DOI:10.19652/j.cnki.femt.2204450

融合特征金字塔和通道注意力的轻量车辆检测算法*

张 奇 陈梦蝶 赵 杰
(西安工程大学电子信息学院 西安 710048)

摘 要:车辆检测是智能交通、无人驾驶等系统得以实现的重要支撑性技术。低精度或低速度的车辆检测器应用受限,因此提出了一种快速准确的车辆检测器。首先,前端特征提取网络 VGG16 由 MobileNetV3_Large 替代,减少了参数量和计算量,并增加了对高维特征的提取能力;其次,利用特征金字塔思想构建双向加权融合网络,有效融合不同尺度的特征,获取多维度的车辆特征;最后在特征提取层引入高效通道注意力,重新标定不同特征通道的重要性,进一步提高模型性能。与 SSD 相比,所提出的模型在 KITTI 数据集和 BDD 100 K 数据集上分别将平均精度提高了 7.50%和 3.50%,并具有实时检测能力(超过 40 fps),在检测精度和速度方面有更好的平衡,说明了方法的有效性。

关键词:车辆检测;SSD;MobileNetV3;特征金字塔;注意力机制
中图分类号: TP391 **文献标识码**:A **国家标准学科分类代码**: 510.4

Lightweight vehicle detection network fusing feature pyramid and channel attention

Zhang Qi Chen Mengdie Zhao Jie
(School of Electronics and Information, Xi′an Polytechnic University, Xi′an 710048, China)

Abstract: Vehicle detection is an important supporting technology for the realization of intelligent transportation, autonomous driving, etc. Poor accuracy or low inference vehicle detectors are limited in application, therefore this paper proposes a fast and accurate vehicle detector. First, the front-end feature extraction network VGG16 is replaced by MobileNetV3_Large, which reduces the number of parameters and computation, and increases the ability to extract high-dimensional features. Next, the feature pyramid idea is used to construct a weighted bi-directional fusion network to obtain multi-dimensional vehicle features; In the end, introducing efficient channel attention in the feature extraction layer to re-calibrate the importance of different feature channels and further improve the model performance. Compared with SSD, our proposed model improves mAP by 7.50% and 3.50% on KITTI dataset and BDD 100 K dataset, and with real-time inference (more than 40 fps), it reports a better trade-off in terms of detection accuracy and speed, illustrating the effectiveness of our method.

Keywords: vehicle detection; SSD; MobileNetV3; feature pyramid; attention mechanism

0 引 言

车辆检测是识别、检测和定位车辆目标,在智能交通、自动驾驶等领域得到广泛的应用。由于车辆图像大多是在实际道路场景中拍摄的,目标车辆类型多样、背景复杂、光照条件多变等问题使得如何获取更具区分性的车辆特征成为一个难点和热点问题。近年来,利用卷积神经网络(convolutional neural networks,CNN)自主地提取车辆深度特征成为解决该问题的主流方法,相较于依靠人工设计特征的传统方法,该方法操作简单易于实现,鲁棒性更好,为车辆检测提供了新的思路。

基于 CNN 的车辆检测模型分为两类,双阶段和单阶段检测算法。双阶段检测算法以 R-CNN[1]、Fast R-CNN[2]、Faster R-CNN[3]以及 Mask R-CNN[4]等为代表,该类算法基于区域候选框,具有良好的检测精度,但速度较慢。单阶段检测算法以 SSD[5]、RefineDet[6]、

收稿日期:2022-10-28
* **基金项目**:西安市碑林区应用技术研发项目(GX2007)资助

YOLO[7]等为代表，此类算法取消了区域候选框，基于分类回归思想端到端地使用卷积神经网络提取图像特征信息，进而定位目标位置并识别目标类别，在检测速度上具有很大优势，能更好地适用于实时交通场景。然而区域建议步骤的取消导致该方法定位和分类结果变差，检测精度不及双阶段检测器。寻找一种快速准确的车辆检测器成为研究难点和热点。本文基于检测速度较快的单阶段方法，提高车辆检测器的准确性和获取更快的检测速度。

为了将车辆检测应用于嵌入式平台或低延迟需求场景，满足自动驾驶领域 40 fps 以上的实时检测速度。刘寒迪等[8]采用轻量级 MobileNet 网络替换 VGG-16 网络，构成 MobileNet-SSD 模型，提高了检测速度。Sim 等[9]使用 ShuffleNet 模块替换跨级局部密集连接网络(cross stage partial dense convolutional network，CSP DenseNet)来压缩网络模型，适用于实时的车辆检测，但车辆检测准确率也略微降低。上述几种模型通过不同的方法提升检测速度，但在复杂的应用场景中，并不能保证其检测精度一定满足。

交通图像和视频中通常包含具有大尺度变化的车辆目标，为了解决车辆检测器不适应不同大小车辆的问题。石欣等[10]额外引入浅层特征改进特征金字塔，提取小目标细粒度特征，提高小目标检测准确率。Hu 等[11]结合上下文感知池化(context-aware RoI pooling，CRP)和多分支决策网络提出了一种对尺度不敏感的卷积神经网络(scale-insensitive convolutional neural network，SINet)，有效检测尺度变化较大的车辆。虽然上述模型证明采用多尺度的特征图用于检测是至关重要的，但没有很好地利用局部细节特征和全局语义特征，导致对小目标的检测效果仍然不好。近年来，采用特征融合或者加入注意力机制是提升模型性能的重要方法。

在车辆检测场景中，通过充分挖掘车辆特征信息，可以强化对车辆的检测能力。刘鸣瑄等[12]将高分辨率的浅层特征和语义较强的深层特征进行融合，通过残差块设计了一个完整的特征融合结构，准确率有明显提高。Zhao 等[13]继承了 SSD 的体系结构，并引入一种特征融合模块从原始特征中学习到更好的特征，在增加很少推理时间的条件下提高了检测精度。梁继然等[14]通过轻量化通道注意力机制加强对有效通道特征的表达，并将其作为深层特征提取层，在复杂交通场景下仍具有良好的检测性能。上述两种方法对特征提取仍不够丰富，在复杂的应用场景中，并不能保证其特征提取性能一定满足检测精度需求。

综上所述，本文提出了一种融合特征金字塔和通道注意力的轻量级车辆检测器(MCE-SSD)，采用 MobileNetV3_Large[15]作为 SSD 的前端特征提取网络，并利用反残差结构改进网络最后五层的标准卷积，既减少了模型参数量，提高了检测速度，又有效地提升了模型的识别准确率；然后在特征提取模块中利用双向加权特征融合网络(weighted bi-directional feature pyramid network，BiFPN)[16]的思想，有效融合不同尺度和同一尺度的特征，更好地形成车辆多维度特征；最后引入高效通道注意力机制(efficient channel attention network，ECA-Net)[17]，增强车辆信息等关键特征的表示，抑制背景及非车辆等无关特征的表示，获取更具判别性的车辆特征，进一步提升模型的识别能力。

1 SSD 检测算法

SSD 目标检测算法是一种端到端的算法，相对于双阶段方法有着更快的检测速率，在 VGG-16 网络的基础上使用两个卷积层替换掉全连接层 fc6 层和 fc7 层，并添加了一系列卷积层获取不同尺度的特征图(尺寸分别是 38×38、19×19、10×10、5×5、3×3、1×1)直接回归目标车辆的位置和类别，比较大的特征图来用来检测相对较小的目标，而小的特征图负责检测大目标，实现多尺度特征图预测。同时借鉴了锚(anchor)的理念，在每个单元设置长宽比不同的先验框，预测的边界框以这些先验框为基准，在一定程度上降低了训练难度。最后通过非极大值抑制算法(non-maximum suppression，NMS)去除多余的候选框，得到最终的检测结果。根据 SSD 算法思想进行合理变换可以提升目标物体的识别定位准确率，同时保持实时性。

2 MCE-SSD 算法

本文所提出的 MCE-SSD 在保持网络轻量化的前提下，引入金字塔网络和注意力机制获取更具判别性的车辆特征，提升模型的识别能力。模型结构如图 1 所示，大小为 300×300×3 的图像经 MobileNetV3_Large 后得到车辆特征图(feature map)；选取其中不同尺度的 6 层特征图(Conv14、IRSConv1、IRSConv2、IRSConv3、IRSConv4、IRSConv5)作为跨级双向加权特征融合网络(cross weighted bi-directional feature pyramid network，CBiFPN)输入，实现车辆多层特征的加权融合；最后在每个特征提取层后接入注意力机制 ECA-Net，重新标定特征通道权重，赋予含有更多车辆信息的通道较大的权重，进一步提高检测精度。

2.1 基于 MobileNetV3_Large 的特征提取网络

传统卷积的计算参数量大，导致检测速度较慢。为了使车辆检测任务应用于嵌入式平台或低延迟需求场景，本文对 SSD 网络改进如下：1)使用 MobileNetV3_Large 网络(去掉分类层和输出层)代替 VGG16 网络作为特征提取基础网络；2)利用反残差结构代替最后五层网络中的标准卷积，构成 IRSConv1、IRSConv2、IRSConv3、IRSConv4、IRSConv5。改进后的 MobileNetV3_Large-SSD 网络结构如图 1(a)所示。其核心结构 MobileNetV3 block 结构示意图如图 2 所示。

MobileNetV3 block 首先经过 1×1 点卷积扩展通道，然后由 3×3 深度卷积得到的车辆特征图，再通过全局平

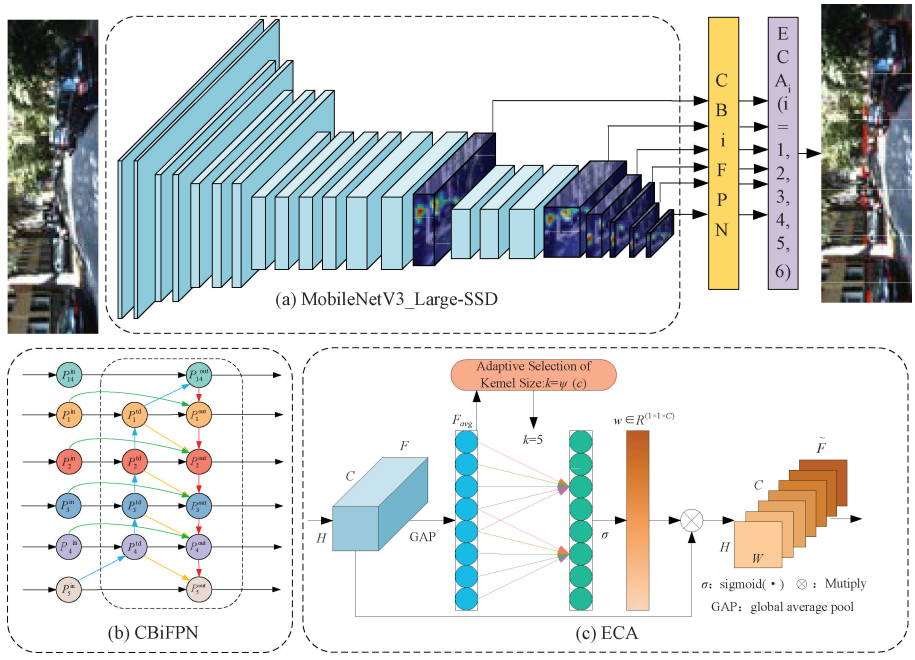

图 1 MCE-SSD 网络结构

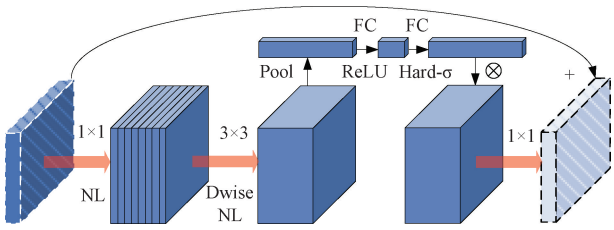

图 2 MobileNetV3 block

均池化整合各个车辆特征图的信息，然后接入挤压和激励网络(squeeze-and-excitation networks，SENet)，通过两个全连接层重新校准输入特征，加强网络的学习能力。最后再经过 1×1 点卷积将其映射到一个低维空间。同时使用分阶段线性函数 ReLU6$(x+3)/6$ 代替计算成本过高的 sigmoid 函数，将 swish 函数改进为 h-swish 函数，该激活函数减少了内存访问次数，显著降低了延迟成本，在深层网络有着更好的表现，其公式为：

$$h-swish(x)=x\frac{ReLU6(x+3)}{6} \tag{1}$$

改进后的 MobileNetV3_Large-SSD 网络参数如表 1 所示。

2.2 双向加权特征融合网络

在实际道路场景中，车辆尺度变化较大。针对 SSD 没有很好地利用局部细节特征和全局语义特征的问题，本文借鉴 BiFPN 算法思想，继承 SSD 体系结构选取 Conv14、IRSConv1、IRSConv2、IRSConv3、IRSConv4、IRSConv5 层的特征图，将浅层特征与深层特征双向加权融合，实现车辆特征的多层复用，提高检测精度。CBiFPN 具体网络结构

表 1 MobileNetV3_large-SSD 网络参数

输入	操作	非线性函数	通道数	s
300×300×3	Conv2d	H-Swish	16	2
150×150×16	IRS，3×3	ReLU	16	1
150×150×16	IRS，3×3	ReLU	24	2
75×75×24	IRS，3×3	ReLU	24	1
75×75×24	SE IRS，5×5	ReLU	40	2
38×38×40	SE IRS，5×5	ReLU	40	1
38×38×40	SE IRS，5×5	ReLU	40	1
38×38×40	IRS，3×3	H-Swish	80	2
19×19×80	IRS，3×3	H-Swish	80	1
19×19×80	IRS，3×3	H-Swish	80	1
19×19×80	IRS，3×3	H-Swish	80	1
19×19×80	SE IRS，5×5	H-Swish	112	1
19×19×112	SE IRS，5×5	H-Swish	112	1
19×19×112	SE IRS，5×5	H-Swish	112	1
19×19×112	SE IRS，5×5	H-Swish	160	2
10×10×160	SE IRS，5×5	H-Swish	160	1
10×10×160	Conv2d，1×1	H-Swish	960	1
10×10×960	IRS，3×3	ReLU6	1 024	1
10×10×1 024	IRS，3×3	ReLU6	512	2
5×5×512	IRS，3×3	ReLU6	256	2
3×3×256	IRS，3×3	ReLU6	256	2
2×2×256	IRS，3×3	ReLU6	128	2
1×1×128		ReLU6		

注：s 表示步长

如图 1(b)所示。一个节点只有一条输入边而没有特征融合，去掉以减少参数(P_5^{td} 和 P_{14}^{td})。自上而下阶段：最深层

特征经过二倍上采样与上一层融合，融合后的 feature map 重复迭代该过程，直至最浅层特征 Conv14。自下而上阶段：处理后的浅层特征经过下采样再向深层特征融合，更多的关注非相邻分辨率的特征。其次为了减少因网络层级过多造成的信息丢失，在同一尺度特征间添加一条额外的边。同时本文在中间层新增跨级数据流，更好的融合不同层级之间的数据。最后通过可学习的权重给不同的尺度特征分配信息比重，特征的加权融合通过快速归一化融合方法实现，如式(2)所示。

$$O=\sum_i \frac{w_i}{\varepsilon+\sum_j w_j}\cdot I_i \tag{2}$$

式中：$w_i \geqslant 0$；ε 取值为 1×10^{-4}，是为了避免数值不稳定而设置的极小值。

以图 1(b)的 IRSConv3 层为例，该特征层融合过程如下：

$$P_3^{td}=Conv\left(\frac{w_1\cdot P_3^{in}+w_2\cdot Resize(P_4^{td})}{w_1+w_2+\varepsilon}\right) \tag{3}$$

$$P_3^{out}=Conv\left(\frac{w'_1\cdot P_3^{in}+w'_2\cdot P_3^{td}+w'_3\cdot Resize(P_2^{out})+w'_4\cdot P_2^{td}}{w'_1+w'_2+w'_3+w'_4+\varepsilon}\right) \tag{4}$$

式中：P_3^{in} 表示 IRSConv3 层的输入特征；P_3^{td} 表示 IRSConv3 层的中间特征；P_3^{out} 表示 IRSConv3 层的输出特征；w_i 表示不同的权重。

2.3 接入 ECA 通道注意力

在车辆检测场景中，背景及非车辆的特征为无效特征，其混杂在网络的通道维度中，对有效物体的特征造成干扰，因此本文在特征提取层后接入高效通道注意力 ECA-Net，从车辆图像深层信息中筛选出有效的特征。ECA-Net 建立了局部跨通道信息融合机制，首先对输入特征图 F 使用全局平均池化，聚合 $\boldsymbol{F}\in\boldsymbol{R}^{W\times H\times C}$ 各通道的空间信息，获得特征 $\boldsymbol{F}_{avg}\in\boldsymbol{R}^{1\times1\times C}$，如式(5)所示。

$$\begin{cases}\boldsymbol{F}_{avg}=GAP(\boldsymbol{F})\\ GAP(\boldsymbol{F})=\dfrac{1}{W\times H}\sum\limits_{i=1,j=1}^{W,H}F_{i,j}\end{cases} \tag{5}$$

式中：$F_{i,j}\in\boldsymbol{R}^{(C)}$ 是 $\boldsymbol{F}$ 在(i,j)位置的全通道特征；$\boldsymbol{F}_{avg}$ 通过卷积核大小为 k 的一维卷积进行卷积计算，对各个通道之间的相关性进行建模。之后使用 Sgimoid 函数计算一维卷积输出的激活值，获得表示特征通道局部关系和重要程度的权重 $\boldsymbol{W}\in\boldsymbol{R}^{1\times1\times C}$。最后，通过逐元素相乘将生成的各个通道权重加入到原输入特征图上，完成对 F 各通道特征的重新编码，赋予关键特征较大的权重，而无关特征则被赋予较小的权重。ECA 结构如图 1(c)所示。

3 实验及结果分析

3.1 实验环境

本实验的环境配置如下：64 位 Ubuntu16.04 操作系统；NVIDIA GTX 1080Ti GPU；深度学习框架 Pytorch1.1.0；编程语言 Python3.6。训练参数的具体设置如下：每批次训练样本为 16 张，采用随机梯度下降法(SGD)进行优化。采用多步衰减学习策略，初始学习率为 10^{-4}，gamma 为 0.1。

3.2 基于 KITTI 数据集测试

KITTI 数据集[18]由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合构建，主要包含城区、乡村和高速公路等场景下采集的真实图像数据，每张图像中最多达 15 辆车，还有各种程度的遮挡与截断。对于车辆检测任务，KITTI 提供了 Car、Van、Tram 以及 Truck 四种不同的车辆类别标签，共含有 7 481 张图片。为了便于 MCE-SSD 算法进行训练，本文将数据集转化为 VOC 数据集格式，并按照 8∶1∶1的比例划分为训练集、测试集和验证集。

本文所提出的方法使用到了 MobileNetV3_Large 网络、CBiFPN 特征金字塔、ECA 通道注意力。为验证上述方法的有效性本文进行了如下实验。

1)MobileNetV3_Large 网络对模型速度的影响

如表 2 所示，对不同基础网络的算法性能，参数量和传输速率在 KITTI 数据集上进行了对比。从表 2 可以看出，本文改进算法 MobileNetV3_Large-SSD 的均值平均精度(mean average precision，mAP)较 VGG16-SSD 网络提高了 1.5%，反残差结构将模型输入改为低维度特征，并用深度卷积替换标准卷积，参数量减少 193 MB，传输速率提高 1.35 倍，达到 78.3 fps。使用 MobileNetV1 和 MobileNetV2 作为基础网络，可以通过深度可分离卷积或反残差结构减少模型参数，提高检测速率，但会使网络层数加深，引起梯度消失问题，导致准确率提升有限甚至下降。综合考虑检测精度和速度，使用 MobileNetV3_Large 作为基础网络在保证车辆信息提取能力的同时加快了网络的计算速度，综合性能最优，故使用此网络作为 SSD 前端特征提取的基础网络。

表 2 不同基础网络模型性能对比

算法	mAP/%	模型大小/MB	帧率/fps
VGG16-SSD	80.4	218.8	57.9
MobileNetV1-SSD	79.8	28.2	69.1
MobileNetV2-SSD	81.5	31.8	66.9
MobileNetV3_Large-SSD	81.9	25.0	78.3

2)多尺度特征融合网络对模型精度的影响

为了验证 CBiFPN 的有效性，在相同基础网络上对特征金字塔网络(feature pyramid network，FPN)[19]、路径聚合网络(path aggregation network，PANet)[20]、全连接特征金字塔网络(full-connected feature pyramid network，FCFPN)、神经架构搜索特征金字塔网络(neural architec-

ture search feature pyramid network，NAS FPN)[21]、BiFPN 以及 CBiFPN 网络进行比较，结果如表 3 所示。传统 FPN 只进行深层到浅层的单向特征融合，精度提升有限；PANet 执行双向特征融合，精度相比于 FPN 进一步提升，但也引入了额外的参数；FCFPN 通过全连接进行特征融合，增加了大量的参数，精度却没有得到较大的提升；NAS FPN 重新组合和融合已经提取的特征图，参数量最小，精度也有所提升，但训练时间过长；BiFPN 与 PANet 类似均采用双向特征融合结构，并按照一定权值将不同尺度特征信息进行累加，在多个金字塔网络中取得较高的精度，同时参数量较小。本文提出的 CBiFPN 通过新增跨级数据流，更好的融合了不同层级之间的数据，mAP 相较于原 BiFPN 提升了 0.4%。

表 3　不同金字塔网络模型性能对比

网络	mAP/%	参数大小/MB
FPN	83.0	18.1
PANet	85.2	21.8
FCFPN	83.8	27.4
NAS FPN	84.4	12.5
BiFPN	86.3	15.7
CBiFPN	86.7	17.7

3)加入 ECA 通道注意力对模型精度的影响

为了使模型重点关注有效特征，同时削弱无效特征的影响，在每个特征提取层后引入高效通道注意力 ECA-Net，通过局部跨通道信息交互策略，获得表示特征通道局部关系和重要程度的权重。以 MobileNetV3_Large-SSD+CBiFPN 网络为基础网络，在相同的实验条件下与 SENet[22]，卷积块注意模块(convolutional block attention module，CBAM)[23]，ECA(手动选取 k 值)等注意力模型进行比较。从图 3 可以看出，模型自适应选取 k 值的精度均高于手动选取 k 值($k=3,5,7,9$)，验证了 ECA 自适应选取 k 值的有效性。在嵌入各种注意力模块的对比中，ECA 提升了 1.2%，提升效果最好，且 ECA 模型参数量最少。因此，本文引入的 ECA 注意力在参数量较少的情况下取得了更高的检测精度。

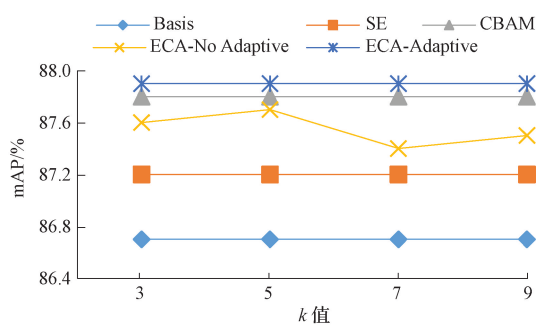

图 3　不同注意力模型对比

4)消融实验

为了验证 MCE-SSD 算法各个模块的效果，对不同模块进行消融实验，实验结果如表 4 所示。可以看出对每个单独模块的改进，都能够有效提升车辆的识别准确率。CBiFPN 模块提升效果最为明显，浅层网络包含比较多的几何信息，但是图像的语义特征并不多，不利于图像的分类；高层网络虽然能响应语义特征，但是由于 feature map 的尺寸太小，拥有的几何信息并不多，不利于目标的检测。通过双向加权融合浅层特征与深层特征，细节特征的提取效果得到加强，使得 MCE-SSD 易于发现和识别远距离小目标；对大目标也有较好的判断和定位。通道注意力 ECA 的引入将特征通道权重重新标定，为关键特征生成较大的权重，并加入到原输入特征图上，完成对各通道特征的重新编码，平均精度达到 82.2%。当融合所有模块时，车辆检测准确率显著提升了 7.5%，表明所有模块的融合能够共同促进车辆特征的学习，更易避免错判、漏判、误判等情况的发生。

表 4　不同模块对网络的影响

MobileNetV3-Large	CBiFPN	ECA	mAP/%
			80.4
√			81.9
	√		85.3
		√	82.2
√	√		86.5
√		√	83.4
	√	√	86.8
√	√	√	87.9

5)与其他方法的对比

如表 5 所示，在 KITTI 数据集上对 R-CNN、Fast R-CNN、Faster R-CNN、DAVE、SSD300、SSD512、SINet 以及本文提出的 MCE-SSD 算法进行评估和比较。从表 5 可以看出，本文提出的 MCE-SSD 算法获得了 87.9%的 mAP，并且传输速率达到 72 fps，均高于其他方法。

R-CNN、Fast R-CNN、Faster R-CNN 分别获得了 60.9%、62.7%、76.9%的 mAP，分别比 MCE-SSD 低 27.0%、25.2%与 11.0%。上述 3 种方法中检测速度最快的是 Faster R-CNN，但 12.8 fps 仍不适用于对实时性有较高要求的场合，这是由于双阶段检测算法增加了区域建议网络提取建议框的步骤，导致速度较慢。DAVE 以 79.1%的 mAP 超越了上述 3 种算法，但仍低于本文的 MCE-SSD 算法。对于 SSD300 算法，速度达到 57.9 fps，检测精度比 MCE-SSD 低 7.5%；SSD512 模型检测速度为 26.8 fps，检测精度也低于 SSD300，本文经过分析认为是 SSD512 网络对 KITTI 数据集图像的尺度和长宽比进行了较大的改变，无效维度增多，导致检测结果不佳。SINet 基于 VGG 网络提取特征，对各种类型车辆均有较好的检测精度，获得 84.2%的 mAP，比 MCE-SSD 低 3.7%，同时速度达到 31.3 fps，实现了实时检测。

表 5 不同车辆检测模型在 KITTI 数据集的检测结果

方法	基础网络	帧率/fps	AP/%				mAP/%
			Car	Van	Tram	Truck	
R-CNN		0.6	62.9	68.4	58.2	54.4	60.9
Fast R-CNN	VGG-16	0.6	64.8	57.9	65.3	62.8	62.7
FasterR-CNN	VGG-16	12.8	77.2	72.2	78.8	79.5	76.9
DAVE		3.9	83.6	71.4	80.3	81.3	79.1
SSD300	VGG-16	57.9	84.8	80.2	78.1	78.3	80.4
SSD512	VGG-16	26.8	76.1	81.1	76.0	74.4	76.9
SINet	VGG	31.3	88.6	83.8	84.0	80.5	84.2
MCE-SSD	MobileNetV3-Large	72.3	88.1	87.6	87.5	88.4	87.9

为了更好地展现算法的性能，进一步可视化了不同方法在 KITTI 测试集上的检测结果，如图 4 所示。可以看到，Faster-RCNN 只能检测到少量车辆目标，检测效果较差。SSD 对中小尺寸目标容易漏检。SINet 一定程度上改善了小目标检测效果，但对遮挡和重叠车辆的鲁棒性较差。MCE-SSD 有效解决了上述问题，可以识别这些较难目标并进行准确的检测。验证了本文提出的 MCE-SSD 的有效性。

(a) Faster-RCNN (b) SSD-300 (c) SINet (d) MCE-SSD

图 4 不同方法在 KITTI 数据集检测对比

3.3 基于 BDD100K 数据集的通用道路目标测试

BDD100K 数据集[24]由伯克利大学 AI 实验室发布，涵盖了 6 种真实路况场景，6 种天气环境以及 3 个时间段的复杂图片，是一个大规模、内容多样性的公开驾驶数据集。该数据集的道路目标数据包含 10 万张图片，包括 Bike、Bus、Car、Motor、Person、Rider、Traffic light、Traffic sign、Train and Truck 十种目标类别标签数据，按照 7∶2∶1的比例划分为训练集、测试集和验证集。

本文提出的 MCE-SSD 算法虽然是解决实时复杂环境下对车辆目标的检测，但算法的泛化性能和适应性较好，同样适用于常见道路目标的检测。因此，为进一步验证模型对一般道路目标的定位与识别性能，本文使用包含 10 种不同道路目标的 BDD100K 数据集进行训练和测试，该部分使用 512×512 输入的 MCE-SSD 算法，检测结果如表 6 所示。MCE-SSD 相比 SSD 提升了 3.5 的 mAP，检测速度为 44.8 fps，检测精度与检测速度均优于原始 SSD 算法，性能也优于其他单阶段方法。此外，MCE-SSD 比 RFBNet 高 3.1 mAP 的同时检测速度快 5.8 fps。与精度最高的 CFENet 方法相比，MCE-SSD 精度较低，但效率要高得多，实验结果表明，本文方法更好地权衡了准确性和推理速度。

表 6 不同车辆检测模型在 BDD100K 数据集的检测结果

算法	mAP/%	帧率/fps	输入尺寸
MS-CNN	5.7	6.0	1 920×576
SINet	9.2	20.2	1 920×576
SSD	14.1	27.6	512×512
ASSD	15.8	27.1	512×512
RefineDet	17.4	22.3	512×512
CFENet	19.1	21.0	512×512
RFBNet	14.5	39.0	512×512
YOLOv3	14.6	42.9	512×512
MCE-SSD	17.6	44.8	512×512

为了更好地理解算法的性能，进一步可视化了不同方法在 BDD100K 测试集上的检测结果，如图 5 所示。显然 MCE-SSD 可以检测到其他方法无法找到的远距离小尺寸对象，执行更精确的目标定位。此外，在夜间环境、遮挡重叠等场景下，本文方法仍然具有很好的检测能力，证明了 MCE-SSD 的优越性。

(a) SINet (b) SSD-512 (c) YOLOv3 (d) MCE-SSD

图 5 不同方法在 BDD100K 数据集检测对比

4 结 论

本文提出了一种快速准确的车辆检测模型 MCE-SSD，使用 MobileNetV3_Large 作为特征提取基础网络，使模型适用于计算资源有限的车辆检测终端且能保持较高的车辆检测精度；通过双向加权融合网络获取多维度的车辆特征，增强模型在多尺度检测场景下的适应性；ECA 通道注意力的引入提升了模型对有效特征的关注，进一步

提高了检测精度。在KITTI和BDD100K两个公开数据上的测试结果表明,本文所提方法在检测精度和检测速度上均可获得较好的检测效果。但当车辆目标出现遮挡等情况时会对算法造成一定影响,在未来的工作中将着重此方面进行改进,进一步提高算法识别性能。

参 考 文 献

[1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [J]. IEEE Computer Society, 2013: 580-587.

[2] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[3] REN S, HE K, 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.

[4] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Computer Society, 2017: 2961-2969.

[5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, 2016(7): 21-37.

[6] ZHANG S, WEN L, BIAN X, et al. Single-shot refinement neural network for object detection[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4203-4212.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[8] 刘寒迪,赵德群,陈星辉,等.基于改进SSD的航拍施工车辆检测识别系统设计[J].国外电子测量技术,2020,39(7):127-132.

[9] SIM I, LIM J H, JANG Y W, et al. Developing a compressed object detection model based on YOLO-v4 for deployment on embedded GPU platform of autonomous system [J]. Computer Science, 2021, arXiv:2108.00392.

[10] 石欣,卢灏,秦鹏杰,等.一种远距离行人小目标检测方法[J].仪器仪表学报,2022,43(5):136-146.

[11] HU X, XU X, XIAO Y, et al. SINet: A scale-insensitive convolutional neural network for fast vehicle detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(3): 1010-1019.

[12] 刘鸣瑄,刘惠义.基于特征融合SSD的远距离车辆检测方法[J].国外电子测量技术,2020,39(2):28-32.

[13] ZHAO Q, SHENG T, WANG Y, et al. CFENet: An accurate and efficient single-shot object detector for autonomous driving[J]. Computer Science, 2018, arXiv:1806.09790.

[14] 梁继然,陈壮,董国军,等.结合注意力机制和密集连接网络的车辆检测方法[J].电子测量与仪器学报,2022,36(3):210-216.

[15] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.

[16] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.

[17] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.

[18] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: The kitti dataset [J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.

[19] LIN T Y, DOLLAR P, GIRSHICK R, et al. Featurepyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[20] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[21] GHIASI G, LIN T Y, LE Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7036-7045.

[22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition, 2018: 7132-7141.

[23] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision, 2018: 3-19.

[24] YU F, CHEN H, WANG X, et al. BDD100K: A diverse driving dataset for heterogeneous multitask learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2636-2645.

作 者 简 介

张奇，硕士研究生，主要研究方向为目标检测、人体姿态估计等。

E-mail: zqzx789@163.com

陈梦蝶，硕士研究生，主要研究方向为行人重识别、图像处理等。

E-mail: 2372699331@qq.com

赵杰，硕士研究生，主要研究方向为3D点云分割、图像处理等。

E-mail: 1569391518@qq.com