结合注意力机制的车型检测算法

谢斌红,赵金朋,张英俊
(太原科技大学 计算机科学与技术学院,山西 太原 030024)

摘 要:针对目标检测算法应用在车辆类型检测的场景中,检测速度较快,但检测精度相对较低的问题,该文对 CenterNet 算法进行改进。首先,使用 ResNet 作为主干网对车型图像进行特征提取,并在特征提取网络中引入通道注意力和空间注意力,对不同通道以及不同位置的特征进行权重划分,获取更多需要关注的特征,抑制无用的特征,进而提升车型检测算法的分类及定位准确率;其次,针对小目标车型检测精度不高的问题,将不同尺度车型特征进行融合,更好地提取细粒度车型特征,提升检测精度。为验证结合注意力机制的车型检测算法的有效性,在 KITTI 车型数据集和 BIT-Vehicle 数据集上进行实验,mAP 值分别达到 94.6% 和 95.5%。结果表明改进后的算法模型在检测速度影响较小的情况下检测精度得到显著提升。

关键词:智慧交通;目标检测;特征融合;注意力机制;残差网络;可变形卷积
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2021)12-0078-07
doi:10.3969/j.issn.1673-629X.2021.12.014

Vehicle Detection Algorithm Combined with Attention Mechanism

XIE Bin-hong,ZHAO Jin-peng,ZHANG Ying-jun
(School of Computer Science and Technology,Taiyuan University of Science and Technology,
Taiyuan 030024,China)

Abstract:Aiming at the problem of fast detection speed and relatively low detection accuracy for the object detection algorithm in the scene of vehicle type detection,we improve the CenterNet algorithm. Firstly,ResNet is used as the backbone network to perform feature extraction on vehicle images,and the channel attention and spatial attention are introduced into the feature extraction network to carry out the weight division of the features in different channels and at different positions,to obtain more features that need attention and suppress useless features,thus improving the classification and positioning accuracy of vehicle detection algorithms. Secondly,in view of the problem of low detection accuracy of small target vehicles,we integrate the features of different scale vehicle to better extract fine-grained vehicle features and improve detection accuracy. To verify the effectiveness of the vehicle detection algorithm combined with the attention mechanism,experiments were conducted on the dataset of KITTI and BIT-Vehicle. The mAP values reached 94.6% and 95.5% respectively. The results show that the improved algorithm can significantly improve the detection accuracy with little influence on the detection speed.

Key words:intelligent transportation;object detection;feature fusion;attention mechanism;residual network;deformable convolution

0 引 言

随着社会的发展,国内的汽车数量不断增加,种类也日益丰富,使用计算机技术对交通图像中的车型进行识别检测已经成为计算机视觉领域的一项重要应用。在不同的场景中检测不同的车型具有广阔的应用前景,例如:在无人驾驶领域,通过识别图像中的车辆类型和位置,可以规避车辆的碰撞;在智能交通管理中,可以用于市区车辆的限行,也可以进行更精准的车流检测等。

近年来,由于深度卷积神经网络(DCNNs)的发展和计算机计算能力的提升,基于深度学习的车型检测技术引起了人们的广泛研究。Sengar[1]等人采用一种基于双向光流块的运动目标检测算法,该算法实验效果较好,但是需要对比前后两帧图像,需要输入视频,不适用于静态检测。孙皓泽等人[2]提出使用 MobileNet 网络对装甲车进行检测识别,该方法适用于计算资源受限场景,但在检测精度上仍有待提高。为了能够直接检测图像中的车辆类型,提高实时检测准

收稿日期:2020-12-15 **修回日期**:2021-04-15
基金项目:山西省重点研发计划(重点)高新领域项目(201703D111027);山西省重点研发计划项目(201803D121048,201803D121055)
作者简介:谢斌红(1972-),男,硕士,副教授,CCF 会员(43374M),通信作者,研究方向为智能化软件工程、机器学习。

确率,该文提出将深度学习的目标检测方法用于实时车型检测任务中,解决实际场景应用中车型的检测精度和速度问题。

目前,基于深度学习的目标检测方法主要分为两类:基于候选区域的双阶段检测器和基于回归框的单阶段检测器。其中,双阶段检测器的检测过程分为两个步骤,第一步从图像中生成候选区域(可能存在目标的区域),然后将候选区域作为输入,输入神经网络提取特征,进行目标的类别和回归框位置的检测,典型的网络有 R-CNN[3]、Fast R-CNN[4]、Mask R-CNN[5]、PANet[6]、TridentNet[7]。而单阶段检测器省去了候选区域的生成过程,将目标检测任务视为回归任务,直接对输入的图像进行回归预测并输出结果,典型的网络有 SSD[8]、YOLO[9-11]、ConerNet[12]、FSAF[13]、FCOS[14]。

两种方法相比较,双阶段检测方法的精度略高于单阶段检测方法的精度,但是其在检测速度方面表现不如单阶段检测方法,不足以满足车型检测的实时要求;而单阶段检测方法在检测速度方面有着很好的表现,故采用单阶段的检测方法进行车型检测。

现在主流的基于单阶段的车型检测方法在检测速度和精度方面都有着较好的表现,但大多数方法是基于 Anchor 框的,需要人为预先设置 Anchor 的一些大小和比例等超参数。

该方法存在以下不足:

(1)算法对预先设定的图像的大小、Anchor 框的长宽比和数量比较敏感;

(2)由于与 Anchor 相关的超参数是预先设定的,使得算法无法自适应检测目标的大小,且对变形较大的目标检测效果不太理想;

(3)该类方法计算量和内存开销较大,因为需要和真实结果多次地计算 IOU(intersection over union);

(4)该方法会使得数据集的正负样本不平衡,为了获得较高的召回率,需要在特征图上密集地部署 Anchor,而其中大部分是负样本,会加剧正负样本的不平衡。

针对上述不足,该文提出基于 Anchor-Free 的车型检测方法。该方法减少了车型检测模型的设计复杂度和超参数的设置难度,从而简化训练过程,提升模型检测速度;取消了 Anchor 的设置,在减少计算量的同时可以更好地适应不同尺寸的车辆特征。同时,为了解决车型检测过程中对车辆关键特征提取能力不足的问题,在 CenterNet[15] 的基础上引入了混合注意力机制;此外为了更好地提取不同尺寸的车型特征,将不同尺度的特征图进行了融合。在增加了极少参数量的同时提升了检测精度。

1　相关工作

1.1　注意力机制

注意力机制(attention mechanism)源于人们对视觉的研究。人类视觉系统的一个重要特性是人们不会一次尝试处理整个场景的信息,而是有选择地聚焦于有重要特征信息的区域。Jaderberg 等人[16] 在 Spatial Transformer Networks 中提出了用于分类任务的空间注意力模块,该模块允许对特征数据进行空间变换。Wang 等[17] 使用编码器式注意模块的残差注意网络,通过细化特征图,使得网络在提升性能的同时增加了对噪声的鲁棒性。注意力机制已被广泛地应用于序列化标注、图像识别和目标检测等场景。使用注意力机制来提升卷积神经网络在大规模图像分类、检测任务中的效果,故该文使用注意力机制提升车型检测效果。

1.2　网络结构

该方法是一种基于 Anchor-Free 的单阶段目标检测算法,在速度和精度方面都有很好的表现,并且在摒弃 Anchor 后,减少了人为设置超参的影响。本研究采用 ResNet-34 作为主干网,其网络结构如表 1 所示,该网络很好地解决了深度神经网络的退化问题。

表 1　主干网的结构

Layer	Output size	ResNet-34
输入	512×512×3	
Conv1	256×256×64	$7\times7,64,S=2$
Conv2	128×128×64	3×3max pool, $S=2$ $\begin{Bmatrix}3\times3,64\\3\times3,64\end{Bmatrix}\times3$
Conv3	64×64×128	$\begin{Bmatrix}3\times3,128\\3\times3,128\end{Bmatrix}\times4$
Conv4	32×32×256	$\begin{Bmatrix}3\times3,256\\3\times3,256\end{Bmatrix}\times6$
Conv5	16×16×512	$\begin{Bmatrix}3\times3,512\\3\times3,512\end{Bmatrix}\times3$

该方法最大的特点就是使用关键点估计网络来寻找车辆的中心点,然后采用回归的方法对车辆的边框大小进行学习。具体来讲,首先将图像 $I\in R^{w\times h\times3}$ 输入到网络中,经过多层卷积最终生成一个大小为 $\hat{Y}\in[0,1]^{\frac{W}{R}\times\frac{H}{R}\times C}$ 的热力图(R 为步长 4,C 为预测类别数),如果 $\hat{Y}_{x,y,c}=1$,则该点为关键点;$\hat{Y}_{x,y,c}=0$ 则为不包含目标的背景。

模型训练时,首先使用 ResNet-34 进行特征提取,然后对提取出来的特征经过多层可变形卷积(deformable convolutional networks),将特征图尺寸进行四次下采样,由 512×512 缩小到 128×128,最后形成

三个并行分支,分别预测车辆的类别损失 L_k 、边框损失 L_{size} 以及车辆中心偏移损失 L_{off} 。损失函数的计算公式如式(1)所示,其中 λ_{size} 为 0.1, λ_{off} 为 1。

$$L_{\text{det}} = L_k + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}} \tag{1}$$

其中,车辆类别损失 L_k 使用了 Focal loss 函数,该损失函数可以有效缓解单阶段检测器中样本类别不平衡的问题,使数量多的类别所贡献的损失被大幅削减,数量少的类别所贡献的损失几乎没有多少降低。其中, p 为关键点,特征图尺寸缩小后的位置为 $\hat{p} = \lfloor \frac{p}{R} \rfloor$,将所有的关键点使用高斯核 $Y_{xyc} = \exp[-\frac{(x-\tilde{p}_x)^2 + (y-\tilde{p}_y)^2}{2\sigma_p^2}]$ 映射到热力图 $Y \in [0,1]^{\frac{W}{R}\times\frac{H}{R}\times C}$ 上,其中 σ_p 是目标大小自适应标准差[12]。分类损失具体计算公式如式(2)所示。

$$L_k = \frac{-1}{N}\sum_{xyc}\begin{cases}(1-\hat{Y}_{xyc})^{\alpha}\log(\hat{Y}_{xyc}), & \text{if } Y_{xyc}=1\\ (1-Y_{xyc})^{\beta}(\hat{Y}_{xyc})^{\alpha}\log(1-\hat{Y}_{xyc}), & \text{otherwise}\end{cases} \tag{2}$$

在分类损失中, α 为 2, β 为 4, N 为关键点个数,该超参的选择依据 Law[18] 等人的实验。

此外,图像下采样过程中,中心点会因为数据离散而产生误差,因此,还引入了局部偏移预测 $\hat{O} \in R^{\frac{W}{R}\times\frac{H}{R}\times 2}$,所有车型类别的关键点共享偏移预测。偏移损失具体计算公式如式(3)所示。

$$L_{\text{off}} = \frac{1}{N}\sum_{p}\left|\hat{O}_p - (\frac{p}{R} - \tilde{p})\right| \tag{3}$$

最后,假设 $(x_1^k, y_1^k, x_2^k, y_2^k)$ 为车辆 k 的边框,中心坐标 $p_k = (\frac{x_1^k + x_2^k}{2}, \frac{y_1^k + y_2^k}{2})$,车辆的大小为 $s_k = (x_2^k - x_1^k, y_2^k - y_1^k)$,则车辆尺寸的损失计算公式如式(4)所示。

$$L_{\text{size}} = \frac{1}{N}\sum_{k=1}^{N}|\hat{S}_{p_k} - s_k| \tag{4}$$

2 改进网络

2.1 注意力模块

目前常见的注意力机制划分方式有三种,按照关注区域可以分为软注意力和硬注意力;按照输入形式可以分为基于项的注意力和基于位置的注意力;如果按照注意力域(attention domain)分类,则包含三种注意力域:空间域(spatial domain)、通道域(channel domain)和混合域(mixed domain)。

通道注意力的作用是通过对特征图的各个通道之间的依赖性进行建模以提高对于重要特征的表征能力。目前生成通道注意力的方式有以下几种:平均池化、最大池化、结合平局池化和最大池化、方差池化。其生成过程类似,首先通过在各层特征图上的池化获得各个通道的全局信息,然后使用全连接层进行特征提取,ReLU 进行非线性激活,最后使用 Sigmoid 进行权重归一化,通过该过程自适应地对各通道特征的相关程度进行建模,最后再将原特征通道的信息与自适应学习建模后的权重进行加权处理,实现特征响应及特征重校准的效果。

使用注意力机制的网络在前向传播的过程中,重要的特征通道将会占有更大的比重,在最终所呈现的输出图像中也能更加明显地表征车型检测网络所重点关注的部分,更加关注图像的内容特征,更好地分辨出车辆的类别。

空间注意力需要为特征图生成一个空间注意力图,用于增强或抑制不同位置的特征。空间注意力的方式有两种:最大池化和平均池化结合、标准卷积($1*1$, $S=1$,不同卷积核大小)。通过空间注意力,能够更好地展示网络所要关注的重点位置,更加关注图像的位置特征,更好地对车辆进行定位。

混合注意力,顾名思义就是将图像的通道特征和空间特征引入到特征提取的过程。Convolutional Block Attention Module (CBAM)[19] 就是使用了混合注意力机制,同时关注通道和空间的特征,以此来提高神经网络在类别以及位置的表征能力。

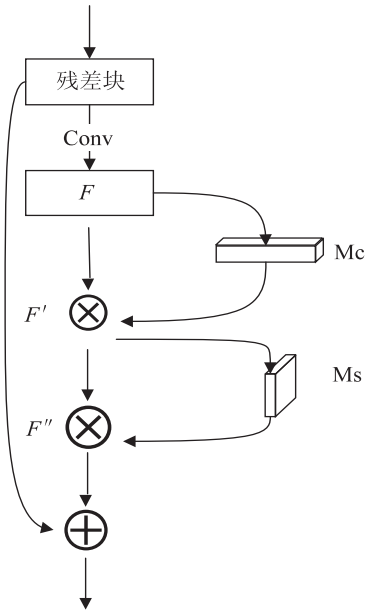

图 1 引入注意力的残差模块

本研究在 ResNet[17] 的残差模块中融入混合注意力机制,用于提升车型类别以及车辆位置的表征能力。图 1 为引入注意力之后的残差模块结构图。从图 1 可知,输入图像经过卷积之后,首先将特征图输入到通道注意力模块,经过全局平均池化和全局最大池化操作

后依次通过两次全连接和 Sigmoid;将通道注意力模块输出的特征图输入到空间注意力模块中,经过通道最大池化和通道平均池化后输入到全连接和 Sigmoid;最后再和残差连接结合一并输出。

通道注意力输入特征图 $F \in R^{c\times h\times w}$ (c 为通道数, h 、w 为图像的高宽),会生成一个一维的通道注意力图 $M_c \in R^{c\times 1\times 1}$ 。生成过程如图 2 所示(图中 S 代表 Sigmoid)。具体注意力特征图计算公式如式(5)所示。

$$\begin{cases} F' = M_c(F) \otimes F \\ M_c(F) = \text{sigmoid}(\text{Avg}_{\text{out}} + \text{Max}_{\text{out}}) \\ \text{Avg}_{\text{out}} = \text{Fc}(\text{ReLU}(\text{Fc}(\text{Avg}(F)))) \\ \text{Max}_{\text{out}} = \text{Fc}(\text{ReLU}(\text{Fc}(\text{Max}(F)))) \end{cases} \tag{5}$$

其中,全局平均池化输出为 Avg_{out} ,全局最大池化输出为 Max_{out} ,Fc 为全连接,ReLU 为激活函数。

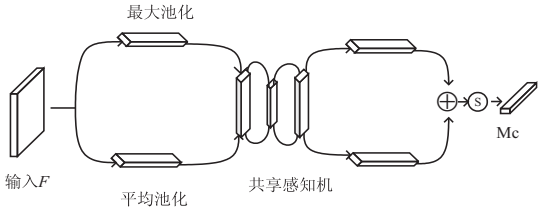

图 2　通道注意力结构

空间注意力将通道注意力的输出作为输入,输入到网络,运算后生成一个二维的空间注意力图 $M_s \in R^{1\times h\times w}$ 。具体注意力特征图计算公式如式(6)所示,Avg 为平均池化操作,Max 为最大池化操作,Cat 为张量拼接运算。生成过程如图 3 所示。

$$F'' = M_s(F') \otimes F'$$
$$M_s(F') = \text{Sigmoid}(\text{conv}(\text{Cat}(\text{Avg}(F') + \text{Max}(F')))) \tag{6}$$

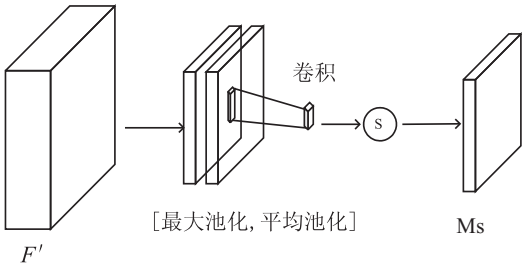

图 3　空间注意力结构

2.2　特征融合

在特征提取过程中,ResNet-34 进行了四次下采样,将图像原始尺寸进行了四次缩放,因此,图像中的一些小目标在进行特征提取时,其分辨率逐渐下降,在网络的末端小目标的特征信息可能就会丢失,从而影响小目标的检测精度。所以为了提高车辆目标检测效果,更好地提取图像中车型的细粒度特征,通过引入特征融合,将可以更好地保留上层的特征,减少特征信息的损失,从而提升识别精度,具体过程如下。

首先,将残差网络中 C3 层的特征进行下采样操作,并通过 1×1 卷积改变通道数,与 C5 层的特征进行融合,然后将融合之后的特征一并进行后续运算。图 4 为引入特征融合的整体网络结构,加粗连接线为引入的特征融合。

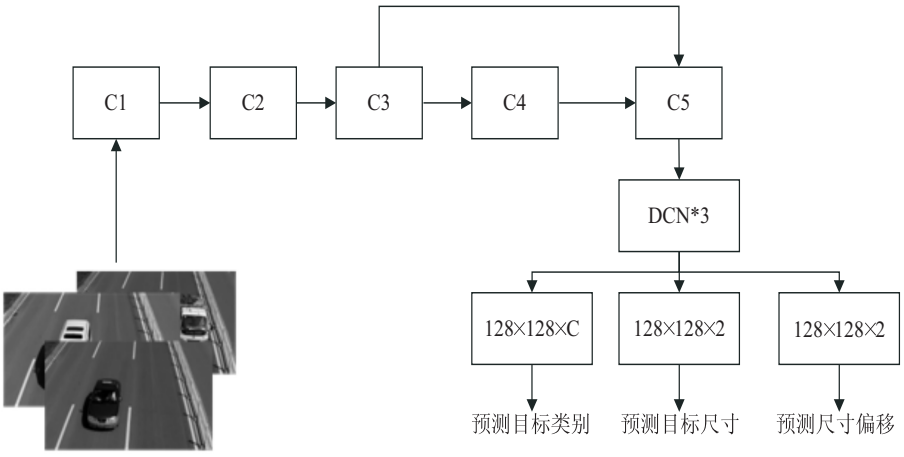

图 4　网络结构

2.3　图像增强

数据增强也称为数据增广,目的是增加数据集的规模,更好地训练模型,让模型有更好的检测能力,防止模型过拟合。为了提升车型检测模型的泛化能力,提升检测性能,从而更好地进行车型检测,该文首先对实验数据集进行了翻转增强,然后再使用增强后数据集进行训练。

3　实验结果与分析

3.1　数据集

文中使用的数据集为 KITTI 车型数据集和 BIT-Vehicle 数据集。其中,KITTI 数据集是由丰田美国技术研究院同德国卡尔斯鲁厄理工学院联合创建,该数据集是目前国际上最大的数据集,主要用于自动驾驶

场景下的计算机视觉算法评测。

KITTI 车型数据集一共有 7 481 张图像,包含小汽车(Car)、厢式货车(Van)、卡车(Truck)和电车(Tram)四种车型。实验中将数据集划分为两部分,其中 5 000 张作为训练集,2 481 作为测试集,训练标签总共有 17 637 个,测试标签有 15 627 个,具体每类车型标签数如图 5 所示。

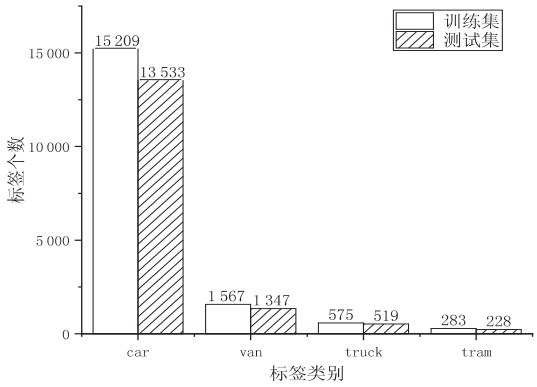

图 5 KITTI 数据集

另一个数据集是 BIT-Vehicle 车型数据集,它包含了公共汽车(Bus)、越野车(SUV)、轿车(Sedan)、小货车(Minivan)、中巴(Microbus)和卡车(Truck)6 种车型,共 9 850 张图像。本次实验将数据集划分为两部分,6 000 张用于训练,3 850 张用于测试,详细类别的标签数如图 6 所示,该数据集中的图像均采自于实际的交通高清摄像头。

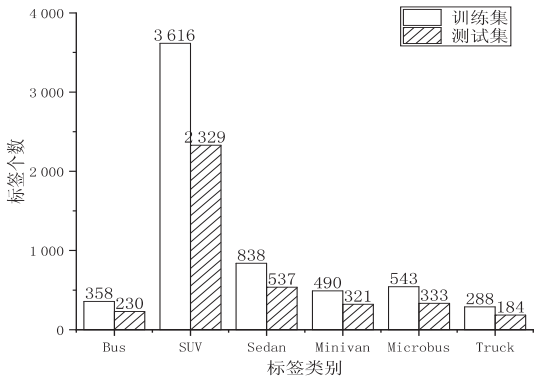

图 6 BIT-Vehicle 数据集

实验中的数据集格式为 COCO,所以需要对原始标签进行数据格式的转化。具体步骤如下:

(1)将 KITTI 转化为 txt 格式;

(2)从 txt 中筛选车辆类别;

(3)txt 格式标签转化为 XML 格式;

(4)将 XML 格式标签转化为 Json 格式用于训练和测试。

3.2 评价指标

实验使用各车型类别 AP 的平均值(mean average precision,mAP)和每秒检测帧数(frames per second,fps)作为评价指标。mAP 通过计算 IOU=0.5 时的精度 precision 和召回率 recall 得到每类车型的 PR(precision-recall)曲线,然后计算 PR 曲线与其下的面积得到该类别的平均精度 AP,最后,计算所有类别 AP 的平均值得到 mAP。而 fps 则是首先通过计算出检测一张图片所消耗的时间,然后计算每秒可以检测多少张图计算而来。

3.3 实验设置

本研究中使用的实验配置如下:CPU:Inteli7 8700K;RAM:16 G;GPU 加速库:CUDA 10.0,CUDNN 7.5.0;GPU:Nvidia GTX1080Ti;实验平台的操作系统为 Ubuntu16.04,实验程序开发使用了基于 Python 机器学习库的 Pytorch 框架。

网络训练过程中,首先在 ImageNet 数据集上进行预训练,然后在车型数据集上进行微调。训练参数设置如下:batch_size 为 32,epoch 为 120,初始学习率 0.000 125,并在第 75 个和 100 个 epoch 时分别下调学习率,每次下调为原来的 1/10。

3.4 实验结果及分析

为验证文中方法的有效性,与现有的方法进行对比,在 BIT-Vehicle 数据集上的实验结果如图 7 所示。

由图 7 可知,文中方法与 Yolov3 相比,在 Truck、SUV、Microbus 三种车型数据集上识别精度有比较明显的提升,同时速度也由 35 fps 提升至 43 fps,能够更好地应用于实时车型检测。

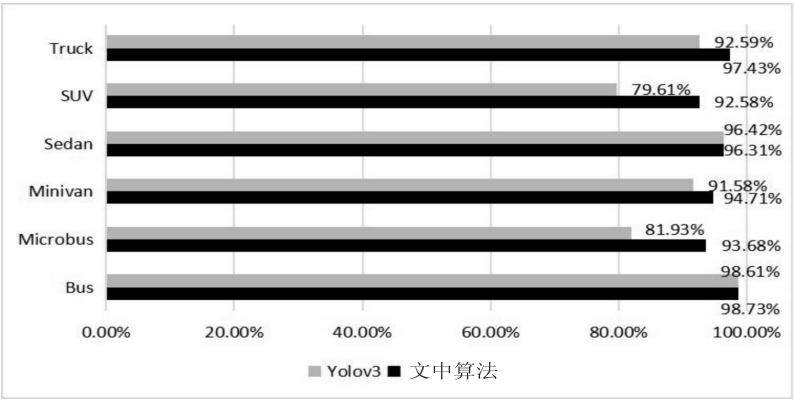

图 7 BIT-Vehicle 数据集实验结果

文中方法与其他方法在 KITTI 数据集上的实验结果如表 2 所示。分析表中数据可知,文中方法与原 CenterNet[15] 方法相比 mAP 提升了 2.3%,而检测速度基本不受影响。这就说明混合注意力的引入能够很好地提升车型检测的精度;并且与现在主流的方法相比,能够在速度与精度之间达到了一个很好的平衡。和 DF-YOLOv3[22] 相比,虽然速度慢了 2 fps,但精度提升接近 1%。

表 2　KITTI 数据集实验结果

方法	输入	fps	mAP/%	Car	Van	Truck	Tram
Faster R-CNN(VGG16)[20]	600×600	11.63	76.90	77.18	72.17	79.48	78.77
DAVE[21]	60×60/224×224	3.86	79.14	83.56	71.44	81.32	80.25
DF-YOLOv3[22]	512×512	45.48	93.61	92.57	93.90	95.08	92.87
SSD300[8]	300×300	58.32	81.00	82.04	74.59	86.42	80.394
SINet(VGG)[23]	384×1 280	31.32	86.02	90.39	85.56	82.32	85.81
CenterNet[15]	512×512	45.45	92.3	91.67	92.54	93.13	91.85
文中方法	512×512	43.48	94.6	94.88	94.63	95.61	93.26

通过对上述实验结果的分析,证明了通过融入注意力模块,对车辆的空间信息以及通道信息进行权重划分,同时进一步融合了不同尺度的车型特征,虽然增加了模型参数,但检测速度不受较大影响,同时提升了车辆检测的精度,从而验证了文中方法的有效性。

此外,为了更好地分析文中方法,对车型中心点检测结果进行可视化展示。图 8 为该方法在 BIT-Vehicle 数据集上的检测结果,其中第一行为原始输入图像,第二行为预测的关键点效果图,最后一行为检测结果图。从图中可以看出该方法能够很好地预测车辆的中心位置。此外,通过观察检测结果发现,在光照充足的情况下,图像中会有车的阴影,这会一定程度上影响检测效果。

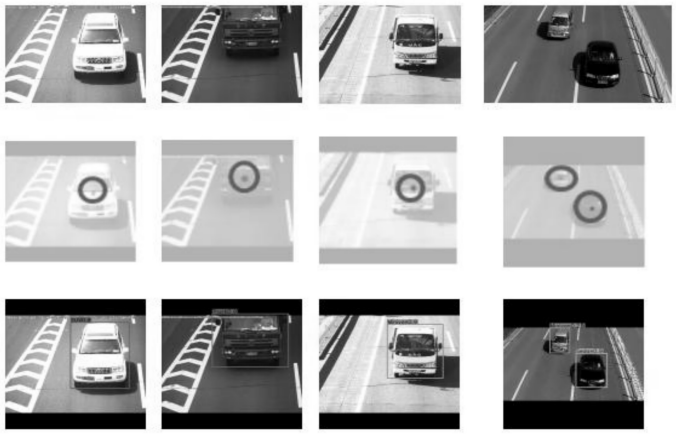

图 8　检测结果

4　结束语

针对当前车型检测方法存在精度、速度较低和数据集少的问题,首先使用图像增强对车型数据集进行数据增强,为车型检测模型提供了规模更大的数据集。同时为了适应不同尺寸的车型以及多目标检测等情况,通过使用混合注意力模块和特征融合对 Centernet[15] 进行改进,最终得到混合注意力卷积神经网络,提高了车型检测精度。在 KITTI 数据集和 BIT-Vehicle 数据集上分别进行实验,其在测试集上的平均检测精度分别达到了 94.6%、95.5%,与现有的一些车型检测算法对比结果显示,该方法更适用于车型检测任务,能够直接对图像进行车型检测,并且能够在速度和精确率上实现了一个很好的平衡。

在未来工作中,将探索更优的注意力模块,同时使用更好的图像处理方法,来适应复杂的应用环境,促进深度学习在车型检测、自动驾驶等任务上的应用。

参考文献:

[1] SENGAR S S, MUKHOPADHYAY S. Motion detection using block based bi-directional optical flow method[J]. Journal of Visual Communication and Image Representation, 2017, 49: 89-103.

[2] 孙皓泽,常天庆,张　雷,等. 基于轻量级网络的装甲目标快速检测[J]. 计算机辅助设计与图形学学报,2019,31

(7):1110-1121.
[3] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Washington,DC,United States:IEEE,2014:580-587.
[4] GIRSHICK R. Fast R-CNN[C]//2015 IEEE international conference on computer vision (ICCV). Santiago, Chile: IEEE,2015:1440-1448.
[5] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]//2017 IEEE international conference on computer vision (ICCV). Venice,Italy:IEEE,2017:2980-2988.
[6] LIU S,QI L,QIN H,et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City,UT, USA:IEEE,2018:8759-8768.
[7] LI Y,CHEN Y,WANG N,et al. Scale-aware trident networks for object detection[C]//2019 IEEE/CVF international conference on computer vision (ICCV). Seoul,Korea (South):IEEE,2019:6053-6062.
[8] LIU W,ANGUELOV D,ERHAN D,et al. SSD:single shot multibox detector[C]//Computer vision - ECCV 2016. Amsterdam,The Netherlands:Springer,2016:21-37.
[9] 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 (CVPR). Las Vegas,NV,USA:IEEE,2016:779-788.
[10] REDMON J,FARHADI A. YOLO9000:better,faster,stronger[C]//2017 IEEE conference on computer vision and pattern recognition (CVPR). Honolulu,HI,USA:IEEE,2017: 6517-6525.
[11] REDMON J,FARHADI A. Yolov3:an incremental improvement[J]. arXiv:1804.02767,2018.
[12] LAW H,DENG J. Cornernet:detecting objects as paired keypoints[J]. International Journal of Computer Vision,2020, 128(3):642-656.
[13] ZHU C,HE Y,SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach,CA,USA:IEEE,2019:840-849.
[14] TIAN Z,SHEN C,CHEN H,et al. FCOS:fully convolutional one-stage object detection[C]//2019 IEEE/CVF international conference on computer vision (ICCV). Seoul,Korea (South):IEEE,2019:9626-9635.
[15] ZHOU X,WANG D,KRÅHENBÜHL P. Objects as points [J]. arXiv:1904.07850,2019.
[16] JADERBERG M,SIMONYAN K,ZISSERMAN A,et al. Spatial transformer networks[C]//Proceedings of the 28th international conference on neural information processing systems. Montreal,Canada:MIT Press,2015:2017-2025.
[17] WANG F,JIANG M,QIAN C,et al. Residual attention network for image classification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Honolulu,HI,USA:IEEE,2017:6450-6458.
[18] LIN T Y,GOYAL P,GIRSHICK R,et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,42(2):318-327.
[19] WOO S,PARK J,LEE J Y,et al. CBAM:convolutional block attention module[C]//Computer vision-ECCV 2018. Munich,Germany:Springer,2018:3-19.
[20] 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.
[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):1-12.
[22] 张富凯,杨 峰,李 策. 基于改进 YOLOv3 的快速车辆检测方法[J]. 计算机工程与应用,2019,55(2):12-20.
[23] 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.