

DOI:10.19651/j.cnki.emt.2313979

面向自动驾驶的多任务环境感知算法

宋绍京¹ 陆婷婷² 孙翔² 龚玉梅¹ 陈建¹

(1.上海第二工业大学计算机与信息工程学院 上海 201209; 2.上海第二工业大学智能制造与控制工程学院 上海 201209)

摘要:为了解决复杂驾驶场景下目标检测精度较低而难以满足自动驾驶需求的问题,提出一种基于YOLOP的高效网络模型MEPNet。MEPNet可同时处理车辆检测、可行驶区域分割和车道线检测三项任务。首先,采用YOLOv7作为主体结构平衡精度与实时性;其次,设计了FRFB模块增大感受野,以增强网络的特征提取能力;并且提出在检测网络的头部添加小目标检测层,有效减轻车辆遮挡和重叠现象对识别结果的干扰;最后使用CARAFE作为上采样算子,精准定位的轮廓的同时更好地保留图片的语义信息。实验表明,该算法推理速度达到42.5 fps,对比基线YOLOP,车辆检测的mAP50和Recall分别提升了6.8%和6.3%,车道线检测的准确率和IoU分别提升了6%和1%,可行驶区域分割的mIoU达到92.5%,大幅度提升了性能,并且进一步设计了MEPNet-s,实现了四任务目标检测,亦满足自动驾驶所需的准确性和实时性。

关键词:复杂驾驶场景;目标检测;可行驶区域分割;车道线检测;小目标;四任务目标检测

中图分类号: TP183 **文献标识码:** A **国家标准学科分类代码:** 520.20

Multi-task environment perception algorithm for autonomous driving

Song Shaojing¹ Lu Tingting² Sun Xiang² Gong Yumei¹ Chen Jian¹

(1. School of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai 201209, China;

2. School of Intelligent Manufacturing and Control Engineering, Shanghai Polytechnic University, Shanghai 201209, China)

Abstract: In order to solve the problem of low object detection accuracy in complex driving scenarios which makes it difficult to meet the needs of autonomous driving, an efficient network model MEPNet based on YOLOP is proposed. MEPNet can simultaneously handle three tasks: vehicle detection, drivable area segmentation, and lane detection. First, YOLOv7 is used as the main structure to balance the accuracy and real-time performance. Second, the FRFB module is designed to enlarge receptive fields and enhance the feature extraction capability of the network. The proposed small object detection layer added to the head of the detection network effectively alleviates interference caused by vehicle occlusion and overlap. Finally, CARAFE is used as the upsampling operator to accurately locate object contours while preserving semantic information in images. Experimental results show that the algorithm achieves a inference speed of 42.5 fps, and compared with the baseline YOLOP, it improves the mAP50 and Recall of vehicle detection by 6.8% and 6.3%, the accuracy and IoU of lane detection by 6% and 1%, and the mIoU of drivable area segmentation reaches 92.5%, which significantly improves performance. Furthermore, MEPNet-s has been further designed to accomplish four-task object detection, while simultaneously meeting the accuracy and real-time requirements of autonomous driving.

Keywords: complex driving scenarios; object detection; drivable area segmentation; lane detection; small object; four-task object detection

0 引言

自动驾驶技术在全球范围内越来越受到重视,其中最主要的原因是它具有能够提高行车安全、减少交通拥堵等方面的优势。为了更好地感知和理解周围环境^[1],自动驾

驶系统采用了激光雷达、视觉传感器和毫米波雷达等传感器^[2],为后续的决策提供必要数据,从而帮助自动驾驶汽车做出准确且及时的决策。根据所做决策的任务不同,环境感知可以包括很多子任务:如目标检测、可行驶区域分割、车道线检测等。

准确的目标检测对于实现自动驾驶来讲至关重要。按照算法类型可以将目标检测分为传统算法^[3-4]和基于深度学习的算法^[5],由于传统算法对光照、亮度等因素较为敏感,加之存在运行速度缓慢等问题,逐渐被深度学习算法所取代。基于深度学习的目标检测算法主要分为两阶段和单阶段,两阶段方法需要先生成候选区域,再对候选区域进行分类和回归,代表算法有 Faster R-CNN^[6]和 Mask R-CNN^[7]等,这类算法精度高,但复杂度也高,速度较慢。单阶段方法在固定位置生成先验框之后进行位置回归及目标分类,速度大幅提升,代表算法有 SSD^[8]和 YOLO 系列^[9-14]等。2022 年,YOLOv7^[14]横空出世,它最大的优势是能够使精度、速度之间达到平衡。例如郑玉珩等^[15]提出的基于 YOLOv4 模型使得车辆检测变得轻量化。由于单阶段方法通常采用密集采样策略,因此在小尺寸物体或者高密度场景中,检测结果容易出现重叠或漏检等问题,定位精度相对较低。

Evan 首次将卷积神经网络引入语义分割领域,提出了全卷积网络 (fully convolutional networks, FCN)^[16],实现了像素级的语义分割,但其没有充分考虑像素与像素之间的关系,导致其在处理边界重叠的物体时表现不佳。在自动驾驶中一般使用语义分割的方法分割可行驶区域,这样能够加深视觉感知系统对场景的理解,帮助车辆保持在正确的车道内。PSPNet^[17]基于 FCN 设计的金字塔池化模块能够聚合不同区域的上下文信息,但也因为它采用了不同大小的池化核对特征图进行操作,所以其对输入尺寸比较敏感。Li 等^[18]基于多传感器融合提出了 BiFNet 来融合点云与图像,提供了道路可行驶区域。对于车道线而言,通常也是使用语义分割,例如 SCNN^[19]中的切片处理将特征信息以叠加的方式进行传递,有利于传递细长形的目标特征,可清晰预测出车道线。

尽管上述工作有效推进了相关领域的研究,但单一的环境感知任务功能受限,需图像分类、目标检测和图像分割等步骤融合完成。然而,简单串连会降低推理速度。因此学者们提出了多任务学习,可以平衡速度和精度。MultiNet^[20]、DLT-Net^[21]、YOLOP^[22]、MLRPNet^[23]、TDL-YOLO^[24]、RPNet^[25]、YOLOP-E^[26]等模型可同时处理多个任务。YOLOP^[22]能实现车辆检测、可行驶区域分割和车道线检测,并在满足实时性的条件下精度超过以往一些先进模型。而 MLRPNet^[23]、TDL-YOLO^[24]、RPNet^[25]与 YOLOP-E^[26]都是受到 YOLOP^[22]的启发进行了各自的改进,例如 MLRPNet^[23]在车道线解码器处采用 Transformer 机制来提取全局信息,TDL-YOLO^[24]利用交叉注意力模块得到分割和检测特征图,RPNet^[25]引入 RepVGG 训练过程的基本组成单元 (RepVGG block, RVB) 改进主体结构中的结构来提升模型精度,YOLOP-E^[26]利用轻量级骨干网络 EfficientNet-V2 进行重构加快运算速度,但是由于其使用的数据集中包含数量众多的小尺寸目标,并且采用的是密

集采样策略,因此面对复杂的交通环境,YOLOP 及其上述改进模型对于小目标检测结果均易出现重叠或漏检、定位精度相对较低等问题,例如远距离车辆检测。然而交通场景中的远距离车辆检测也是至关重要的。

针对以上问题,本文基于 YOLOP 进行优化设计,提出了一种改进的多任务环境感知网络模型 (multi-task environment perception network, MEPNet),该模型创新性的设计了四分支感受野模块 (four branches of receptive field block, FRFB)、小目标检测层,并结合轻量级的通用上采样算子模块 (content-aware reassembly of features, CARAFE)^[27],进行网络构建。在 BDD100K 数据集上的实验结果表明,本文的方法优于 YOLOP 算法。同时考虑到标识牌能够向驾驶员传递交通信息和警示信息,本文又在此基础上设计了 MEPNet-s,可同时进行车辆及标识牌检测,实现了四任务目标检测,大大提高行车安全性,亦满足准确性和实时性的要求。

1 模型介绍

MEPNet 的网络架构如图 1 所示,该模型以 YOLOv7 为主体结构,将下采样后的特征图送入 FRFB 模块,该模块可通过多尺度分支和残差连接的方式,有效增加网络对目标物体的感受野,同时提升网络模型对不同尺寸目标的特征表达能力。

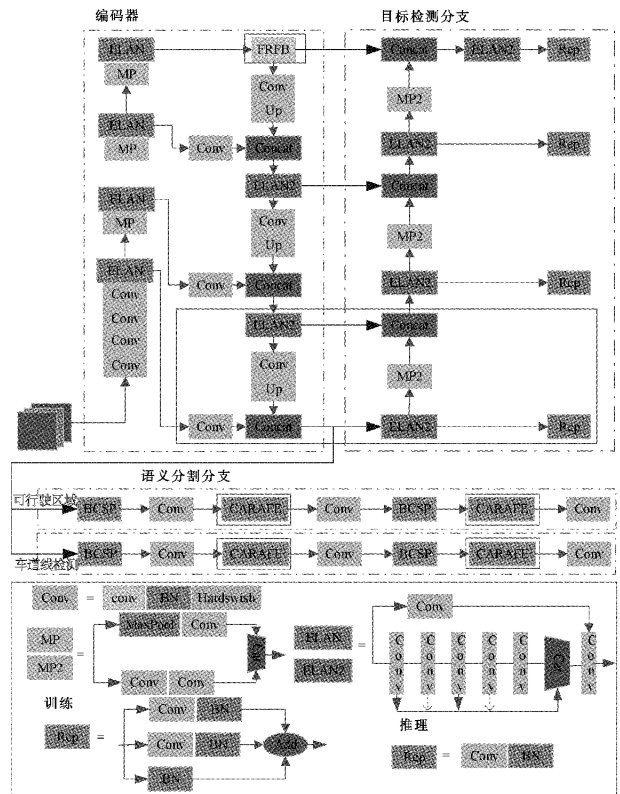


图 1 MEPNet 网络架构

对于目标检测分支,在预测层中添加了一个小目标检

测层,获得的 160×160 大小的特征图可检测到像素小于 8×8 的图像,改善了 YOLO 系列由于采用密集采样策略而易对小目标产生重叠或漏检等问题,有效提高了检测精度。

对于可行驶区域分割分支和车道线检测分支进行同样的操作,将编码器中被下采样成 $(W/4, H/4, 128)$ 的图用 CARAFE 进行上采样操作,该算子会根据输入特征来指导重组过程,在特征重组时拥有较大的感受野,更利于特征的表达。两次上采样之后变回原图大小,通道数变为 2,对是否为可行驶区域/车道线进行判断并将值返回。

1.1 FRFB 模块

本文在特征感受野模块 (receptive field block, RFB) 的基础上进一步探索效率更高、参数更少的网络,本文结合空洞卷积思想对其进行了优化,设计了 FRFB 模块,如图 2 所示。RFB 模块以模拟人类视觉的感受野来加强网络的特征提取能力,从而适应不同大小和比例的输入图像。

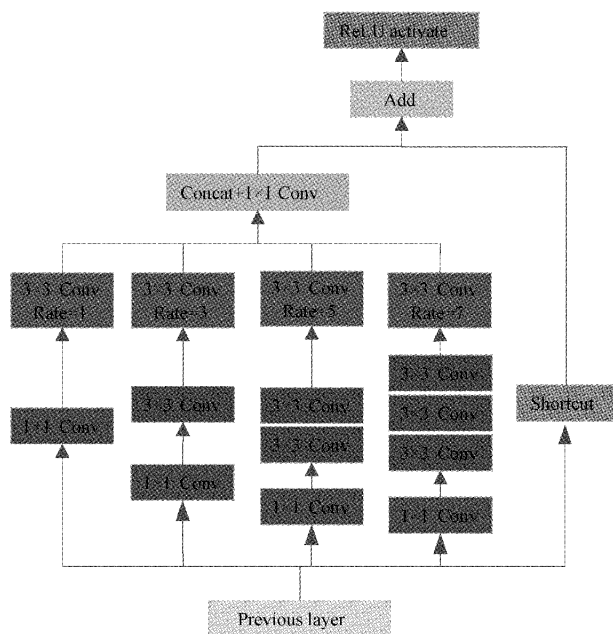


图 2 FRFB 结构

原始模块使用的空洞卷积的膨胀率只有 1、3、5,扩张尺寸范围有限,而较大的感受野可以提供更全局的上下文信息、提高目标检测和定位的准确性、增强特征表示的丰富性,并能够处理多尺度信息,从而提升卷积神经网络在各种计算机视觉任务中的性能和表现。同时空洞卷积能在不丢失分辨率的情况下扩大感受野并且调整膨胀率获得多尺度信息,因此为了让感受野变得更大,FRFB 先将膨胀率扩展到了 7,离心率增加的同时要增加视野大小,使用相应的 7×7 卷积核。为了减少计算量,同时考虑到堆叠的 3×3 卷积能提供更多数量的激活函数,增加了网络的非线性特性,另一方面卷积操作本身并没有破坏图像的空间信息,因此本文将 7×7 卷积替换成 3 个 3×3 卷积,并且也将膨胀

率为 5 的空洞卷积对应的 5×5 卷积替换成两个 3×3 卷积。具体来说,先通过 1×1 的卷积降维;其次,分别进行 1×1 、 3×3 、2 个 3×3 和 3 个 3×3 卷积,模拟不同尺度的感受野,不同大小的卷积核对应连接膨胀率为 1、3、5、7 的空洞卷积;最后,把各个分支的输出进行拼接,达到融合不同特征的目的,提升网络模型对不同尺寸目标的特征表达能力。此外,在 FRFB 模块中还采用了 ResNet 中的 Shortcut 连接方式,可以有效减轻梯度消失,使梯度能够很好地传递到浅层,降低网络的复杂度,提升网络的训练效果。FRFB 模块相比于 RFB 模块具有更大的感受野和对不同尺度物体更强的敏感性,能接触到的原始图像范围更大,因此蕴含更为全局、语义层次更高的特征,有效提高语义分割的准确性,轻量又高效。

1.2 小目标检测层

原模型提取多尺度特征进行目标检测时一共提取了 3 个特征层,特征图大小为 20×20 、 40×40 、 80×80 ,分别用于检测大小在 32×32 、 16×16 、 8×8 以上的目标。然而现实中交通场景复杂,对于远距离的车辆这种小目标容易出现误检或漏检、定位精度相对较低等问题。为了精确捕捉到小目标,本文提出增加小目标检测层对较浅特征图与深特征图拼接后进行检测,可大大降低小目标被误检或漏检的概率。本文在原有的 ELAN 模块之后继续进行上采样等处理,使得特征图继续扩大,扩大后的特征图大小为 160×160 ,能够进一步捕获物体的细节特征和空间位置特征,避免低分辨率的特征图中信息丢失问题可能会导致模型无法学到足够的细节信息。接着将获取到的大小为 160×160 的特征图与骨干网络中第 2 层特征图进行 Concat 操作,将较浅特征图与深特征图拼接可使得模型在保持高分辨率的同时拥有更多的语义信息,其中包含丰富和复杂的特征表达,能够适应图像中的不同物体尺度和姿态,提高了模型的鲁棒性和稳定性,使得模型在不同场景和环境都能够准确地检测目标。故对于本文设计的小目标检测层,如输入图片大小仍为 640×640 ,那么可以检测到像素为 4×4 的小目标,有效解决了小目标易被忽略等问题,大大提高了检测精度。对于新的 4 个对应不同尺寸的检测层,本文使用 K-means 聚类来代替人工设计的锚框尺寸,通过对训练集的真值框进行聚类,更具针对性地自动生成适合数据集车辆的锚框尺寸,进一步提高网络的检测精度,有助于提升自动驾驶安全性。

1.3 CARAFE

传统的上采样操作通常使用双线性插值等方式进行像素点的复制和补全,导致车道线分割图像的精度不高。而上采样算子 CARAFE 的设计思想是基于深度特征的多通道信息来指导重组过程,在重组时可以有较大的感受野。为了提供更好的特征表达,CARAFE 在不同位置使用不同的上采样核来进行特征的重组,这项操作在处理细小结构对象(例如线条、角落、曲线等)时尤为重要,一方面可以增

加特征图的分辨率以更准确地定位物体的轮廓,另一方面可以保留更多的语义信息以更好地分类每个像素所属的物体或区域。因此,使用 CARAFE 作为上采样算子可以提高可行驶区域与车道线检测的准确性和鲁棒性。

CARAFE 分为上采样核预测模块和特征重组模块两个主要模块。上采样核预测模块的主要作用是为后续的特征重组模块提供合适的上采样核。为了实现算法的轻量化,先对形状为 $H \times W \times C$ 的输入特征图进行操作,使用 1×1 的卷积将其通道数压缩到 C_m ,接着假设上采样核是 $k_{up} \times k_{up}$,并且上采样比是 σ ,利用 $k_{encoder} \times k_{encoder}$ 的卷积层来预测上采样核,此时输出通道数变为 $\sigma^2 k_{up}^2$,然后将通道维在空间维展开,得到形状为 $\sigma H \times \sigma H \times k_{up}^2$ 的上采样核,最后为了使得重组核权重和为 1,对上一步得到的上采样核进行归一化操作,这里用的是 Softmax。那么对于 X' 的每个位置 l' ,都有 k_{up}^2 个参数对应,用于重组特征权重。特征重组模块用于预测重组核的特征,实现对特征图的精准恢复和重构。对于输出特征图中的每个位置,将其映射回输入特征图,取出以它为中心的 $k \times k$ 区域和预测出的该点的上采样核进行点积操作,相同位置的不同通道共享同一个上采样核。最终得到 $\sigma H \times \sigma H \times C$ 的新特征值。具体结构如图 3 所示。对于目标位置 l' ,对应以 $l = (i, j)$ 为中心的放行区域 $N(\mu_l, k_{up})$,计算方法如下:

$$X'_{l'} = \sum_{n=-r}^r \sum_{m=-r}^r W_{l'(n,m)} \cdot X_{(i|n,j|m)} \quad (1)$$

其中, X' 表示新特征图; $W_{l'}$ 表示重组核; $r = k_{up}/2$ 。在本实验中,本文的参数设置如下: $\sigma = 2, k_{encoder} = 3, k_{up} = 5$,这能使此模块起作用的同时又不增加过多的计算量。

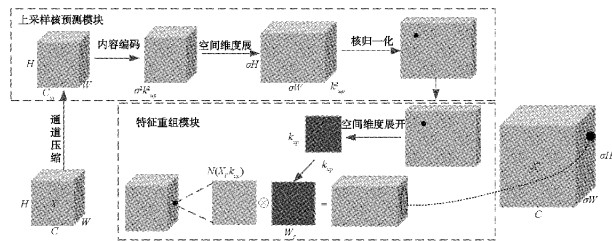


图 3 CARAFE 网络结构

1.4 Focal Loss

Focal Loss 是一种应用于目标检测任务的损失函数,可解决单类别样本数量远大于其他类别的类别不平衡问题。通常情况下,目标检测中使用交叉熵损失函数来训练模型,然而,当检测小目标时,由于其目标较小,可能只覆盖了检测特征图上的几个网格单元,导致很难与大目标区分开来,难以正确预测。此外,交叉熵损失函数对于样本分类错误的惩罚是固定的,无法区分哪些样本更有价值,这会容易导致在处理类别不平衡和难易样本权重不一致的情况下,容易出现过拟合或欠拟合等问题。为了解决这些问题,Focal Loss 在交叉熵损失函数的基础上添加了一个缩放因子 γ ,该缩放因子用于控制样本数量不平衡对损失函数的影响程

度。由于缩放因子的存在,Focal Loss 具有下降得更快、聚焦在难以分类的样本上的性质,使模型更专注于学习困难的样本,从而提高模型性能。公式如下:

$$FL(p_i) = -(1 - p_i)^\gamma \log(p_i) \quad (2)$$

其中, p_i 表示模型预测结果为正样本的概率; γ 为可调参数,用于调整难分类样本对损失函数的影响程度。当 $\gamma = 0$ 时,Focal Loss 退化为普通的交叉熵损失函数,本文设其为 2。

2 实验与分析

2.1 实验环境及细节

本文的实验环境是基于 Ubuntu18.04 操作系统,使用的是显存为 24 G 的 NVIDIA GeForce RTX 3090 显卡。深度学习框架为 Pytorch,版本号是 1.10.1; CUDA10.1; Python3.9.7。本文的所有实验都未使用预训练权重,优化器选择 Adam, batch size 取 12,初始学习率设置为 0.001,权重衰减以及动量衰减分别是 0.0005 和 0.937,并且前 3 轮采用训练热身的方法能使得训练出来的模型收敛速度变快,同时效果更佳。

2.2 数据集

本文实验采用 BDD100k 作为数据集,它是迄今为止规模最大、最多样的自动驾驶数据集之一。该数据集中有 10 万张已标注的图片,这些照片涵盖不同时间、不同天气条件和真实驾驶场景,并且其中包含了图像标记、可行驶区域、车道标记线等信息。

2.3 实验结果

1) 比对实验

本文对比了当前先进的目标检测算法,同时采用 mAP50 和召回率 Recall 作为检测精度的评价指标。由表 1 可知,MEPNet 的检测精度远远超越先前的网络,达到了 83.3% 的 mAP50 和 95.5% 的 Recall,这主要归功于 MEPNet 增加了小目标检测层,不仅可以检测到小目标,而且使得模型具有鲁棒性,能在夜晚灯光不足、光线反射、多径干扰背景下有效检测。

表 1 目标检测结果对比

网络	mAP50/%	召回率/%
MultiNet	60.2	81.3
DLT-Net	68.4	89.4
Faster R-CNN	64.9	81.2
YOLOv5s	77.2	86.8
YOLOP	76.5	89.2
TDL-YOLO	78.0	88.6
RPNet	78.3	89.8
YOLOP-E	79.2	—
MEPNet	83.3	95.5

本文的模型将图像区分为可行驶区域或背景,其中PSPNet是单语义分割网络。同时利用mIoU来评价不同模型的分割性能。由表2可知,MEPNet比MultiNet、DLT-Net、PSPNet、基线YOLOP以及基于YOLOP改进后的MLRPNet、TDL-YOLO和RPNNet的性能分别高出20.9%、21.2%、2.9%、1%、1.2%、1.1%和1.2%,表现最优。

表2 可行驶区域分割结果对比

网络	可行驶区域 mIoU/%
MultiNet	71.6
DLT-Net	71.3
PSPNet	89.6
YOLOP	91.5
MLRPNet	91.3
TDL-YOLO	91.4
RPNNet	92.3
MEPNet	92.5

在表3列出的这5个网络中,前3者都为单任务网络,仅检测车道线。由此可知,MEPNet不仅可以端到端训练,并且结果比单任务网络更好。准确率方面分别提高

42.38%、40.71%、39.94%、6%、7.83%、4.2%以及相比较与YOLOP-E提升3.1%,IoU方面对比基线YOLOP提升1%,均领先于其余先前的网络。

表3 车道线检测结果对比

网络	准确率/%	车道线 IoU/%
ENet	34.12	14.64
SCNN	35.79	15.84
ENet-SAD	36.56	16.02
YOLOP	70.50	26.20
MLRPNet	68.67	—
TDL-YOLO	72.30	26.50
RPNNet	—	27.20
YOLOP-E	73.40	—
MEPNet	76.50	27.20

2) 消融实验

为了验证本文提出的模块的有效性,进行了消融实验。在实验中,通过对模块进行了逐步调整,观察不同配置下的性能变化,以便更好地理解模块各参数的影响程度。具体结果如表4所示。

表4 消融实验结果对比

网络	速度/fps	mAP50/%	召回率/%	mIoU/%	准确率/%	IoU/%
YOLOP	61.3	76.5	89.2	91.5	70.5	26.2
v7-YOLOP	53.1	79.1	90.4	91.7	72.1	26.0
v7+FRFB	51.5	79.4	90.0	92.3	73.7	27.1
v7+CARAFE	42.1	79.5	90.5	92.2	75.3	26.6
v7+小检测层	45.9	83.5	95.2	92.4	72.5	27.2
v7+Focal loss	56.3	80.0	91.5	92.6	76.2	26.3
MEPNet	42.5	83.3	95.5	92.5	76.5	27.2

以上所有消融实验均是在以YOLOv7为主体结构的条件下单独进行的,各模块之间没有叠加,为了验证模块的有效性,后续添加的模块均与v7-YOLOP相比。首先,考虑到YOLOv7具有高效、准确、易于优化和可扩展性等优势,本文将替代基线YOLOP中的主体结构,速度略微下降但是车辆检测精度均上升,其中mAP50提升3.6%,并且其对车道线的精确度也有帮助。在v7-YOLOP基础上分别添加FRFB模块和采用Focal Loss损失函数来训练模型,分别增加了车道线检测1.1%的IoU和5.2%的准确率。将分割分支中的上采样改为CARAFE算子,有效提升两项分割任务的性能。通过添加了小目标检测层,mAP50和Recall分别提升4.4%和4.8%。最后将所有改进融合,实验结果表明本文提出的网络MEPNet对比基线YOLOP,速度达到42.5 fps的情况下,目标检测的mAP50和Recall分别提升6.8%和6.3%,可行驶区域分割的

mIoU提升1%,车道线检测准确性和IoU提升6%和1%。

3) MEPNet-s

在自动驾驶中,标识牌作为智能交通系统与先进辅助驾驶系统的重要组成部分,包括限速标志、禁止标志等,它们提供了必要的交通信息和规则,帮助驾驶员理解和掌握道路情况,降低了驾驶员不遵守停车标志等交通法规的可能,避免了违法左转或者其他交通违法行为,从而实现更加安全和高效的驾驶。鉴于其重要性,本文进一步设计了MEPNet-s,该模型在MEPNet的基础上增添了标识牌的检测,实现了四任务目标检测。首先本文调整了网络架构,将输出层的类别预测个数增加到多个类别数量,并添加相应的类别嵌入层,同时还修改了损失函数,以处理多个类别的目标,并使用目标分配策略将真实标签分配给相应的边界框,最后使用BDD100K中的测试集评估训练模型的性能。由表5可知,MEPNet-s的性能相比于

MEPNet 在可行驶区域分割以及车道线检测方面并没有太大的变化, mIoU 达到 92.3%, 但是由于多加了标识牌检

测, 不同类别之间的样本数量存在不平衡, 并且多类别的存在会导致类别之间的相互干扰, mAP50 略微下降。

表 5 MEPNet-s 实验结果

网络	速度/fps	mAP50/%	召回率/%	mIoU/%	准确率/%	IoU/%
MEPNet-s	38.2	80.5	94.8	92.3	76.2	27.1

MEPNet-s 模型在速度和 mAP50 性能略微下降的情况下能够同时进行车辆检测、标识牌检测、可行驶区域分割和车道线检测这四项任务, 很大程度上节省了系统开销, 利用有限的资源完成了多种感知任务, 为自动驾驶感知提供一体化思路。

4) 效果对比

图 4、5 显示了原图以及在 BDD100K 数据集上 YOLOP 和本文的 MEPNet 的可视化比较, 其中绿色区域表示可行驶区域, 红线表示车道线, 框表示检测到的车辆。图 4 显示了昼间的结果。图 5 显示了夜间的结果。

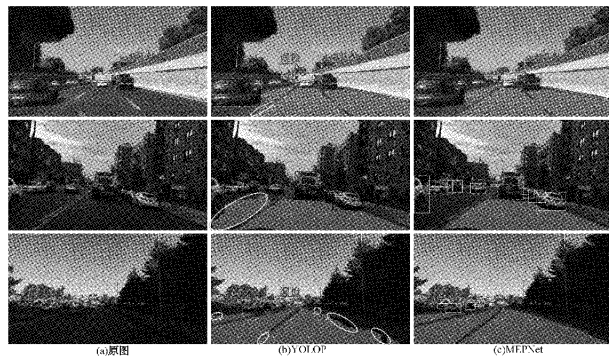


图 4 昼间效果图

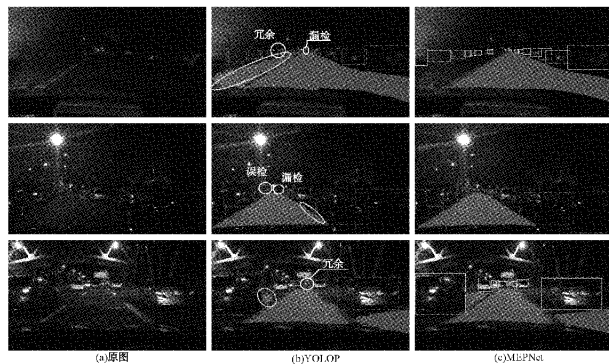


图 5 夜间效果图

图 4(b) 展示了昼间 YOLOP 可视化结果, 第 1 个场景的车道线显示不全, 有缺失, 且高架上有 1 辆车, 对小目标出现了错误检测, 同时将栏杆误检成车辆; 第 2 个场景的对向车道存在一些虚假的可行驶区域; 第 3 个场景的车道线不连续, 且部分可行驶区域缺失, 同时漏检了栏杆外的车辆。

图 5 则是夜晚可视化结果。对于 YOLOP 来说, 第 1

个场景夜晚时没分辨出对向车道从而错误识别大段可行驶区域, 同时对于左侧车道的小型车辆存在一些误检和冗余检测, 漏检了前方远处的车辆; 第 2 个场景将对面的建筑物误检成车辆, 又将正前方的两辆车漏检了一辆, 并且对于右侧的可行驶区域边界划分的不平滑; 第 3 个场景未分割出左侧一大段空路间的可行驶区域, 同时将小型车辆重复检测。

可视化结果表明, 昼间环境呈现出良好的效果, 即使面对夜晚灯光不足、光线反射、多径干扰背景下的交通环境, 本文的模型 MEPNet 能很好的划分可行驶区域, 并且针对远距离车辆目标尺寸较小、遮挡严重等问题, 做出的改进取得了明显的改善, 拥有更好的性能。

根据图 6 所示, 不论是在昼夜场景下, MEPNet-s 都能准确地划分道路的可行驶区域, 并完整地标出车道线。此外, 该模型还具备对远处车辆的有效检测以及同时检测车辆加标识牌的能力。这些结果表明 MEPNet-s 在各种光照条件下都展现出了优秀的场景理解和物体检测性能, 为自动驾驶领域提供了可靠的解决方案。



图 6 MEPNet-s 效果图

3 结 论

本文基于 YOLOP 算法做出改进, 设计了一种高效的多任务网络模型 MEPNet。通过采用 YOLOv7 作为主体结构, 有效平衡了准确性、实时性和鲁棒性之间的关系; 并且设计了 FRFB 模块, 有效提取特征; 接着增加了小目标检测层, 减少远处车辆较小存在漏检的现象; 最后, 利用 CARAFE 作为上采样算子, 提供更好的特征表达和有用信息的保留。实验表明, 该算法对比基线 YOLOP, 在各任务分支上都有明显的改善, 能够基本满足自动驾驶所需的准确性和实时性, 并且进一步设计的 MEPNet-s, 增添了标识牌检测功能, 实现了四任务目标检测。此外, 这项工作也可以为其他研究者提供不同网络的设计思路, 有利于创建更高效的网络结构。在下一步的工作中, 将更多类别纳入检测范畴, 为自动驾驶系统带来更加全面和可靠的场景感知能力。

参考文献

- [1] 韩胜明,肖芳,程纬森. 深度强化学习在自动驾驶系统中的应用综述[J]. 西华大学学报(自然科学版), 2023, 11(2): 1-7.
- [2] 张新钰,高洪波,赵建辉,周沫. 基于深度学习的自动驾驶技术综述[J]. 清华大学学报(自然科学版), 2018, 58(4): 438-444.
- [3] NISHIMURA J, KURODA T. Versatile recognition using haar-like feature and cascaded classifier [J]. IEEE Sensors Journal, 2010, 10(5): 942-951.
- [4] WU S, NAGAHASHI H. Parameterized AdaBoost: Introducing a parameter to speed up the training of real AdaBoost[J]. IEEE Signal Processing Letters, 2014, 21(6): 687-691.
- [5] 杨锋,丁之桐,邢蒙蒙,等. 深度学习的目标检测算法改进综述[J]. 计算机工程与应用, 2023: 1-17.
- [6] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Proceeding of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 21-37.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2016: 779-788.
- [10] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2017: 7263-7271.
- [11] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. ArXiv Preprint, 2020, ArXiv:2004.10934.
- [13] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv Preprint, 2022, ArXiv:2209.02976.
- [14] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [J]. ArXiv Preprint, 2022, ArXiv:2207.02696.
- [15] 郑玉珩,黄德启. 改进 MobileViT 与 YOLOv4 的轻量化车辆检测网络[J]. 电子测量技术, 2023, 46(2): 175-183.
- [16] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2015: 3431-3440.
- [17] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2017: 2881-2890.
- [18] LI H, CHEN Y, ZHANG Q, et al. BiFNet: Bidirectional fusion network for road segmentation[J]. Institute of Electrical and Electronics Engineers (IEEE), 2021: 8617-8628.
- [19] PARASHAR A, RHU M, MUKKARA A, et al. SCNN: An accelerator for compressed-sparse convolutional neural networks[J]. ACM SIGARCH computer architecture news, 2017, 45(2): 27-40.
- [20] TEICHMANN M, WEBER M, ZOELLNER M, et al. Multinet: Real-time joint semantic reasoning for autonomous driving [C]. 2018 IEEE Intelligent Vehicles Symposium(IV), 2018: 1013-1020.
- [21] QIAN Y, DOLAN J M, YANG M. DLT-Net: Joint detection of drivable areas, lane lines, and traffic objects [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(11): 4670-4679.
- [22] WU D, LIAO M W, ZHANG W T, et al. YOLOp: You only look once for panoptic driving perception[J]. Machine Intelligence Research, 2022, 19(6): 550-562.
- [23] 赵旋,黄崇栋,刘云涛等. 复杂交通环境下基于多任务学习的道路感知方法[J/OL]. 无线电工程:1-8[2023-09-02]. <http://kns.cnki.net/kcms/detail/13.1097.TN.20230815.1656.002.html>.
- [24] 牛国臣,王晓楠. 基于交叉注意力的多任务交通场景检测模型[J/OL]. 北京航空航天大学学报:1-12[2023-09-02]. <http://10.13700/j.bh.1001-5965.2022.0610>.
- [25] 宫保国,陶兆胜,赵瑞,等. 基于改进 YOLOv5s 的道路场景多任务感知算法[J]. 齐齐哈尔大学学报(自然科学版), 2023, 39(3): 19-29.
- [26] LIU Y, LI G, HAO L, et al. Research on a lightweight panoramic perception algorithm for electric autonomous mini-buses [J]. World Electric Vehicle Journal, 2023, 14(7): 179.
- [27] WANG J, CHEN K, XU R, et al. Carafe: Content-aware reassembly of features[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.

作者简介

宋绍京,博士,教授,主要研究方向为智能驾驶、图像处理。

E-mail: sjsong@sspu.edu.cn

陆婷婷,硕士研究生,主要研究方向为智能驾驶。

E-mail: lutingting1103@163.com

龚玉梅(通信作者),博士,讲师,主要研究方向为信号与信息处理。

E-mail: ymgong@sspu.edu.cn