real-time Object Detection[J]. International Journal of Computer Vision, 2001(57) :137–154.
- [4] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2):121–167.
- [5] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), 2014:580–587.
- [6] Girshick R. Fast R-CNN[C]. The IEEE International Conference on Computer Vision (ICCV 2015), 2015:1440– 1448.
- [7] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time object Detection with Region Proposal Networks[C]. 29th Neural Information Processing Systems (NIPS 2015), 2015:91-99.
- [8] Dai J, Li Y, He K, et al. R-FCN: Object detection via regionbased fully convolution networks[C]. 30th Conference on Neural Information Processing Systems (NIPS 2016), 2016:379-387.
- [9] 王宇宁,庞智恒,袁德明.基于YOLO算法的车辆实时检测[J]. 武汉理工大学学报,2016,38(010):41-46.
- [10] 吴水清,王宇,师岩.基于SSD的车辆目标检测[J].计算机与现 代化,2019(05):39-44.
- [11] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[C]. The IEEE International Conference on Computer Vision (ICCV 2017), 2017:2980–2988.
- [12] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[C]. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), 2017:4700-4708.