

DOI:10.19651/j.cnki.emt.2209236

基于改进 CenterNet 的自动驾驶小目标检测^{*}于方程 张小俊 张明路 赵天亮
(河北工业大学机械工程学院 天津 300401)

摘要: 自动驾驶领域主流目标检测算法对小目标检测效果差,给行车安全带来了威胁,对单阶段无锚框 CenterNet 算法进行改进以解决此问题。首先,替换原主干网络为具有分裂注意力机制的 ResNeSt50 网络,并将 ReLU 激活函数升级为 FReLU,以极少的额外计算开销强化了特征提取效果;然后提出轻量级网络 PASN 融合不同尺度的语义特征,并在浅层特征输入端引入空间池化金字塔(SPP)模块强化小目标信息的表达;最后在 Kitti 数据集进行随机多尺度输入训练。验证集结果表明改进后算法的 FPS 达到 37.7 满足实时性要求,小目标检测精度较原算法提 12.9%,平均检测精度提升 13.9%,同时检测速度与精度均高于主流算法 YOLOv4;在实车上每秒可检测 31 帧图像,为自动驾驶技术发展提供有力支持,具有工程应用价值。

关键词: 自动驾驶;小目标;无锚框;分裂注意力;CenterNet

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.20

Automatic driving small target detection based on improved CenterNet

Yu Fangcheng Zhang Xiaojun Zhang Minglu Zhao Tianliang
(College of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China)

Abstract: The mainstream target detection algorithms in the field of automatic driving have poor detection effect on small targets, which poses a threat to driving safety. The one-stage anchor-free CenterNet algorithm is improved to solve this problem. Firstly, the original backbone network is replaced by ResNeSt50 network with split-attention mechanism, and the ReLU activation function is upgraded to FReLU, which strengthens the effect of feature extraction with little additional computational overhead. Then, a lightweight network PASN is proposed to fuse semantic features of different scales, and spatial pooling pyramid (SPP) module is introduced into the shallow feature input to enhance the expression of small target information. Finally, random multi-scale input training is carried out on Kitti data set. The verification set results shows that the FPS of the improved algorithm reaches 37.7, meets the real-time requirements, the average precision of small targets is improved by 12.9% and the mean average precision is improved by 13.9%. At the same time, the detection speed and average precision are higher than the mainstream algorithm YOLOv4. It can detect 31 images per second on the real vehicle, which provides strong support for the development of automatic driving technology and has engineering application value.

Keywords: automatic driving; small target; anchor-free; split-attention; CenterNet

0 引言

近些年,基于深度学习的目标检测算法发展快速,学者们的不断研究使得算法在综合性能与工程应用等方面取得巨大进步^[1-3]。伴随着研究的深入,一个难题被发现:常规检测时小目标的检测精度明显低于其他尺寸目标^[4]。原因是目标检测领域中规定所占像素面积小于总面积 0.12% 的目标为小目标,而算法设计时,随着网络层数的加深,网

络感受也迅速增大,此时小目标因为分辨率低、像素占比小,其有效的特征难以被提取,导致识别率很低^[5]。且此问题在自动驾驶领域的危害更加突出,因为自动驾驶场景目标检测时小目标多为行人尤其是儿童或其他较小障碍物,一旦检测不及时或者是不准确会对人身安全造成巨大威胁,为了降低行车安全隐患,保证行车的主动安全性,小目标问题需要被重点考虑。

国内外学者关于小目标问题的研究方法主流是基于深

收稿日期:2022-03-11

* 基金项目:天津市新一代人工智能科技重大专项(18ZXZNGX00230)资助

度学习的方法,且解决手段分为两类:增强小目标信息与改善小目标训练方法^[6]。增强小目标信息常通过提升网络的特征提取能力以及特征融合能力达到目的,如:He 等^[7]在 2015 年提出一种空间金字塔池化层(spatial pyramid pooling, SPP),其使神经网络可处理不同大小的候选区,一定程度上克服了多尺度提取以及小目标特征表达能力弱的问题。Lin 等^[8]在 2017 年提出了特征金字塔网络(feature pyramid networks, FPN),其利用上下采样的方式实现多尺度特征融合。孙备等^[9]则在 2021 年提出了基于多尺度卷积融合结构和空间注意力加强的改进型 SSD 目标检测算法,提升对弱纹理小目标特征保持性,小目标检测率相对原生 SSD 算法提升明显。而对于改善小目标训练方法方面,常使用训练过程中平横大、小样本及创建小目标专项数据集等特殊手段,如:Singh 等^[10]提出了一个不变尺度的目标检测架构 SNIP,其通过学习不同尺度目标,提升了小目标的检测精度。

学者们虽然对小目标问题的研究有了诸多进展,但对小目标在自动驾驶领域特有的危害性研究的仍不充分,优化算法时没有着重强化对小目标的检测能力,增加了行车安全隐患。如王新等^[11]在 2021 年提出了一种高分辨率网络 CenterNet-DHRNet,其通过迭代聚合的方式对不同分辨率的特征图进行融合,提高网络的分辨率,最终提升了算法检测精度,但其这种方式没有考虑小目标问题的危害性,难以实现实车应用。

因此,为了提升目标检测算法小目标检测精度,降低智能汽车行车安全隐患,本文对实时性较强的单阶段无描框目标检测算法 CenterNet^[12]改进,有效提升算法小目标检测精度并高效部署于实车。

本文首先将 CenterNet 主干网络 ResNet18 替换为 ResNeSt50^[13]以高效平衡算法实时性及检测精度,引入 FReLU(funnel ReLU activation)^[14]激活函数以极低的额外计算开销强化了特征提取效果;然后提出 PASN 网络融合不同尺度语义特征,引入 SPP 模块增强小目标特征表达,并使用深度可分离卷积对网络轻量化处理,降低实车部署难度。最后通过 Kitti 数据集及实车实验双重验证,证明本文提出算法对小目标检测精度高,算法实时性与检测精度的平衡性好,综合性能强,较主流算法更适用于智能汽车实车应用,有效提升了智能汽车行车主动安全性,加速了自动驾驶技术全面落地。

1 CenterNet 算法原理

CenterNet 为单阶段无描框目标检测算法,不仅检测速度优于双阶段及预描框系列,其特殊的无描框机制亦可避免小目标因尺寸较小而难以匹配大多数先验框的问题,对小目标检测具有优势^[15]。CenterNet 使用对象边界框的中心点作为唯一的表示对象的关键点,即在热力图中找到物体的中心点,然后将找到的中心点进行宽、高回归得到预

测框,完成图像检测工作。CenterNet 共设计了 3 个主干网络模型分别为 ResNet18、DLA-34 以及 Hourglass-104,常设置输出下采样因子为 4,取热力图峰值的前 100 为检测目标的中心点,并设置筛选值得到最终检测目标的中心点^[16]。CenterNet 输入图片大小为 $W \times H \times 3$, W 和 H 分别为图像的宽高,经 R 次下采样倍数(本文设定为 4),并经过 3 次反卷积整合通道,最后输入 3 个解码头部(Center-Head)对特征图进行预测。

本文选取模型最小的 ResNet-18 网络为 CenterNet 主干网络优先保证实时性。CenterNet 主干网络采用 ResNet18 时设定输入图像大小为 $512 \times 512 \times 3$,经过卷积预处理、ResNet18 主干网络特征提取及 3 次反卷积后得到 $128 \times 128 \times 64$ 的特征图(128×128 为特征图尺寸,64 为特征图通道数,后文均以此种方式描述),将其送入 3 个预测头部得到 C 个类别的中心点热力图、2 张用来预测目标中心点偏置的特征图及 2 张用来预测目标框大小的特征图。网络结构如图 1 所示。

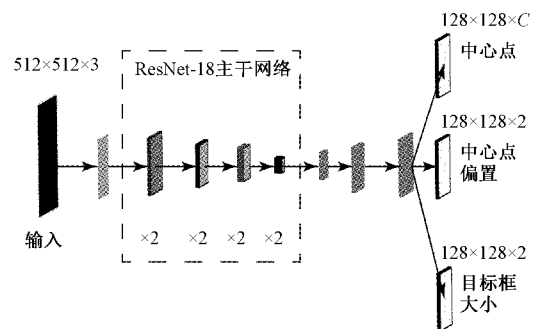


图 1 CenterNet-ResNet18 网络结构

2 改进的 CenterNet 算法

小目标所占像素面积小且所含语义特征少,其纹理特征难以被有效提取,但仅注重于选取特征提取能力较强的网络往往意味着更深的层数以及更多的计算资源消耗,从而影响算法实时性以及并增加实车部署的难度,针对此问题本文将首先使用较少的计算资源设计一个可较好平衡实时性与检测精度的高性能主干网络。

2.1 F-ResNeSt50 主干网络

ResNeSt 系列性能较强,且在不同下游任务中均体现出较好的泛化性,其在保持了 ResNet 最为代表性的残差连接的基础上,对内部残差块进行了较大的改进,提出了分裂注意力架构,增强网络的特征提取能力。而 ResNeSt 系列与 ResNet 相比参数量与计算量没有明显增加,检测速度也趋于一致,但预测精度提升约 3%。通过衡量 ResNeSt18、ResNeSt50 等网络的实时性及精度,本文使用 ResNeSt50 替换 ResNet-18 以平衡算法的精度与实时性,提升网络对不同目标特征提取能力。同时将 ResNeSt50 小残差块中 1×1 与 3×3 卷积后的 ReLU 激活函数升级为 FReLU,命

名为 F-ResNeSt50 网络。

FReLU 是专门面向视觉任务的激活函数,其通过添加一个“漏斗结构”对 ReLU 进行扩展使其具有空间像素级的建模能力如图 2 所示,其增强了非线性激活函数的空间依赖性,加强了对目标特征的捕获能力。同时其只增加了可忽略的计算开销,对实时性几乎无影响。FReLU 对输入端的激活方式如下:

$$y = \text{Max}(x_{s,i,j}, T(x_{s,i,j})) \quad (1)$$

式中: $x_{s,i,j}$ 为第 s 个通道 (i, j) 位置处激活函数的输入值, $T(x_{s,i,j})$ 则为本节“漏斗结构”,本质为一个空间上下文特征提取器,计算方法如下式所示:

$$T(x_{s,i,j}) = \mathbf{x}_{s,i,j}^a \cdot \mathbf{p}_s^a \quad (2)$$

式中: $\mathbf{x}_{s,i,j}^a$ 表示以 $x_{s,i,j}$ 为中心的参数池化窗口,此窗口大小本文设置为 3×3 , \mathbf{p}_s^a 表示该窗口参数,则同一通道的所有池化窗口 $\mathbf{x}_{s,i,j}^a$ 均共享同一参数。此 3×3 的池化窗口可增强不同特征的捕获能力,尤其对特征表达较弱的小目标改善良好;而额外的计算开销仅为加入的此池化窗口的开销,额外参数量为 $3 \times 3 \times C_{in}$, 计算量为 $3 \times 3 \times W' \times H' \times C_{in}$, C_{in} 为任一特征图输入通道数, W' 与 H' 为该特征图宽高。

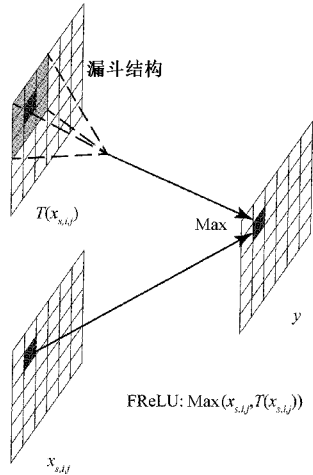


图 2 FReLU 激活函数

F-ResNeSt50 的每个小残差块如图 3 所示。其采用分组卷积的思想,将输入的通道数总计为 c 的 $h \times w$ 尺寸特征图均分为 k 个基组,同时每个基组中再次均分 r 个分裂单元(split),总分裂单元个数为 $G = k \times r$,多次分组可充分扩宽网络,增强网络的特征提取能力。随后对每个分裂单元首先进行 1×1 调整维度和 3×3 整合通道卷积操作,将此处两次卷积由 Conv+BN+ReLU(CBR)的组合替换为本文设计的 Conv+BN+FReLU(CBF)的组合。

此时每个分裂单元输出结果为:

$$U_i = \mathcal{F}_i(\mathbf{X}), t \in \{1, 2, \dots, G\} \quad (3)$$

式中: U_i 为每个分裂单元输出特征图, $\mathcal{F}_i(\mathbf{X})$ 为两次卷积过程的函数映射。

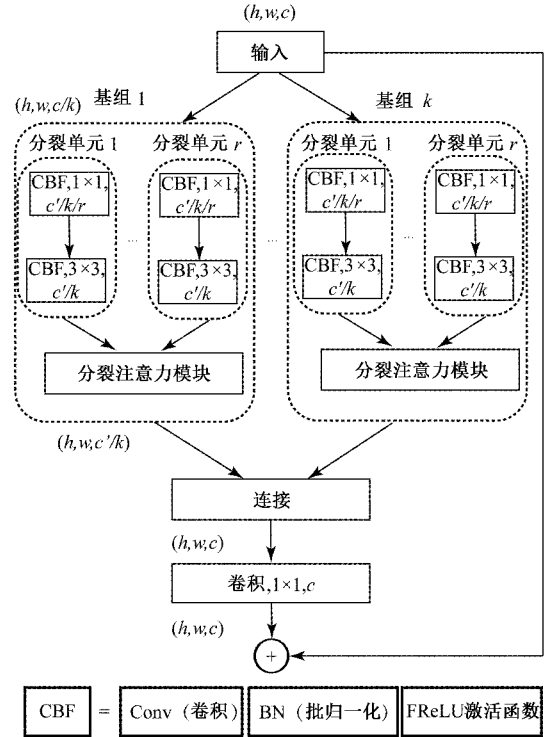


图 3 F-ResNeSt 小残差块网络结构

每个分裂注意力模块(split-attention module)首先引入 r 个分裂单元的输出特征图 U_i , 随后对特征图执行通道软注意力加权以得到此基组输出,分裂注意力模块可建模各个特征之间的重要性,对于不同的任务可根据输入进行特征分配,简单而有效。将 k 个基组的输出叠加(concat)并结合一个残差(shortcut)连接,得到一个 F-ResNeSt 的小残差块,堆叠多个小残差块堆叠得到一个大残差块。本文 F-ResNeSt50 采用“3、4、6、3”四个大残差块堆叠。

将 ResNet50 替换为 ResNeSt50 网络,并将 ReLU 升级为 FReLU 后算法复杂度变化如表 1 所示,发现 ResNeSt50 较 ResNet50 参数量仅增加 2 M, 计算量增加 1.4 GFLOPs,而本文提出的 F-ResNeSt50 较 ResNeSt50 增加可忽略。

表 1 不同主干网络复杂度对比

主干网络	激活函数	参数量/ M	计算量/ GFLOPs
ResNet50	ReLU	25.5	20.4
ResNeSt50	ReLU	27.5	21.8
F-ResNeSt50	FReLU	27.5	22.0

2.2 PASN 特征融合网络

CenterNet 设计时仅使用最深层特征图进行检测,小目标信息对浅层特征图有较强的依赖性,使用特征融合网络融合深、浅层语义信息可增强网络对小目标特征保留能力。

对比主流的特征提取网络发现:特征金字塔(FPN)被设计出来之后以其优秀的特征融合功能广受各大目标检测算法的追捧,PAN(pyramid attention networks)^[17]在 FPN 基础上加入了上采样过程,使深浅层不同尺度语义信息充分的融合,在保证网络深度的同时不会失去包含较多小目标信息的底层语义信息,传统的 PAN 对每个输入的特征层均采用同等方式处理,而小目标语义信息往往在较浅的特征层更加丰富,本节将对 PAN 进行重构,首先在较浅的 512 维度的特征层使用一个小目标特征增强模块 SPP,以引进较少参数的方式强化特征融合时的小目标语义丰富度;随后设定所有 3×3 的卷积为轻量化的深度可分离卷积,以大幅度降低算法参数量以及计算量。

首先对 SPP 小目标特征增强模块设计如图 4 所示。将上一节设计的 F-ResNeSt50 中第 2 个大残差块特征图 P_1 输出,通过 1×1 卷积整合通道至 $16 \times 16 \times 128$ 维度,连接 SPP 空间池化金字塔模块对输入的通道进行 3 次独立的最大池化以及一个残差连接操作,SPP 由于采用多次最大池化操作,本身不会引入额外的参数量,仅会引入池化过程中极少的计算量,对实时性几乎无影响,但多次池化能够极大地增加感受野,分离出最显著的上下文特征,增加对小目标的信息提取能力。3 次最大池化大小分别为 3×3 、 5×5 以及 9×9 ,将池化后的结果以及残差边进行叠加连接,重新拼接为 512 维度的特征层。

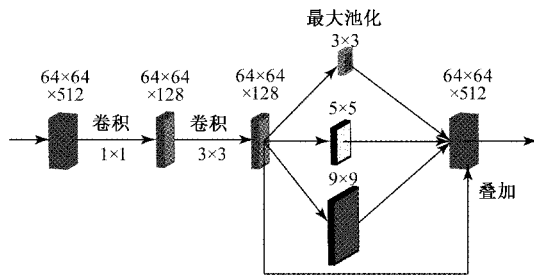


图 4 SPP 模块结构

考虑到简单的引入特征融合网络会造成参数量以及计算量的大量增加,同时观察卷积过程,可发现 5×5 的卷积块造成的计算资源消耗数倍于 3×3 与 1×1 的卷积块,针对这一问题,本节仅使用 1×1 与 3×3 的卷积并将 3×3 设定为深度可分离卷积。在 3×3 卷积操作中深度可分离卷积参数量及计算量约为普通卷积的 $\frac{1}{9}$,极大节省了计算资源的开销^[18]。

借鉴 PAN 前向传播与上、下采样方式,对本文特征融合网络结构设计如图 5 所示。将 P_1 特征图联合主网络 ResNeSt50 中第 3、4 大残差块输出的 2 个尺度特征图 P_2 、 P_3 ,同时输入特征融合网络,此网络命名为 PASN。所有卷积流程均为上文提出的 CBF 卷积块,即由 Conv 卷积操作、BN 归一化以及上文选取的 FReLU 激活函数组合而成,计算流程共分为 4 部分:SPP 特征加强、特征图降维输入、不同尺度特征图特征融合以及对融合结果整合输出。所有卷

积方式如下:

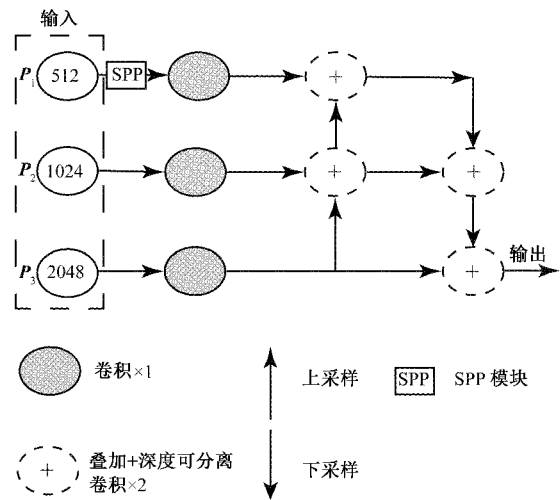


图 5 PASN 网络结构

- 1) 设定降维输入卷积为 1×1 普通卷积。
- 2) 设定通道整合卷积为两个 3×3 深度可分离卷积堆叠,输出通道数相同。
- 3) 设定 2 次上采样卷积为 1×1 的反卷积。
- 4) 设定 2 次下采样卷积为 1×1 普通卷积。

对比 PAN 与本文提出的 PASN 的算法复杂度如表 2 所示,可发现本文提出的特征融合网络参数量与计算量均低于原 PAN 网络,下降幅度约为 82.4% 与 79.1%。

表 2 不同特征融合网络复杂度对比

融合网络	参数量/M	计算量/GFLOPs
PAN	57.9	35.4
PASN	10.2	7.4

2.3 CenterNeSt 算法

CenterNeSt 算法输出端使用原解码方式经由 3 次反卷积整合信息进行预测,改进后全算法结构如图 6 所示。输入端采用 $512 \times 512 \times 3$ 的单帧图像,对输入信息使用 3 次卷积进行预处理;处理后的数据使用最大池化操作对数据整合,输入本文提出的 F-ResNeSt50 主干网络;已知 F-ResNeSt50 主干网络为 4 个大残差块堆叠的而成,各小残差块数量为 3、4、6 与 3,除第 1 个大残差块外每个大残差块之间尺寸均为 2 倍下采样倍率,通道数则为 $256 \sim 2048$ 成倍增加;将后 3 个大残差块输出的特征图输入本文设计的 PASN 特征融合模块,对 P_1 使用 SPP 模块以小目标特征增强,使深浅特征充分融合,最终融合后特征图大小为 $16 \times 16 \times 2048$;将融合后的特征图输入最后的解码头部 Center-Head 模块,操作为首先经过 3 次反卷积进行通道信息整合以及增大图像分辨率操作,每次上采样倍率为 2,最后将 $128 \times 128 \times 64$ 特征层输入 3 个 Center-Head 分别预测,得到中心点(key point)、中心点偏置(offset)与目标框

大小(size)。改进后的 CenterNeSt 算法损失函数上沿用了原 CenterNet 的计算方式,总的损失函数为 3 个损失的加权之和,分别为中心点的预测损失、目标中心的偏置损失以及目标的大小损失。中心点的预测损失应用变体的 Focal Loss^[19]损失,偏置损失以及大小损失则应用损失 L_1 来计算。

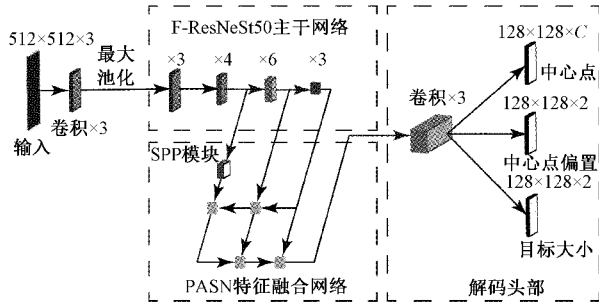


图 6 CenterNeSt 网络结构

3 实验与结果分析

3.1 数据集准备

本文选取自动驾驶专用公开数据集 Kitti 验证改进的 CenterNet 算法,Kitti 数据集中包含乡镇、市区和高速公路等采集的真实图像数据,每张图像中最多达 15 辆车和 30 个行人,还有各种程度的遮挡与截断。原始数据集中目标检测分为 8 类,分别为“car”、“truck”、“van”、“pedestrian”、“pedestrian(sitting)”、“cyclist”、“tram”、“misc”,在日常生活中,“tram”以及“misc”在常规道路难以见到。本文将类别重组,将“car”、“truck”、“van”合并为“car”类,将“pedestrian”、“pedestrian(sitting)”与“cyclist”合并为“person”类,“tram”与“misc”去掉,整个数据集则被分为 2 类:“car”、“person”。采用的全部图片为 Kitti 数据集中 7 481 张图片,所有图像均为彩色 png 格式,按照 8 : 1 : 1 划分为训练集、验证集与测试集,将 Kitti 数据集图片标注的 txt 格式转换为 VOC 数据集的 xml 格式,同时训练 CornerNet^[20]、CenterNet、CenterNeSt、Yolov3^[21]、Yolov4^[22]、Faster-rcnn^[23]以及 CenterNet-DHRNet。

3.2 算法训练

本文的所有训练均是在 Ubuntu 18.04 系统下进行,CPU 为 Intel Core i7-10700F,GPU 采用英伟达的 RTX 3070 独立显卡,采用 Pytorch 1.8.0 框架,安装 CUDA 11.1、CUDnn 8.0.4 以及 OpenCV 4.5 等扩展包。

训练过程中采用随机多尺度训练的方式,即每隔 10 个世代(epoch)随机变换一次输入图片的尺寸,输入图片大小为 256~416~618,使用随机多尺度训练可解决大、小样本不平衡的问题,使算法学习到更多的小目标特征;对所有输入图片采用 Mosaic 数据增强的方式;初始学习率 Learning rate(Lr)设定为 0.001,采用余弦退火学习率(cosine annealing)衰减方式,Batch size 设定为 8,优化器则采用的

Adam(adaptive moment estimation)进行梯度下降。

经过 220 个训练批次后,算法损失值已经完全收敛,如图 7 所示。在前 100 训练批次,训练集损失值(train Loss)与验证集损失值(val Loss)均快速下降,第 100 个批次的损失值大约为 6.0,随后收敛速度放缓,在第 200 个批次左右基本收敛完全,总体损失函数值最终收敛在 5.0 左右,算法模型参数已到达最优,选取训练过程中得到的最优权重作为最终权重文件。

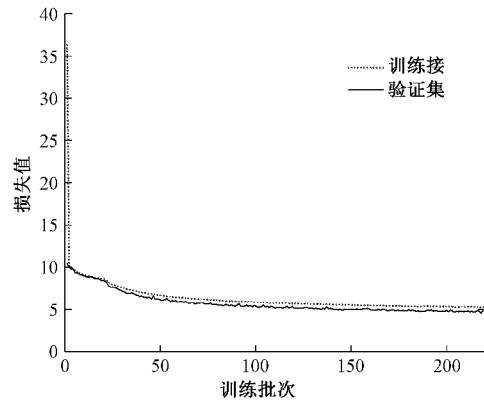


图 7 训练损失收敛

3.3 结果分析

本文使用每秒检测帧数(frames per second,FPS)来作为算法检测速度的评价指标,同时考虑到智能汽车算法设计需首先满足理论实时性,而车载视觉传感器频率多为 10~30 Hz,本文规定 FPS 大于 30 即为满足实时性(real-time)要求;使用 mAP、mAP₅₀ 以及针对小目标的检测精度 mAP_s 指标评价算法的准确性,如表 3 所示。

表 3 算法评价指标

指标	含义
mAP	IoU=0.50 : 0.05 : 0.95 时,平均检测精度
mAP ₅₀	IoU=0.50 时,平均检测精度
mAP _s	像素面积<315 时,小目标平均检测精度
FPS	每秒检测帧数
实时性	FPS>30

本文所有算法图像输入尺寸均设定为 512×512×3,在 Kitti 测试集上进行性能对比测试,当计算 mAP₅₀ 与 mAP_s 以及 mAP 时,将“car”与“person”两类合并并表示。测得不同算法检测速度及精度如表 4 所示,观察发现,改进后的 CenterNeSt 检测速度 FPS 可达到 37.7 满足实时性要求,平均检测精度 mAP 值原算法 CenterNet-ResNet18 为 46.4%,改进后的 CenterNeSt 达到 60.3%,提升 13.9%,小目标的检测精度 mAP_s 较原算法提升了 12.9%;对比主流实时目标检测算法 YOLOv4,本文提出的 CenterNeSt 检测精度较之高出 0.9%,且小目标检测精度高出 3.6%,检测

速度每秒也可多检测 1.1 帧;对比文献[11]中学者提出的 CenterNet-DHRNet 算法,本文算法 mAP₅₀ 检测精度高

2.5%,每秒可多检测 6.8 帧图片,性能更强。测试集算法表现证明本文算法具有有效性。

表 4 不同算法测试集表现

算法	主干网络	mAP/%	mAP ₅₀ /%	mAP _s /%	FPS	实时性
Faster-rcnn	ResNet101	53.7	81.2	38.2	12.5	×
CornerNet	Hourglass-104	56.1	84.2	40.5	10.7	×
Yolov3	DarkNet53	52.3	79.6	37.9	58.8	√
Yolov4	CSPDarkNet53	59.4	88.6	42.5	36.6	√
CenterNet	ResNet18	46.4	70.1	33.2	122.5	√
CenterNet	DLA-34	56.0	83.9	40.2	58.7	√
CenterNet	Hourglass-104	59.1	87.1	42.2	17.4	×
CenterNet-DHRNet	DHRNet	—	87.1	—	30.9	√
CenterNeSt	F-ResNeSt50	60.3	89.6	46.1	37.7	√

将 CenterNeSt 与原算法在 Kitti 测试集中检测结果进行可视化展示如图 8 所示,图中为 Kitti 测试集中挑选出的 4 个场景,分别为多车辆、多行人、光线较暗以及目标较小 4 种特殊情况,左侧为原算法识别图,右侧为改进后的 CenterNeSt。可以发现,第 1 个场景光线正常但较多车辆如图 8(a)、(b)所示,图 8(a)中原算法检测出 8 个车辆,而图 8(b)中改进后的算法较图 8(a)中原算法多识别出 7 个车辆,远处极小车辆目标也成功检出;第 2 个场景光线正常但行人密集如图 8(c)、(d)所示,图 8(c)仅检测出 2 个行人与 1 辆汽车,而图 8(d)中本文算法共计检出 6 个行人与 2 辆汽车,可知改进后的 CenterNeSt 算法多检出了 4 个行人和 1 个车辆,且图像左侧较小目标行人被成功检出;第 3 个场景光线较差如图 8(e)、(f)所示,2 个车辆高度重叠也增加了检测难度,图 8(e)中原算法仅检出 1 个车辆,图 8(f)中改进后的 CenterNeSt 多识别出了 1 个车辆;第 4 个场景如图 8(g)、(h)所示,所有车辆目标均较小,且相互重叠遮挡增大了检测难度,图 8(a)中原算法仅检测出 4 个目标相对较大的车辆,而图 8(b)中改进后的 CenterNeSt 识别出了 11 个,且最远处极小的 2 个车辆以及右下角仅被拍摄一小部分的汽车也成功检出。可知本文算法识别精度提高,对小目标的识别效果良好。

3.4 实车试验验证

为了验证所提出的 CenterNeSt 算法于真实智能汽车上的表现,本文将算法基于 TensorRT 进行模型量化后部署至嵌入式开发平台 NVIDIA Jetson TX2,并将 TX2 与实车工控机相连如图 9 所示。

实车试验场地选择在某汽车实验所实验场内,将车载单目相机视频流以 30 Hz 频率接入,实车保持 20 km/h 匀速向前行驶。使用本文部署后的 CenterNeSt 算法进行实车检测如图 10 所示,选取随机两帧检测结果如图 10(a)、(b)所示,观察可知图 10(a)中共计检测出 4 个车辆以及 1 个行人,且较远处目标很小的车辆也被检测出,可证明本文对

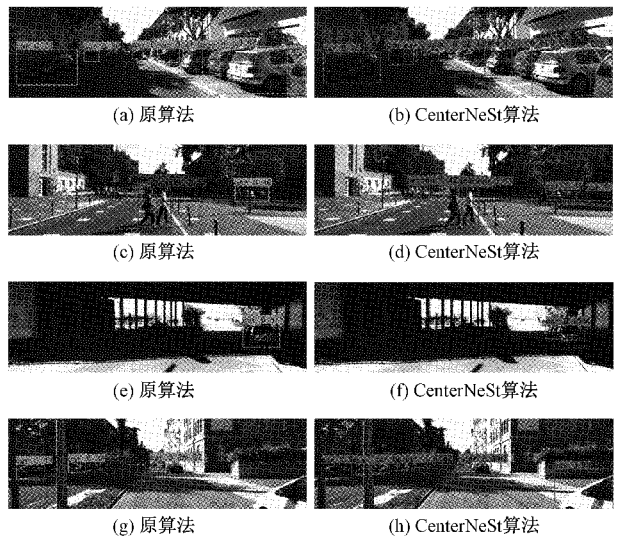


图 8 不同场景检测效果对比

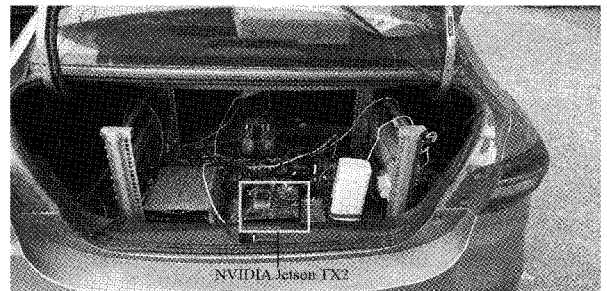


图 9 实车硬件设备

小目标检测效果良好;图 10(b)中右侧多个车辆高度重叠,加大了检测难度,本文算法仍成功检测出 9 个车辆,证明本文算法对重叠目标也有较好的分辨检测能力;实车平均检测速度可保持约 31 fps,满足车载实时性要求,证明本文所提出算法在实车上表现满足仿真预期,具有实际工程应用价值。

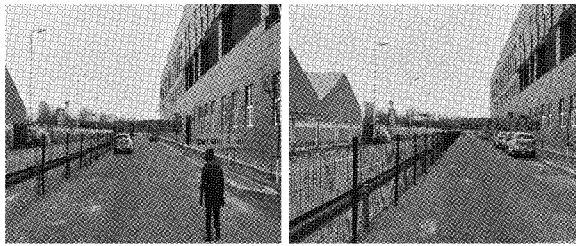


图 10 实车检测结果

对比试验证明了 CenterNeSt 算法的有效性,本文最后探究各改进部分对原算法的影响,以原算法 CenterNet-ResNet18 为基础算法进行消融实验,依次使用替换主网络

为 ResNeSt50、升级 FReLU 激活函数为 F-ResNeSt50 主干网络、加入 PAN 特征融合网络、将 PAN 替换为本文提出的 PASN 特征融合网络 4 个方法,测得每次网络结构变动后算法性能的变化如表 5 所示。观察可知,首先采用 ResNeSt50 主干网络后检测精度上升 7.8%,小目标检测精度提升 6.1%,为所有改进方式中精度提升最多,代价则为 FPS 牺牲了 51.3;采用 FReLU 激活函数后,平均检测精度上升 0.9%,小目标检测精度提升 0.8%,而 FPS 仅损失 2.4;加入 PAN 特征融合模块后,以牺牲 29.5 的 FPS 代价提升了平均检测精度 3.9%与小目标检测精度 3.4%;将 PAN 替换为本文的 PASN 后,FPS 牺牲 1.6 提升了平均检测精度 1.3%与小目标检测精度 1.6%。

表 5 消融实验

方法	ResNeSt50	F-ResNeSt50	PAN	PASN	mAP/%	mAP _s /%	FPS
CenterNet-ResNet18					46.4	33.2	122.5
方法 1	✓				54.2	39.3	71.2
方法 2		✓			55.1	40.1	68.8
方法 3		✓	✓		59.0	43.5	39.3
CenterNeSt(方法 4)		✓		✓	60.3	45.1	37.7

分析可知本文所使用方法均可平衡算法实时性与检测精度并增强算法综合性能,同时对小目标精度改善良好。

4 结 论

本文将 CenterNet 算法应用到自动驾驶场景目标检测领域中,重点解决算法检测时对小目标识别效果较差的问题,以增强行车的主动安全性。对 CenterNet 算法改进并命名为 CenterNeSt,针对原主干网络无法高效平衡实时性与精度的问题,提出了新型主干网络 F-ResNeSt50;针对原算法仅采用最深层特征图进行检测,浅层语义信息易丢失的问题,提出新型特征融合网络 PASN,提升了网络对不同尺度特征的纹理的保留能力并重点加强了小目标特征的表达。在满足实时性前提下,改进后的模型不仅优于原算法,而且性能强于主流算法 YOLOv4。实车实验也表明本文算法具有实际工程应用价值,可加速自动驾驶技术全面落地。

未来工作可通过其他手段重点增加算法对光线极暗的场景的检测能力,以达到较好的夜视效果,提升算法的鲁棒性。

参考文献

- [1] 宋宝玉,王波涛. 基于树莓派嵌入式平台的车道线检测算法[J]. 电子测量技术,2021,44(23):93-98.
- [2] 张昌凡,刘佳峰,何静,等. 基于改进卷积双向门控循环网络的轴承故障诊断[J]. 电子测量与仪器学报,2021,35(11):61-67.
- [3] 彭豪,李晓明. 基于改进 Faster R-CNN 的小目标检测模型[J]. 电子测量技术,2021,44(24):122-127.
- [4] 刘紫燕,袁磊,朱明成,等. 融合 SPP 和改进 FPN 的 YOLOv3 交通标志检测[J]. 计算机工程与应用,2021,57(7):164-170.
- [5] 董美琳,任安虎. 基于深度学习的高速公路交通事件检测研究[J]. 国外电子测量技术,2021,40(10):108-116.
- [6] 赵亚男,吴黎明,陈琦. 基于多尺度融合 SSD 的小目标检测算法[J]. 计算机工程,2020,46(1):247-254.
- [7] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9):1904-1916.
- [8] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [9] 孙备,左震,吴鹏,等. 面向无人艇环境感知的改进型 SSD 目标检测方法[J]. 仪器仪表学报,2021,42(9):52-61.
- [10] SINGH B, DAVIS L S. An analysis of scale invariance in object detection snip[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3578-3587.
- [11] 王新,李喆,张宏立. 一种迭代聚合的高分辨率网络

- Anchor-free 目标检测方法[J]. 北京航空航天大学学报, 2021, 47(12): 2533-2541.
- [12] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points[J]. ArXiv Preprint, 2019, ArXiv:1904.07850.
- [13] ZHANG H, WU C, ZHANG Z, et al. Resnest: split-attention networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 2736-2746.
- [14] MA N, ZHANG X, SUN J. Funnel activation for visual recognition [C]. European Conference on Computer Vision, Springer, Cham, 2020: 351-368.
- [15] 李科岑, 王晓强, 林浩, 等. 深度学习中的单阶段小目标检测方法综述[J]. 计算机科学与探索, 2022, 16(1): 41-58.
- [16] 魏玮, 杨茹, 朱叶. 改进 CenterNet 的遥感图像目标检测[J]. 计算机工程与应用, 2021, 57(6): 191-199.
- [17] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [18] 李一凡, 袁龙健, 王瑞. 基于 OpenPose 改进的轻量化人体动作识别模型[J]. 电子测量技术, 2022, 45(1): 89-95.
- [19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [20] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints [C]. Proceedings of the European conference on Computer Vision (ECCV), 2018: 734-750.
- [21] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv Preprint, 2018, ArXiv:1804.02767.
- [22] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv Preprint, 2020, ArXiv: 2004.10934.
- [23] 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.

作者简介

于方程, 硕士研究生, 主要研究方向为智能汽车感知算法。

E-mail: 525555046@qq.com

张小俊(通信作者), 博士生导师, 博士, 教授, 主要研究方向为人工智能、自动驾驶与特种机器人等。

E-mail: xjzhang@hebut.edu.cn

张明路, 博士生导师, 博士, 教授, 主要研究方向为机器人智能感知、智能决策与控制技术等。

E-mail: zhangml@hebut.edu.cn

赵天亮, 硕士研究生, 主要研究方向为智能汽车规划与决策算法。

E-mail: 896868906@qq.com