基于自适应多尺度特征融合网络的车辆检测方法

申铉京[1],李涵宇[2],黄永平[1],王 玉[1]
(1. 吉林大学计算机科学与技术学院,吉林长春 130012;2. 吉林大学软件学院,吉林长春 130012)

摘 要: 为了提高车辆检测精度,解决小目标车辆难以检测的问题,本文提出了自适应多尺度特征融合网络(Adaptive Multi-scale Feature Fusion Network,AMFFN),并基于该网络对YOLO v4进行了改进,取得了更好的检测效果. 该网络通过使用多个空间金字塔池化,提高特征的表示能力. 提出的AMFFN跨层融合了多尺度的特征,并为不同尺度的特征层分配可学习的权重. 为了更好地获得特征的细节信息,本文选择了DY-ReLU作为激活函数,它可以随输入动态变化. AMFFN可以被视为一个可重用的模块,通过反复融合特性来获得更精细的特性. 为了避免复杂的网络结构导致的巨大参数量,使用深度可分离卷积替换普通卷积,以降低参数量,提高网络检测速度. 实验结果表明,本文提出的方法相比YOLO v4提高了1.90%的AP,检测速度提高了5 FPS.

关键词: 车辆检测;多尺度融合;深度可分离卷积;YOLO v4
基金项目: 吉林省科技发展计划项目(No.20180201064SF)
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(XXXX)XX-0001-09
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20220281

A Vehicle Detection Method Based on Adaptive Multi-Scale Feature Fusion Network

SHEN Xuan-jing[1], LI Han-yu[2], HUANG Yong-ping[1], WANG Yu[1]
(1. *College of Computer Science and Technology, Jilin University, Changchun, Jilin* 130012, *China*;
2. *College of software, Jilin University, Changchun, Jilin* 130012, *China*)

Abstract: In order to improve the vehicle detection accuracy and solve the problem that small vehicles are difficult to detect, an adaptive multi-scale feature fusion network (AMFFN) is proposed and better performance is achieved after applying it to YOLO v4. To improve the representation capability of features, spatial pyramid pooling modules are employed on each feature map. The proposed AMFFN fuses features of multiple scales across layers and assigns learnable weights to layers of different scales. In order to achieve detailed information better, we select DY-ReLU as activation function, which can change dynamically with input. AMFFN can be treated as a reusable module to obtain more refined features by repeatedly fusing features. To avoid the huge amount of parameters caused by the complex network, depthwise separable convolution is used to replace the normal convolution in order to reduce the amount of parameters and increase the speed of detection. Experimental results show that the proposed method improves the AP by 1.90% and the detection speed by 5 FPS.

Key words: vehicle detection; multi-scale feature fusion; depthwise separable convolution; YOLO v4
Foundation Item(s): Jilin Provincial Science and Technology Development Plan Project (No.20180201064SF)

1 引言

车辆检测是一项在捕获的图像或视频中框定出车辆位置的具有挑战性的任务. 随着经济的快速发展,车辆在人们的日常生活中已经越来越普及[1,2],车辆检测在停车场管理、自动驾驶[3]以及城市智能交通建设[4,5]等领域上发挥着重要作用,因此车辆检测也变得更加有意义. 许多科研人员一直在寻找一种高效、鲁棒性强的检测方法.

车辆检测由于相机拍摄角度、环境等因素,面临许多挑战. 例如,(1)尺度多样性,由于拍摄距离的不同,车辆尺度具有较大差异;(2)角度多样性,不同视角的车辆会呈现不同的外观;(3)小目标问题,一些距离拍摄位置较远的车辆包含较少像素点,信息不足难以检测;(4)背景复杂问题,由于拍摄角度的不同,车辆图片

收稿日期:2022-03-18;修回日期:2022-09-08;责任编辑:朱梅玉

经常会出现遮挡,使得车辆特征难以检测. 因此,车辆检测是一个具有挑战性的任务.

车辆检测方法主要基于目标检测,目标检测方法分为两种:基于人工特征的检测方法和基于深度学习的检测方法. 基于人工特征的检测方法使用人为设计的特征,例如尺度不变特征转换(Scale-Invariant Feature Transform,SIFT)[6],梯度方向直方图(Histogram of Oriented Gradient,HOG)[7],haar-like 特征[8],通过滑动窗口在待测图像上提取候选区域,然后提取候选区域的特征传入支持向量机[9]、Adaboost 分类器[10]等分类器进行训练与分类. 该类方法的缺点是,特征是人工设计,鲁棒性差,不能适应复杂的环境. 随着计算机性能的提高,基于深度学习的检测方法逐渐成为目标检测领域的主流. 基于深度学习的检测方法主要包含 R-CNN[11], Fast R-CNN[12], Faster R-CNN[13], YOLO 系列[14~18],SSD[19]等,这些方法在目标检测领域取得了较好的效果. YOLO v4 是 YOLO 第四代算法,在检测速度和精度之间取得了较好的均衡. 尽管 YOLO 系列算法已经发展到了第五代,但是 YOLO v5 往往被认为集成了过多优化策略,以至于无法区分影响网络效果的到底是网络结构本身还是各种优化策略. 所以,本文将 YOLO v4 作为主体框架,结合本文提出的自适应多尺度特征融合(Adaptive Multi-scale Feature Fusion Network,AMFFN),以探索网络结构本身对检测结果带来的影响.

本文的主要目的是提出一种鲁棒、高效的车辆检测方法. 现有的目标检测方法往往用于检测所有物体,缺乏对于车辆的针对性,检测精度较低. 尤其对于小目标,提取的特征不够丰富,检测效果难以令人满意,且检测速度慢不能满足车辆检测实时性的要求.

本文的主要贡献有以下几点.

(1)为了增强特征的表示能力,将主干网络输出的每一层特征传入空间金字塔池化(Spatial Pyramid Pooling,SPP),并且设计了一种池化核初始化方法以更好地适应不同特征层,扩大了感受野,提取了重要的上下文信息.

(2)提出了 AMFFN 结构并替换 YOLO v4 的特征融合模块. AMFFN 包含自下而上和自上而下的融合路径,跨层融合多个尺度的特征并且为不同尺度特征分配可学习的权重,以自适应地关注对融合贡献大的特征层. 在特征融合时使用 DY-ReLU[20]作为激活函数,增强特征表示能力. 使用深度可分离卷积替换普通卷积,在维持精度的同时降低模型参数量,提升检测速度. AMFFN 是一个可复用的模块,通过复用该模块融合得到更精细的特征,提升检测精度.

(3)本文在 COCO 数据集的车辆类别上进行实验,所提出的方法达到了 54.53%AP,检测精度高于现有的方法. 尤其与 YOLO 系列算法相比,本文方法的各项评价指标都高于 YOLO v4 与 YOLO v5,在小目标检测精度 APS 上提升最为明显,并且具有更快的检测速度.

2 相关工作

2.1 基于深度学习的车辆检测方法

近年来,由于深度学习的快速发展,基于深度学习的车辆检测方法已经成为主流. 基于深度学习的车辆检测方法主要分为两个分支:两阶段检测方法与一阶段检测方法.

两阶段检测方法以 R-CNN, Fast R-CNN, Faster R-CNN 等为主导,该类方法首先根据区域建议网络生成候选框,然后使用深度学习网络对这些区域进行特征提取和分类判别. Zhou 等[21]首先使用一个全卷积网络提取车辆的候选框,然后使用一个深度属性学习网络验证每个候选检测对象. Yuan 等[22]提出了一种基于图的车辆建议定位和检测算法,用于估计边界框中车辆的可能性,解决了不同尺寸和形状的车辆定位不准确的问题. 两阶段检测方法具有较高的准确率,但是检测速度较慢,不能满足车辆检测实时性的要求.

一阶段检测方法作为端到端的方法,将目标的回归和预测作为一个联合的任务,在检测速度上拥有较好的表现. Chen 等[23]基于 SSD 提出了 Inception-SSD 结构,在 SSD 预测网络前额外增加了 Inception 模块,提升对于小目标的检测. Zhang 等[24]提出了一个改进的 YOLO v3 模型,拥有更深的特征提取网络并且使用四个不同尺度特征层进行检测,以更好的解决无人机航拍车辆图片的多尺度以及小目标检测问题. Chen 等[25]结合了 DenseNet, YOLO 与 MobileNet 的优点,提出了轻量级网络 DLNet. 该网络平衡了检测速度与检测精度,能够较好的应用在嵌入式设备上. Rani 等[26]提出了基于 YOLO v3-tiny 的轻量车辆检测网络 LittleYOLO-SPP,通过改进特征提取网络和损失函数以更好的适应车辆检测. 与两阶段方法相比,一阶段方法更快,但检测精度也相对较低.

2.2 YOLO v4

本文将 YOLO v4 作为主体框架,YOLO v4 的主干特征提取网络采用了 CSPDarknet53,基于 YOLO v3 的主干网络 Darknet53 以及 Cross Stage Partial Networks[27]. CSPDarknet53 主要包含五个 CSPResblock,CSPResblock 的结构如图 1 所示,包含多个参差块和一条跨过所有残差块的残差边. 特征经过多个参差块后与残差边融合提取特征. 提取最后三个 CSPResblock 的输出得到三个不同尺度的特征图,将它们作为特征融合网络的输入.

SPP 模块位于 CSPDarknet53 最后一层输出后,结构

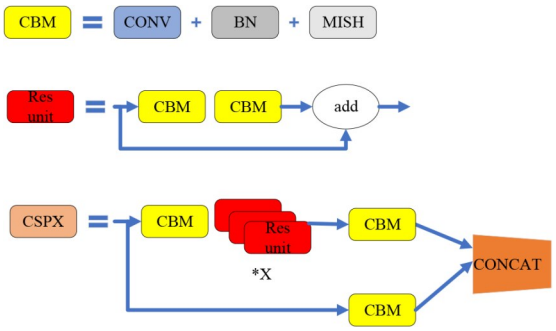


图1 CSPResblock结构图

图如图2所示,采用三个不同大小的池化核,并将池化后的特征图与原特征图进行拼接. 得到的特征图具有更大感受野,具有更强的特征表示能力.

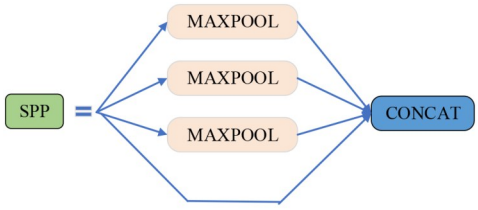


图2 SPP结构图

Path Aggregation Network(PANET)将主干特征提取网络得到三个不同尺度的特征图作为输入进行融合. PANET包含自下而上和自上而下的融合路径. 在自下而上路径中,浅层特征融合高层特征,增强了语义信息;在自上而下路径中,高层特征融合浅层特征丰富了纹理信息. 通过反复融合,每一层特征的上下文信息都得到了增强.

最终PANET输出三个不同尺度特征图传入预测网络进行预测. YOLO v4采用了和YOLO v3一样的预测网络,首先为每层特征分配三个先验框,然后对物体进行分类和回归.

3 本文方法

首先,为了提取特征重要的上下文信息,对SPP进行了改进,并将其应用在了多个特征层上. 然后通过AMFFN自适应跨层融合多个特征层,增强特征的表示能力,以更好的检测小目标. 最后,利用深度可分离卷积提升网络速度.

3.1 SPP的改进

考虑到车辆检测任务中包含大量小目标车辆,低层特征包含丰富的纹理信息,利于小目标的检测,低层特征的信息非常重要. YOLO v4将主干特征提取网络的最后三个输出用于特征融合,本文方法额外将一层低层特征加入融合,提取更丰富的纹理信息. YOLO v4将CSPDarknet53最后一层的输出传入SPP网络中,通过使用多个不同尺寸的池化,有效的增加了特征的感受野,提取了重要的上下文信息. 然而,其他层特征为了融合后通道数不改变,在融合前会降低通道数,这会造成信息的丢失,降低特征的表示能力. 为了增强特征的表示能力,将SPP应用到了每一层特征层上. 图3展示了本文的网络结构,在主干特征提取网络输出的四个特征层后使用了SPP. 因为低层特征感受野和高层特征感受野有较大差距,不适合使用同一尺寸的池化核,低层特征感受野较小,适合使用尺寸较小的池化核,而高层特征感受野较大,适合使用尺寸较大的池化核. 本文设计了一个池化核尺寸初始化的方式:

$$S_j^i = 1 + i \times (j-1) \tag{1}$$

其中,S是池化核大小,j是特征层编号,i是池化核编号. 以特征层P3为例,对应的池化核大小分别为3,5,7. 经过SPP后,每层的特征层拥有更大的感受野,具有更强的特征表示能力.

3.2 AMFFN

车辆检测任务由于其场景的复杂性,车辆具有尺度性多样强、小目标数量多的特点,对于特征的多尺度

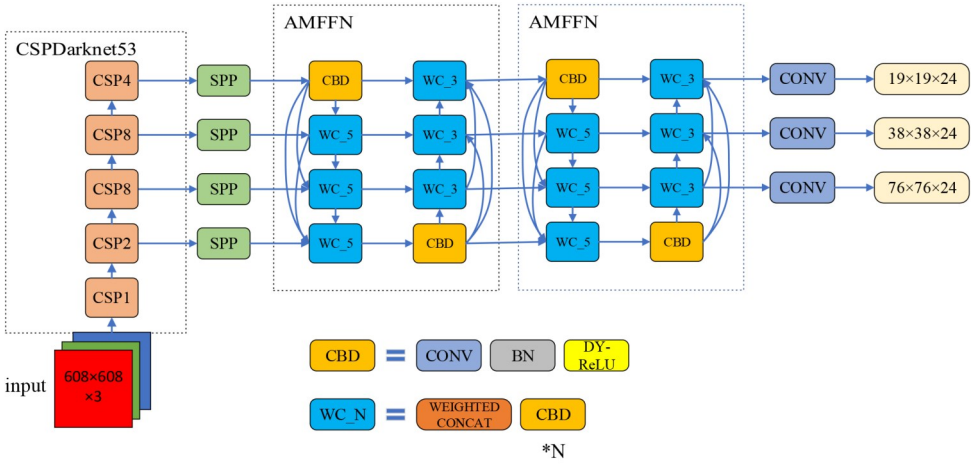


图3 本文网络结构图

信息要求较高. 传统FPN只包含一条单向的特征融合路径,融合得到的特征信息不够丰富. PANET包含自上而下和自下而上的两条特征融合路径,通过反复的特征融合得到更精细的特征. 但是PANET只融合了相邻特征层,忽略了其他特征层对于特征融合的贡献. 本文提出的AMFFN继承了PANET自上而下和自下而上的融合路径,为不同尺度的特征分配可学习的权重,融合多个尺度的特征,丰富了特征的多尺度信息. DY-ReLU是一个动态激活函数,其参数随输入而动态变化,采用DY-ReLU来增强特征表示能力. 除此之外,AMFFN可以作为一个可复用的模块,通过重复融合特征获取更精细的特征. 同时,为了避免由于复杂网络带来的巨大参数量,使用深度可分离卷积替换普通卷积,以降低参数量,提高网络检测速度.

AMFFN将SPP输出的四个特征层作为输入,当输入图像大小为608×608×3时,四个特征层大小分别152×152×128,76×76×256,38×38×512,19×19×1 024. 多层特征融合网络包含自上而下和自下而上的两条特征融合路径. 在自上而下的特征融合路径中,考虑到不同层特征对于融合的贡献是不同的,当前层特征会与上层所有特征加权后进行融合,融合方式为拼接,其中权重是可以学习的参数. 融合结果经过五次卷积后一部分作为自下而上融合路径的输入,另一部分上采样后与下层特征进行融合. 在自下而上的特征融合路径中,当前层特征会与下层所有特征层加权后进行融合,融合结果经过三次卷积后,一部分作为多层特征融合网络的输出,另一部分经过下采样与上层特征层进行融合. 以特征层P3为例:

$$P_3^{\text{td}}=\text{conv}\left(\text{cat}\left(w_1\times P_3^{\text{in}},w_2\times\text{up}\left(P_4^{\text{in}}\right),w_3\times\text{up}\left(P_5^{\text{in}}\right)\right)\right) \tag{2}$$

$$P_3^{\text{out}}=\text{conv}\left(\text{cat}\left(w_1'\times\text{down}\left(P_2^{\text{out}}\right),w_2'\times P_3^{\text{td}}\right)\right) \tag{3}$$

其中,P_3^{in},P_4^{in},P_5^{in}为对应特征层的输入,P_3^{td}是中间特征层,P_2^{out},P_3^{out}为对应特征层的输出,w是可以学习的权重,up是上采样,down是下采样,cat是特征层拼接操作. 其中,权重的计算方式为:

$$w_i=\frac{\text{Relu}\left(w_i\right)}{\sum_j\text{Relu}\left(w_j\right)} \tag{4}$$

其中,w_i首先初始化为1,经过ReLU激活函数以及归一化操作后,作为一个可以学习的参数参与网络训练.

YOLO v4采用了Mish ReLU激活函数,它具有固定的参数,不能灵活适应不同分布的输入. Chen提出了动态线性修正单元(Dynamic Relu,DY-ReLU),它能够依据输入动态调整对应分段函数,与ReLU及其静态变种相比,仅仅需要增加一些可以忽略不计的参数就可以带来大幅的性能提升. 小物体的检测需要特征的细节信息,因此本文在AMFFN中使用DY-ReLU激活函数来增强特征的表示能力. DY-ReLU是一个分段函数,其参数根据$\boldsymbol{x}$计算得到. 输入$\boldsymbol{x}$在进入激活函数前分成两个流分别输入$\theta(\boldsymbol{x})$和$f_{\theta(\boldsymbol{x})}(\boldsymbol{x})$,前者用于获得激活函数的参数,后者用于获得激活函数的输出值. 超函数能够编码输入$\boldsymbol{x}$的各个维度的全局上下文信息来自适应激活函数.

将$f_{\theta(\boldsymbol{x})}(\boldsymbol{x})$定义为

$$f_{\theta(\boldsymbol{x})}(x_c)=\max_{1\le k\le K}\left\{a_c^k(\boldsymbol{x})x_c+b_c^k(\boldsymbol{x})\right\} \tag{5}$$

其中,K为激活函数分段数量,x_c为第c个通道的输入,a_c^k,b_c^k为激活函数的参数,由超函数$\theta(\boldsymbol{x})$生成.

超函数$\theta(\boldsymbol{x})$通过使用类似SE模块的轻量级网络实现. 对于一个输入$C\times H\times W$的输入$\boldsymbol{x}$,空间信息通过全局平均池化压缩,然后经过两个全连接层和标准化层,输出$2KC$个元素,分别对应$\Delta a_{1:C}^{1:K}$,$\Delta b_{1:C}^{1:K}$,经过归一化操作后,最终可以表示为

$$a_c^k(\boldsymbol{x})=\alpha^k+\lambda_a\Delta a_c^k(\boldsymbol{x}) \tag{6}$$

$$b_c^k(\boldsymbol{x})=\beta^k+\lambda_b\Delta b_c^k(\boldsymbol{x}) \tag{7}$$

其中,α^k和β^k是a_c^k和b_c^k的初始值,λ_a和λ_b是控制因子. 对于K=2的情况,默认设置α^1=1,$\alpha^2=\beta^1=\beta^2$=0,λ_a=1,λ_b=0.5.

简而言之,AMFFN首先通过可学习的权重自适应调整不同层特征的重要性,然后使用DY-ReLU激活函数根据不同的输入自适应调整激活函数的参数. 它可以很好地克服融合过程中不同尺度特征的差异带来的影响,充分利用特征的细节信息,这对小目标检测至关重要.

为了融合更精细的特征,将AMFFN视为一个可复用的模块,将第一个AMFFN的输出作为下一个AMFFN的输入. 通过堆叠AMFFN模块,得到的特征图包含更精细的特征,检测精度得到了提高. 随着模块复用次数的提高,模型越来越复杂,往往要以大量的计算开销为代价换取微弱的精度提升. 为了均衡检测精度与速度,本文选择复用AMFFN模块两次.

融合网络最终输出四个特征层,但是本文只将最后三层送入预测网络. 三层特征层已经覆盖了足够多的先验框,能够满足检测的需要,特征层P2只用于特征融合丰富其他特征层的信息. 如果将特征层P2也用于预测,先验框的数量会急剧增加,带来不必要的计算开销.

3.3 深度可分离卷积的应用

车辆检测在实际应用中对于实时性具有一定要求,由于融合多层特征使得网络变得复杂,为了降低模型参数量,本文使用了深度可分离卷积[28]. 将AMFFN中所有卷积替换为深度可分离卷积,深度可分离卷积结构如图4所示. 首先进行逐通道卷积,对于一个$W\times$

$H\times C$的特征图,按通道数划分为C个$W\times H\times 1$的特征图,在每个特征图上进行一个$3\times 3\times 1$卷积. 将得到的C个特征图拼接,得到$W\times H\times C$的特征图. 然后进行逐点卷积,特征图进行N次$1\times 1\times C$卷积,最终得到$W\times H\times N$的特征图. 使用深度可分离卷积后,特征融合网络部分的参数量减少到1/3. 尽管深度可分离卷积会带来精度上略微下降,但在对检测速度和精度之间权衡之后,本文选择使用深度可分离卷积.

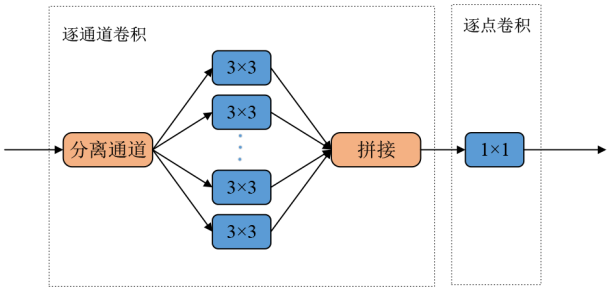


图4 深度可分卷积结构图

4 实验结果与分析

4.1 实验环境与设置

本文在两个数据集上进行了实验,以验证算法的性能,包括COCO 2017数据集和UA-DETRAC[29]数据集. 在四块GeForce RTX3090显卡上进行实验,输入尺寸为,选择adam优化器,为了加速模型训练,加载预训练的CSPDarknet53模型参数作为预训练权重,首先冻结CSPDarknet53部分,批量大小设置为64,初始学习率为0.001,采用余弦退火衰减法,进行50 000次迭代. 然后解冻整个网络,批量大小调整为32,初始学习率调整为0.000 1,进行50 000次迭代.

4.2 实验结果

4.2.1 COCO数据集实验结果

为了验证网络性能,本文首先在COCO数据集上进行实验. COCO数据集包含真实场景下的80种类别图片,常用于目标检测任务. 由于本文的工作针对车辆检测,将所有图片都用于训练不是个好的选择,所以本文只用其中的car,bus,truck类别进行训练与测试. 提取COCO数据集中的车辆图片共16 270张,按比例7:1:2划分训练集,验证集,测试集.

本文将所提出的方法与现有的方法进行了比较,因为这些方法都是针对一般目标检测,为了公平起见,本文复现这些方法并使用相同的实验设置进行训练与测试. 将平均精度(Average Precision,AP)作为评价指标,平均精度是多个召回值的所有类的平均精度的平均值. AP_S为尺寸小于32×32的目标的平均精度,AP_M为尺寸大于32×32小于96×96的目标的平均精度,AP_L为尺寸大于96×96的目标的平均精度. 检测结果见表1,首先本文与两阶段检测方法Faster-RCNN进行了比较,提出的方法在AP指标上提升了6.98%. 在一阶段检测方法中,本文与SSD,Efficientdet,Retinanet,Centernet,YOLO系列进行了对比,提出的方法具有最好的检测效果. 尤其与YOLO v4相比,提高了1.9%AP,在AP_S中提高了2.16%,AP_M提高了1.3%,AP_L提高了0.73%,AP_S提升最大,证明了本文的方法对于小目标的检测效果提升更明显.

为了更全面的评估本文提出的AMFFN的效果,本文还计算了在iou阈值为0.5下的AP,F1得分,精度,召回率和检测速度作为评价指标,与YOLO v4比较. 如表2所示,提出的网络在AP指标上比YOLO v4提高了2.3%,F1得分提高了2.53%,精度提高了1.71%,召回率提高了1.05%. 本文方法各项指标都得到了提高,因为通过SPP模块,每层特征具有更大感受野,上下文信息得到增强. 经过两次AMFFN模块,特征跨层融合并且自适应关注对融合贡献大的特征层,融合得到的特征具有更强的特征表示能力. 模型大小从254M降低到了

表1 本文方法与现有方法在COCO数据集上的实验结果比较

Method	AP/%	AP_{50}/%	AP_{75}/%	APs/%	AP_M/%	AP_L/%
Faster R-CNN[13]	46.85	67.75	49.27	27.28	48.84	63.18
SSD[19]	41.35	61.42	42.35	21.24	45.52	58.80
Efficientdet-d0[30]	43.08	62.22	45.85	22.05	48.35	61.24
Efficientdet-d1[30]	49.06	68.64	52.34	28.85	54.32	66.04
Retinanet[31]	50.85	71.15	54.16	34.15	54.29	61.21
Centernet[32]	51.64	72.45	54.24	32.52	53.14	64.17
YOLOv3[16]	43.05	67.95	44.42	28.31	45.47	51.92
YOLOv4[17]	52.63	75.84	58.01	35.75	56.14	65.25
LittleYOLO-SPP[26]	52.95	—	—	—	—	—
YOLOv5m[18]	53.01	76.24	58.52	36.21	56.52	65.45
YOLOv5l[18]	53.64	77.24	58.69	36.54	56.61	65.70
Ours	54.53	78.14	59.50	37.91	57.44	65.98

198M，检测速度从25FPS提高到了30FPS. 因为AMFFN中使用了深度可分离卷积，相比普通卷积减少了大量参数，速度得以提高. 可见本文提出的网络在精测精度和速度上相比YOLO v4都有一定的提升.

表2 YOLO v4与本文方法的AP_{50}，F1得分，精度，召回率，FPS，模型大小

Method	AP_{50}/%	F1/%	Precision/%	Recall/%	FPS	Model size/M
YOLO v4	75.84	74.59	88.24	63.77	25	245
Ours	78.14	77.12	89.95	64.82	30	198

图5(*a*)是YOLO v4在coco数据集上的检测效果图，虽然能够检测出部分小目标车辆，但仍存在漏检情况. 图5(*b*)是本文方法的检测效果图，对于一些遮挡较严重的小目标车辆仍然有较好的效果，说明本文提出的方法对小目标检测有较好的检测效果.

4.2.2 UA-DETRAC数据集实验结果

为了测试网络在真实道路场景下的检测效果，本文额外在UA-DETRAC数据集上进行了实验. UA-DETRAC数据集由Cannon EOS 550D摄像头在24个不同地点拍摄的10个小时的视频组成，视频以每秒25帧的速度录制，训练集包含82 085张图片，测试集包含56 167张图片. 本文将YOLO v4、YOLO v5与本文提出的方法在该数据集上训练与测试，实验结果见表3. 本文提出的方法在AP上比YOLO v4提高了3.02%，比YOLO v5m提高了1.72%，比YOLO v5l提高了1.26%，能够更好的适应真实道路场景下的车辆检测. AP_S相比AP_M与AP_L提升最大，证明了本文的方法对于小目标的检测效果较好. 这由于本文的方法通过多个SPP丰富了特征的上下文信息，并通过重复的AMFFN模块融合得到更细粒的特征，特征具有更丰富的信息.

图6(*a*)和(*b*)分别为YOLO v4和本文方法在UA-DETRAC数据集上的检测效果. YOLO v4对于远处的小目标车辆存在漏检，而本文方法能够正确的检测出所有车辆，说明本文提出的自适应多尺度特征融合能够增强特征的表示能力，对于小目标车辆的检测具有更强的鲁棒性[33].

表3 在UA-DETRAC数据集上的实验结果

Method	AP/%	AP_{50}/%	AP_{75}/%	APs/%	AP_M/%	AP_L/%
YOLO v4	56.85	81.24	62.12	38.55	60.10	66.90
YOLOv5m	58.15	82.26	62.52	39.15	60.45	67.68
YOLOv5l	58.61	83.15	62.99	41.12	61.15	68.40
Ours	59.87	83.84	64.24	42.85	61.82	68.87

(*a*) YOLO v4在COCO数据集检测结果 (*b*) 本文方法在COCO数据集检测结果

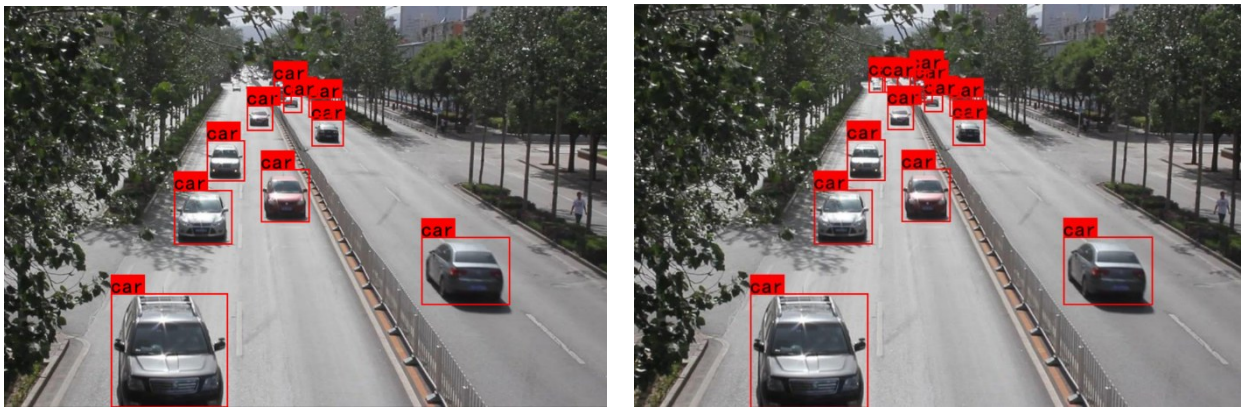
图5 COCO数据集检测结果

5 消融实验

在本节，本文进行了消融实验来验证所提出的方法的有效性.

5.1 各模块重要性的消融实验

为了验证所提出的方法的各模块的重要性，SPP，AMFFN，深度可分离卷积逐步的应用于网络中来验证性能. 消融实验的基准是YOLO v4，实验结果见表4. SPP将YOLO v4的AP提高了0.32%，这得益于SPP使用了不同大小的池化核，获取了不同尺度的特征，增大了特征的感受野，丰富了特征的上下文信息. 由于SPP的计算量较小，FPS仅下降了1. 经过SPP，特征得到了增强，具有更丰富的语义信息. AMFFN提高了1.42%的AP，这表明了AMFFN通过多层特征的融合并自适应学习权重，能够更关注对特征融合贡献大的特征层，并通过DY-ReLU增强特征的表示能力. 将SPP和AMFFN模块同时使用，FPS下降了7，但是检测精度得到了较大提升，AP提升了1.98%. 最后，使用深度可分离卷积替换AMFFN中的普通卷积，以0.08%AP的细微损失换来了FPS上12的提升. 最终的网络在检



(*a*) YOLO v4在UA-DETRAC数据集检测结果 (*b*) 本文方法在UA-DETRAC数据集检测结果

图6 UA-DETRAC数据集检测结果

表4 各模块的消融实验

SPP	AMFFN	DP CONV	FPS	AP%
			25	52.63
√			24	52.95
	√		19	54.05
√	√		18	54.61
√	√	√	30	54.53

测速度和精度上相比YOLO v4都得到了提升.

5.2 SPP的消融实验

SPP可以增强特征上下文信息,为了验证将SPP应用在不同特征层的效果本文进行了消融实验,实验结果见表5. 首先本文将SPP用在了单层特征,AP在SPP用在P2到P5特征层上分别提高了0.14%, 0.2%, 0.31% and 0.37%. 这表明SPP对于特征的上下文信息增强有一定的作用,特征层数越深效果越明显. 最后本文将SPP用在所有四个特征层上,AP达到了54.53%,比不使用SPP提高了0.55%.

表5 SPP的消融实验

SPP	Level	AP%
No SPP		53.98
Single level	P2	54.12
Single level	P3	54.18
Single level	P4	54.29
Single level	P5	54.35
All level	ALL	54.53

5.3 AMFFN的消融实验

为了验证所提出的策略的有效性,本节进行了消融研究. 首先,验证可学习权重和DY-ReLU对AMFFN性能的提升. 未使用可学习权重和DY-ReLU的AMFFN是基线,可学习权重和DY-ReLU逐步应用在网络中. 如表6所示,可学习权重和DY-ReLU使AP分别提升了0.56%和1.07%. 同时使用可学习权重和DY-ReLU的AMFFN将AP从52.89%提高到54.53%,说明了提出的方法的有效性.

表6 AMFFN的消融实验

Learnable weights	DY-ReLU	AP/%
		52.89
√		53.45
	√	53.96
√	√	54.53

然后,进行了消融实验来验证AMFFN最佳重复次数,因为重复AMFFN可以得到更精细的特征但是过多的参数量会造成模型的臃肿. 本文分别将AMFFN重复使用了1,2,3次,实验结果见表7,AP分别达到了53.65%,54.53%,54.65%,证明了重复使用AMFFN可以提高特征的表示能力. 然而,过多的使用AMFFN会带来参数量的增加,降低模型的速度,使用三次AMFFN仅提升了0.12%AP,FPS却下降了8. 在衡量了检测速度和精度之后,本文选择重复使用两次AMFFN模块.

表7 AMFFN重复次数的消融实验

AMFFN	AP/%	Model size/M	FPS
1	53.65	154	40
2	54.53	195	30
3	54.65	236	22

6 结论

本文提出了一个自适应多尺度特征融合网络,并将其应用在了YOLO v4网络中. 首先通过在每一层特征上使用SPP,扩大了感受野,增强了特征的上下文信息. 然后通过使用AMFFN,将浅层特征的纹理信息与深层特征的语义信息充分融合,增强了特征的表示能力. 实验证明,提出的方法在车辆检测任务中具有较好的检测效果,提高了车辆的检测精度与检测速度,并且对于小目标车辆具有一定鲁棒性,在检测精度与精测

速度之间做到了很好的平衡.

参考文献

[1] PIAO J, MCDONALD M. Advanced driver assistance systems from autonomous to cooperative approach[J]. Transport Reviews, 2008, 28(5): 659-684.

[2] ANURADHA R. RETRACTED ARTICLE: Feature selection and classification methods for vehicle tracking and detection[J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(3): 4269-4279.

[3] LI P L, CHEN X Z, SHEN S J. Stereo R-CNN based 3D object detection for autonomous driving[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2020: 7636-7644.

[4] WANG L, LU Y, WANG H, et al. Evolving boxes for fast vehicle detection[C]//2017 IEEE International Conference on Multimedia and Expo (ICME). Hong Kong: IEEE, 2017: 1135-1140.

[5] YAN Z Y, YUAN Y C, ZUO W M, et al. Perspective-guided convolution networks for crowd counting[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2020: 952-961.

[6] SRIVASTAVA S, . Video-based real-time surveillance of vehicles[J]. Journal of Electronic Imaging, 2013, 22(4): 041103.

[7] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005: 886-893.

[8] MITA T, KANEKO T, HORI O. Joint haar-like features for face detection[C]//Tenth IEEE International Conference on Computer Vision (ICCV´05). Beijing: IEEE, 2005: 1619-1626.

[9] SÁNCHEZ A V D. Advanced support vector machines and kernel methods[J]. Neurocomputing, 2003, 55(1/2): 5-20.

[10] BAEK Y M, KIM W Y. Forward vehicle detection using cluster-based AdaBoost[J]. Optical Engineering, 2014, 53 (10): 102103.

[11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.

[12] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2016: 1440-1448.

[13] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 779-788.

[15] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6517-6525.

[16] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08) [2022-03]. https://arxiv.org/abs/1804.02767.

[17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2023-03-13]. https://arxiv. org/abs/2004.10934.

[18] GLENN J. YOLOV5. (2021) [2022-03]. https://github.com/ultralytics/yolov5.

[19] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision - ECCV 2016. Amsterdam: Springer, 2016: 21-37.

[20] CHEN Y P, DAI X Y, LIU M C, et al. Dynamic ReLU [C]//European Conference on Computer Vision - ECCV 2020. Glasgow: Springer, 2020: 351-367.

[21] ZHOU Y, LIU L, SHAO L, et al. Fast automatic vehicle annotation for urban traffic surveillance[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(6): 1973-1984.

[22] YUAN X, SU S, CHEN H J. A graph-based vehicle proposal location and detection algorithm[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(12): 3282-3289.

[23] CHEN W P, QIAO Y T, LI Y J. Inception-SSD: An improved single shot detector for vehicle detection[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(11): 5047-5053.

[24] ZHANG S L, CHAI L, JIN L Z. Vehicle detection in UAV aerial images based on improved YOLOv3[C]// 2020 IEEE International Conference on Networking, Sensing and Control (ICNSC). Nanjing: IEEE, 2020: 1-6.

[25] CHEN L, DING Q W, ZOU Q, et al. DenseLightNet: A

light-weight vehicle detection network for autonomous driving[J]. IEEE Transactions on Industrial Electronics, 2020, 67(12): 10600-10609.

[26] JAMIYA S S, RANI E P. LittleYOLO-SPP: A delicate real-time vehicle detection algorithm[J]. Optik, 2021, 225: 165818.

[27] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle: IEEE, 2020: 1571-1580.

[28] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2022-03]. https://arxiv.org/abs/1704.04861.

[29] LONGYIN, WEN, . UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking[J]. Computer Vision and Image Understanding, 2020, 193: 102907.

[30] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 10778-10787.

[31] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2999-3007.

[32] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-16)[2022-03]. https://arxiv.org/abs/1904.07850.

[33] 王玉, 李涵宇, 申铉京, 等. 一种基于多层特征融合的车辆检测方法: CN113420706B[P]. 2022-05-24.
WANG Y, LI H Y, SHEN X J, et al. Vehicle detection method based on multilayer feature fusion: CN113420706B[P]. 2022-05-24. (in Chinese).

作者简介

申铉京 男,1958年出生,吉林和龙人. 吉林大学计算机科学与技术学院教授. 主要研究方向为多媒体技术,计算机图像处理,智能测量系统,光电混合系统.
E-mail: xjshen@jlu.edu.cn

李涵宇 男,1997年出生,江苏扬州人. 吉林大学软件学院硕士研究生. 主要研究方向为图像处理与模式识别.
E-mail: hanyul19@mails.jlu.edu.cn

黄永平 男,1964年出生,吉林白城人. 吉林大学计算机科学与技术学院副教授. 主要研究方向为嵌入式软件与信息物理融合系统.
E-mail: hyp@jlu.edu.cn

王 玉(通讯作者) 男,1983年出生,黑龙江双鸭山人. 吉林大学计算机科学与技术学院副教授. 主要研究方向包括多媒体技术、机器学习与智能系统.
E-mail: wangyu001@jlu.edu.cn