

基于YOLOP-L的多特征融合道路全景驾驶检测

吕嘉璐, 周力, 巨永锋

引用本文

吕嘉璐, 周力, 巨永锋. 基于YOLOP-L的多特征融合道路全景驾驶检测[J]. 计算机科学, 2024, 51(6A): 230700185-8.

LYU Jialu, ZHOU Li, JU Yongfeng. Multi Feature Fusion for Road Panoramic Driving Detection Based on YOLOP-L [J]. Computer Science, 2024, 51(6A): 230700185-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[RM-RT²NI:融合评论时效与可信近邻影响力的推荐模型](#)

RM-RT²NI:A Recommendation Model with Review Timeliness and Trusted Neighbor Influence
计算机科学, 2024, 51(6A): 230800160-7. <https://doi.org/10.11896/jsjcx.230800160>

[基于多尺度视觉感知特征融合的显著目标检测方法](#)

Salient Object Detection Method Based on Multi-scale Visual Perception Feature Fusion
计算机科学, 2024, 51(5): 143-150. <https://doi.org/10.11896/jsjcx.230100132>

[基于边缘智能感知的无人机空间航迹规划方法](#)

Edge Intelligent Sensing Based UAV Space Trajectory Planning Method
计算机科学, 2023, 50(9): 311-317. <https://doi.org/10.11896/jsjcx.220800032>

[基于边缘智能的频谱地图构建与分发方法](#)

Construction and Distribution Method of REM Based on Edge Intelligence
计算机科学, 2022, 49(9): 236-241. <https://doi.org/10.11896/jsjcx.220400148>

[SDFA:基于多特征融合的船舶轨迹聚类方法研究](#)

SDFA:Study on Ship Trajectory Clustering Method Based on Multi-feature Fusion
计算机科学, 2022, 49(6A): 256-260. <https://doi.org/10.11896/jsjcx.211100253>

基于YOLOP-L的多特征融合道路全景驾驶检测

吕嘉璐 周力 巨永锋

长安大学电子与控制工程学院 西安 710064

(ll510102@163.com)

摘要 目前,驾驶员视角下的交通图像检测技术成为交通领域的重要研究方向,同时提取车辆、道路、交通标志等多种特征已经成为驾驶员理解道路信息多样性的亟需任务。以往研究已在单类目标检测的特征提取方面取得了长足进步,然而,这些研究不能很好地联合应用于其他区别较大的特征检测任务中,且融合训练过程中会损失个别特征检测的精度。针对驾驶员视野范围内道路信息多样且复杂的特点,本文提出了一种基于多特征融合训练的检测模型YOLOP-L,它能够同时对多种不同特征交通目标进行融合训练,同时保证单项检测任务的精度。首先,为了解决特征融合中语义信息表达不完整的问题,设计的SP-LNet模块通过FPN与双向特征网络结合实现网络更深层次的融合,使得提取的信息更完整,从而提升道路小目标的检测性能;其次,设计新的分割头深度可分离卷积,将语义信息与局部特征融合促使多特征融合的训练准确度与速度得到进一步提升;再次,体系中设计的GDL-Focal多类混合损失函数更专注于困难样本,可用于解决样本特征不平衡的问题。最后,对比实验表明:YOLOP-L相比原YOLOP网络运行的速度更快;在车辆目标检测任务下召回率提升了2.2%;在车道线检测任务下准确率提升2.8%,车道线IoU的值较HybridNets网络下降2.45%,但较YOLOP-L网络提升1.95%;在可行驶区域分割任务下其整体检测性能提升1.1%。结果表明,在具有挑战性的BDD100K数据集上,YOLOP-L可以在复杂场景下有效解决检测精度不足和分割缺失的问题,提高了车辆识别、车道线检测以及道路行驶区域联合训练的准确性和鲁棒性。

关键词: 全景驾驶;多特征融合;车辆检测;可行驶区域检测;车道线检测;双向特征金字塔

中图分类号 TP391

Multi Feature Fusion for Road Panoramic Driving Detection Based on YOLOP-L

LYU Jialu, ZHOU Li and JU Yongfeng

School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China

Abstract In recent years, traffic image detection technology from the driver's perspective has become an important research direction in the field of transportation, and extracting various features such as vehicles, roads, and traffic signs has become an urgent task for drivers to understand the diversity of road information. Previous studies have made significant progress in feature extraction for single class object detection. However, these studies cannot be well applied to other feature detection with significant differences, and the accuracy of individual feature detection will be lost during fusion training. In response to the diverse and complex road information within the driver's field of view, this paper proposes a detection model YOLOP-L based on multi feature fusion training. It can simultaneously fuse and train multiple different feature traffic targets, while ensuring the accuracy of individual detection tasks. The results indicate that YOLOP-L can effectively solve the problems of insufficient detection accuracy and missing segmentation in complex scenes on the challenging BDD100K dataset, improving the accuracy and robustness of vehicle recognition, lane line detection, and joint training of road driving areas. Finally, comparative experiments show that YOLOP-L runs faster than the original YOLOP network. The recall rate increases by 2.2% under the vehicle target detection task. In the lane detection task, the accuracy improves by 2.8%, and the IoU value of the lane line decreases by 2.45% compared to the HybridNets network, but increases by 1.95% compared to the YOLOP-L network. Its overall detection performance improves by 1.1% under the task of driving area segmentation. The results indicate that YOLOP-L can effectively solve the problems of insufficient detection accuracy and missing segmentation in complex scenes on the challenging BDD100K dataset, improving the accuracy and robustness of vehicle recognition, lane line detection, and joint training of road driving areas.

Keywords Panoramic driving, Multi feature fusion, Vehicle inspection, Travelable area detection, Lane line detection, Bidirectional feature pyramid network

1 引言

近年来,根据世界卫生组织(WHO)发布的一份报告,每年约有130万人于道路交通事故中发生不幸^[1]。由于驾驶员

需要长时间行驶在正确的道路区域,并且必须与前车保持适当的间距,因而道路信息的检测与获取显得尤为重要。如何避免安全隐患,保证安全驾驶,已成为科研人员研究的重点。然而,尽管计算机视觉和深度学习的方法在单项交通检测任

务中已经取得了巨大进展,但驾驶员视野下同时进行车辆识别、道路区域与车道线检测的多特征融合训练仍然是一项挑战^[2]。

自动驾驶通过感知系统获取道路信息,依靠决策系统分析信息、规划行驶路径,最后利用控制系统保证车辆的安全行驶。该技术的应用已经吸引了学术界以及行业研究人员的广泛关注。为了确保汽车在行驶过程中的安全,自动驾驶系统需要使汽车沿着道路上的车道行驶,这就需要准确的车道感知。车辆检测在提供交通物体的位置和大小信息、帮助自动驾驶汽车在驾驶阶段做出准确及时的决策方面发挥着重要作用。此外,可行驶区域段和车道段为路线规划和提高行车安全提供了丰富的信息^[3]。在这种情况下,多特征任务学习网络为节省计算成本提供了一种潜在的解决方案,其中网络被设计成编码器-解码器模式,并且编码器被不同的任务有效地共享^[4]。目前针对道路行驶的特征学习任务的研究主要分为以下3种:基于深度学习的车辆检测的研究、基于深度学习的车道线检测研究,以及可行驶的道路区域检测研究。

(1) 基于深度学习的车辆检测研究

随着深度学习的快速发展,已经出现了许多性能显著的车辆检测算法。目前主流的目标检测算法可以分为两阶段方法和单阶段方法。两阶段方法分两步完成检测任务。首先获得区域建议,区域提案的产生经历了多个发展阶段^[5-6],然后利用区域建议中的特征对对象进行定位和分类。Ross Girshick 提出了 Region-CNN(R-CNN),使用选择搜索^[9]算法替代了滑动窗口,解决了窗口冗余的问题,降低了算法时间复杂度,然而也存在不少问题,例如特征提取较为耗时,需要固定尺度的输入,图像信息容易丢失。为了解决 R-CNN 的问题,He 等提出了 Spatial Pyramid Pooling(SPPNet)^[10],该方法不需要输入固定大小的图像,减少了图像信息的丢失,并且避免了重复计算卷积特征。

SSD^[11]和 YOLO 系列算法是单阶段方法中的里程碑,通过该算法只需一次提取特征即可实现目标检测,使其成为真正的端到端网络,这种算法同时执行边界框回归和对象分类。当 Feature Pyramid Network(FPN)出现时,它最初通过提供自上而下的途径来从语义丰富的层构建更高分辨率的层,从而支持 Ma 等^[12]采用的 Region Proposal Network(RPN)方法。而 YOLO^[13]将图片划分为 $S \times S$ 网格,而不是使用 RPN 网络提取区域建议,这显著加快了检测速度。随后,Bi-directional Feature Pyramid Network(BiFPN)^[14]提出了一种具有高效快速特征的融合金字塔结构,可以在提高分辨率的同时扩展网络深度宽度,被用于骨干网络、分类和构架以及预测性网络,展现了双向特征融合对单级检测器检测性能提升的特点。其只需提取一次就可以获得多个尺度的特征图,弥补了 YOLO 系列算法只能训练固定尺度图像的明显缺陷。

上述深度学习的车辆检测网络算法皆未针对多特征融合联合训练的任务且通常需要高性能的硬件。多特征融合网络可以广泛应用于低成本的设备中,这类模型的推理速度比其他车辆检测的算法都要快。因此,设计一种轻量级、高效的多特征融合网络及语义分割模型是未来研究的重点。

(2) 基于深度学习的车道线检测研究

深度学习由于其鲁棒性和实时性迅速获得了关注,大致可分为4类:基于分割的方法、基于检测的方法、基于

参数曲线的方法、基于关键点的方法。

基于分割的方法主要是将车道线的像素分类,从而识别出图像中的目标物体,这类方法对于单目标识别效果较好,但是对于多目标识别效果较差。在这类方法中,一般是对车道线进行区域划分,然后再根据不同的区域特征来进行目标检测。为了区分不同的车道线,Spatial CNN(SCNN)^[15]将不同的车道线作为不同的类别,从而将车道检测转化为多类分割任务,并提出一个切片 CNN 结构,以使消息跨行和列传递。Recurrent Feature-Shift Aggregator(RESA)^[16]则是对 SCNN 切片 CNN 结构进行改进,加入切片间的不同步幅大小的信息传递,同时解耦相邻层之间时序上的依赖,增加并行处理能力。基于检测的道路标线预测一般是自上而下的,该方法充分利用了道路标线在驾驶员视野中由近到远的先验信息,从而构造出标线样本。在此基础上,研究者利用拟合分析法,设计了线性拟合分析器,并将拟合得到的拟合值与拟合曲线拟合;采用非极值抑制技术(NMS)对最大信任度进行筛选。已有的工作包括 LineCNN^[17],LaneATT^[18]等。这种自上而下的方法可以有效地利用道路轮廓的先验信息,保证道路轮廓识别的实时性,并可以有效地解决道路轮廓识别中存在的严重遮挡等问题。然而,预先设定的 Anchor 将会影响检测的灵活性。基于参数曲线的方法使用预先设定的参数曲线,对车道线进行检测,相关工作有 PolyLaneNet^[19],BézierLaneNet^[20]等。基于曲线的道路检测算法能够很好地实现对道路的全局描述,且检测速度快,但精度不高。

以深度学习为基础的车辆识别算法的开发,为交通道路算法的落地提供了一条切实可行的途径。从传统的基于样本分割的道路识别到融合先验信息的自上而下的可行驶道路区域识别,其准确率和速度都得到了极大的提升。因此,如何更好地利用道路线型的先验信息,是一个很有意义的研究课题。

(3) 基于可行驶的道路区域检测研究

早期,有研究者提出了一种以单目视觉检测为基础,通过构造车辆周围环境的低秩特征集,对其进行训练的分类器。该分类器能够识别出可行驶道路区域,但存在实时性差、鲁棒性差等缺点。Dong-Si 等^[21]提出了一种机器视觉与神经网络相结合的可行范围探测方法,但该方法仅适合固定路线无人驾驶的车辆(如大巴、巴士及其它沿固定线路行驶的车辆),且必须要有一台车固定在上方的照相机、一张地图,以及一台实时的汽车定位器。由 Ma 等^[22]提出的联合多级传递是一种传感器融合的方法,但是这种方法在实际应用中易受传感器融合的影响。显著性检测旨在从图像中找到有用和有吸引力的区域。然而,道路中存在多个物体的情况下,这些检测通常会降低具有相同吸引力的其他物体的显著程度^[23]。现有的道路网络图形检测工作一般可分为两类:基于分割的方法和基于图形的方法。基于分割的方法首先预测道路网络图的概率分割图,然后进行一系列处理以获得道路网络的图结构,如骨架化和过滤。道路网络图检测的早期工作大多属于这一类^[24]。

尽管这些基于图的方法可以增强拓扑正确性,但它们通常由两个独立的阶段组成,使得它们很难以端到端的方式进行优化。各个阶段可能会累积误差,上述多特征融合训练的语义信息不足会导致网络对于道路区域细节部分检测的缺失,从而降低其有效性和效率。为了解决以往工作的不足,

本文提出了一种多特征融合联合训练的驾驶感知网络模型YOLOP-L,其检测性能显著提升。

2 YOLOP-L:多特征融合驾驶感知模型

本章将详细介绍所提出的YOLOP-L多任务特征融合的网络架构。为了研究一个高效的前馈网络以协同完成道路检测的任务,本文提出了YOLOP-L——一种有效、高效的多任务学习网络,可同时执行交通目标检测、可行驶道路区域分割和车道检测任务,架构中主要包括设计的SP-LNet模块与Generalised Dice Loss(GDL-Focal)多类混合损失函数两部分内容。本文算法首先在训练前使用图像增强技术提高图像的清晰度,自适应增强图像的亮度;其次改进双向特征融合网络,解决多尺度检测的问题,通过在网络的每层分割头前加入特征金字塔,取得了较好的检测效果,降低了训练成本;通过改进多类混合损失函数,改善了车道线以及道路区域检测的效果。

2.1 SP-LNet 模块

本文在一组现有工作,如YOLOP^[25],HybridNet^[26]等的基础上设计了一个更高效的网络架构。YOLOP-L的网络架构如图1所示,其中SP-LNet为本文中改进的网络部分,将FPN与双向特征融合网络结合,然后颈部Spatial Pyramid Pooling(SPP)与Path Aggregation Network(PAN)网络融合形成路径增强进而提升网络分割体系的检测精度。所提出的网络架构由一个用于从输入图像中提取特征的共享编码器和3个用于相应任务的解码器头组成,YOLOP-L体系结构有一个编码器,即骨干网络和颈部网络;利用解码器头的3个分支来执行特定任务,而不是在同一分支中执行可行驶区域分割和车道检测的任务,3个分支为检测头、行驶区域以及车道线分割网络头。改进的原因主要是我们发现交通区域分割和车道检测的任务难度完全不同,这意味着这两项任务在特征级别上有不同的要求,因而具有不同的网络结构是更好的选择。本节将详细介绍YOLOP-L网络模型。

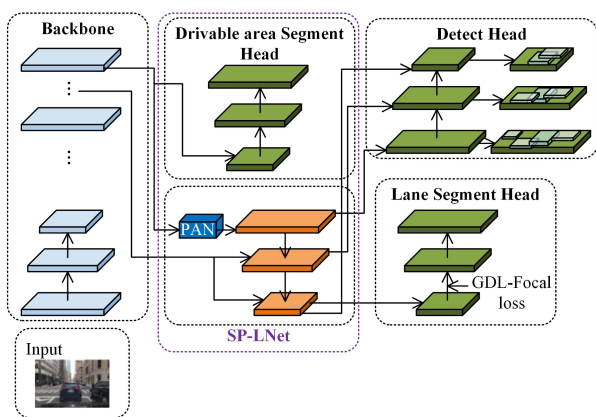


图1 YOLOP-L网络架构图

Fig.1 Diagram of YOLOP-L network architecture

2.1.1 Encoder

来自骨干网络的特征图被馈送到颈部网络通道,如何利用多尺度特征实现多特征融合是一大挑战。其中,多尺度特征图中的每个网格将被分配3个不同纵横比的锚点,检测头将预测位置的偏移量和缩放的高度和宽度,以及每个类别预测的概率和相应的置信度。YOLOP-L中的特征融合网络采用自下而上和自上而下的FPN+BiFPN特征聚合网络,以往

的特征融合是平等地对待不同尺度特征,而本文引入了权重,来更好地平衡不同尺度的特征信息。BiFPN是一种加权的的双向特征金字塔网络,它允许简单和快速的多尺度特征融合,该网络不仅可以融合强语义信息,保留了不同尺度大小的特征,给各个层赋予了不同权重去进行融合,促使网络更加关注重要的层次,而且还减少了一些不必要层的结点连接。与YOLOP网络类似,网络应用空间金字塔池模块来融合不同尺度的特征,并使用FPN模块来融合具有不同语义水平的特征。作为骨干网络的特征提取是模型的重要组成部分,可以帮助网络在各种任务中获得优异的性能。其融合的网络结构图表示如图2所示。

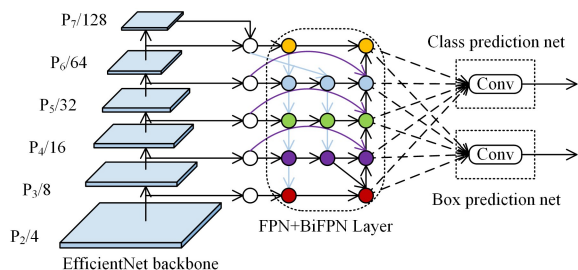


图2 FPN+BiFPN特征融合图

Fig.2 Diagram of FPN+BiFPN feature fusion

对于BiFPN的深度采用线性变换的方式向下取整;对于宽度采用指数变换的方式,在 $\{1.2, 1.25, 1.3, 1.35, 1.4, 1.45\}$ 中采用网格搜索确定1.35作为宽度的缩放因子。形式上,BiFPN的宽度和深度通过以下等式进行缩放。完整的缩放公式如下:

$$W_{\text{bifpn}} = 64 \cdot (1.35^\theta), D_{\text{bifpn}} = 3 + \theta \quad (1)$$

其中, W 表示网络宽度, D 为网络深度。宽度固定为和BiFPN的宽度相等即 $W_{\text{pred}} = W_{\text{bifpn}}$,但是随着层数的增加深度会线性上升,因而其网络深度 D 按下式进行线性变换:

$$D_{\text{box}} = D_{\text{class}} = 3 + \lceil \theta/3 \rceil \quad (2)$$

以上公式表明,本文模型只使用了一个参数 θ ,对整个目标检测网络进行分类预测与边界框回归预测,以及针对图像的分辨率进行联合缩放。

本文借鉴HybridNets模型,采用融合特征的方法,选择EfficientNet-B3作为骨干网络,它通过神经结构搜索找到深度、宽度和分辨率参数来解决网络优化问题,以设计稳定的网络;在颈部收集不同阶段生成的特征,通过拼接进行融合。BiFPN通过每个双向(自上而下和自下而上)路径,基于每个节点的跨尺度连接,以不同的分辨率融合特征,并为每个特征添加权重,以了解每个级别的重要性。

2.1.2 Decoder

如上所述,本文为每个单独的任务设计了3个单独的解码器头,分别为检测网络头以及两个分割网络头。与YOLOv7类似,本文采用了基于锚点的多尺度检测方案,该网络可以很好地处理复杂的数据集。首先,SP-LNet模块中使用路径聚合网络(PAN),它是自底向上的路径增强,特征的复杂度增加,同时图像的空间分辨率降低。与只有一条自顶向下和一条自底向上路径的PANet不同,本文将每个双向路径视为一个特征网络层,并多次重复同一层以实现更高级的多特征融合。通过结合PAN和SPP的特征能够将语义信息与这些局部特征融合,然后直接在PAN中的多尺度融合特

征图上进行检测。其次,对于区域行驶和车道线分割网络头,本文发现,与其他两个任务相比,从更深的网络层中提取的特征对于道路可行行驶分割网络是不必要的,而且这些更深的特征并没有提高预测性能,反而增加了训练期间模型收敛的难度。对于车道分割,为了提取到更深层次的特征,我们将任务分支连接到FPN的末端。此外,在解码器阶段,车道检测应用卷积以进一步提高检测性能。由此,本文对车道线分割网络的每个级别进行上采样,以获得相同大小 $(\frac{W}{4}, \frac{H}{4}, 64)$ 的输出特征图。这项研究将特征图缩放到强语义特征图P2级别,从P2低级特征实现最终的特征融合,将浅层特征与深层特征相结合,提高最终的输出精度,如图3所示。

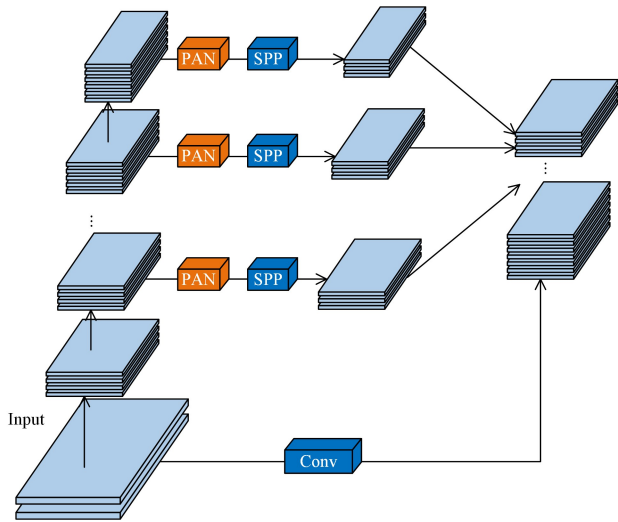


图3 YOLOP-L体系结构分支

Fig. 3 YOLOP-L architecture branch

本文保留了从Neck网络到分割分支的5个特征级别 $\{P_3, \dots, P_7\}$,上采样获得输出特征图后将它们组合起来,通过对所有级别进行求和来获得更好的特征融合。在本文中,为了回归网络能够顺利学习,设定了特定大小的一些锚框,表示为:

$$b_p = \begin{cases} 1, & \text{if } IOU(c_i, b_j) \geq 0.5, i=1, \dots; j=1, \dots; p; \sum_{k=1}^5 n_k m_k \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

其中, c 是锚框,锚框的总数是每个特征图 n 级别的组合, m 是特征图的分辨率, p 是每个输入图像的地面实况边界框的总数。

2.2 GDL-Focal 多类混合损失函数

本文采用多任务损失函数来训练YOLOP-L多特征融合网络,在YOLOP设计的基础上,在检测部分保留了损失函数的设置。分类损失、对象损失和边界损失的加权和损失表示为 L_{det} ,如式(4)所示:

$$L_{det} = \delta_1 L_{class} + \delta_2 L_{obj} + \alpha L_{box} \quad (4)$$

由于focal loss聚焦于训练一个困难样本的稀疏集,通过直接在标准的交叉熵损失函数基础上做改进,可以使得模型在训练过程中更专注于困难样本。 L_{class} 与 L_{obj} 使用focal loss,并额外增加 α 平衡因子,来处理样本不平衡的场景。 L_{class} 用于惩罚分类,而 L_{obj} 用于预测置信度。 L_{box} 反映了预测结果与实况尺度相似性的距离,即衡量模型预测的边界框与

真实边界框之间的差异,这有助于确保模型能够准确地定位对象。由于道路背景的复杂性,车道线分类任务较为困难,而对于车道分割,本文使用focal loss而不是交叉熵损失,这可以有效地引导模型关注复杂的示例,从而提高检测精度。此外,数据分布具有不平衡性,这对于对象分割任务是一个挑战,所以本文采用GDL-Focal多类混合损失函数,即结合 L_{GDL} Generalized Dice loss(GDL)与 L_{Focal} Focal loss来预测所属类别。当车道分割过程针对罕见的观察值时,候选标签之间很可能出现严重的类别失衡,从而导致性能不佳。本文利用GDL的类重平衡特性作为一种鲁棒且准确的深度学习损失函数来处理不平衡任务,从而减轻不平衡问题。Focal loss则最大程度地学习分类特征,进而最大可能地减少分类误差。如上所述,最终的损失函数可以按照式(5)进行计算。

$$L = L_{GDL} + \beta L_{Focal} \\ = C - \sum_{c=0}^{C-1} \frac{TP_p(c)}{TP_p(c) + \delta FN_p + \alpha FP_p(c)} - \frac{\beta}{N} \sum_{c=0}^{C-1} \sum_{n=1}^N g_n(c) (1 - p_n(c))^\gamma \log(p_n(c)) \quad (5)$$

$$TP_p(c) = \sum_{n=1}^N p_n(c) g_n(c) \quad (6)$$

$$FN_p(c) = \sum_{n=1}^N (1 - p_n(c)) g_n(c) \quad (7)$$

$$FP_p(c) = \sum_{n=1}^N p_n(c) (1 - g_n(c)) \quad (8)$$

其中, C 为类别的总数, β 是generalized dice loss和focal loss之间的权重,通过调节 β 值可以实现最佳平衡。对象分割任务分为车道分割和道路可行行驶区域分割两类,因而可将 γ 设置为2。如式(6)所示, TP_p 表示交并比大于0.5的检测框数量,代表实际正类预测为正类的数量;式(7)中, FN_p 表示交并比小于等于0.5的检测框数量,或者是检测到同一个GT框的多余检测框的数量,代表实际正类预测为负类的数量;而式(8)中 FP_p 是实际负类预测为正类的数量。

2.3 Gamma 自适应亮度增强

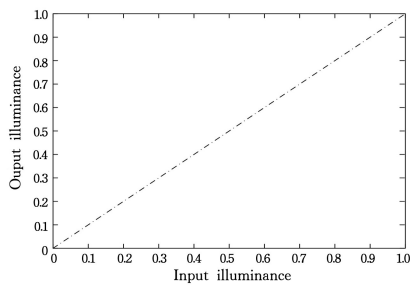
图像增强可以提高图像的质量,增强其视觉效果,提高图像的清晰度并丰富图像中的信息。在本文中,除了使用随机旋转和平移的图像增强外,还使用了一种自适应图像亮度增强算法,以减轻光线不足对交通目标、车道线以及可行行驶区域检测的影响。Gamma变换用于图像增强,其提升了暗部细节,通过非线性变换促使图像从曝光强度的线性响应变得更加接近人眼感受的响应。

$$V_{out} = V_{in} \gamma \quad (9)$$

其中, V_{in} 取值范围为 $0 \sim 1$,因此需要先进行归一化,然后取指数。 $\gamma > 1$ 时,较亮的区域灰度被拉伸,较暗的区域灰度被压缩得更暗,图像整体变暗; $\gamma < 1$ 时,较亮的区域灰度被压缩,较暗的区域灰度被拉伸得较亮,图像整体变亮。

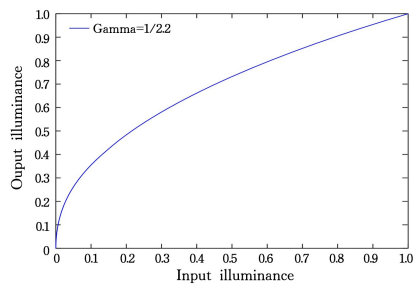
图4(a)为 $\gamma = 1$ 时原图的性能指标,图4(b)为校正后 $\gamma = 1/2.2$ 的指标。由此可知,校正后图像的低灰度区域对比度得到增强。图4(c)为原始白天中的交通道路行驶图像,主要表现为图像中亮的区域光线足够,而暗的区域光照不足,导致一些重要的道路及车辆细节信息无法凸显甚至被掩盖掉,严重影响了图像的视觉效果和应用价值。而图4(d)是经过亮度增强处理后的图片,对比表明,开展光照不均匀图像的校正研究可以有效地消除不均匀光照对图像的影

响,提升道路交通图像方面的视觉效果,给驾驶员提供更清晰



(a)

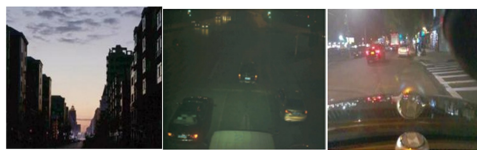
的道路前方视野及行驶环境。



(b)



(c)原始图像



(d)亮度增强图

图4 图像增强效果图

Fig.4 Image enhancement effect

本文采用改进的双向融合特征网络对原有算法进行改进,并对相应任务使用3个解码器头。改进后的网络对车辆目标的特征提取更加显著,对于车道线与道路区域分割也有优秀的性能提升。改进后的YOLOP-L网络不仅弱化了特征图中复杂的背景信息,增强了网络模型的目标提取能力,同时提高了网络特征融合和检测的能力。

3 实验分析

3.1 实验基础

本文实验环境配置如表1所列,其中原始算法和改进后的算法都基于Pytorch框架且使用Python语言完成,采用Pycharm平台训练,同时使用CUDA11.3.1进行加速,使用AMD Ryzen 7 6800H with Radeon Graphics配置一块rtx3060显卡进行网络训练和测试。

表1 实验环境配置

Table 1 Experimental environment configuration

软硬件环境	参数
处理器(CPU)	AMD Ryzen 7 6800H with Radeon Graphics
主频	3.20 Hz
硬件核心	8
内存(RAM)	64 GB
独立显卡	NVIDIA RTX3060
机带RAM	16.0 GB
操作系统	Windows 11 64 位

3.1.1 硬件参数

本小节介绍了实验的数据集设置和参数配置,所有实验都是在NVIDIA GeForce RTX 3060 LaptopGPU上和torch 2.0的配置环境下进行的。“余弦退火”策略用于调整训练过程中的学习率,其中初始学习率设置为0.01,学习率呈线性衰减趋势。网络训练的batch size为8,共训练200个Epoch。

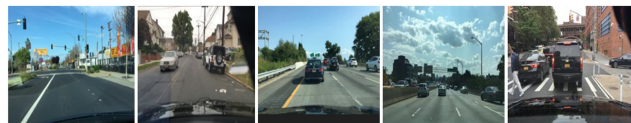
本文在训练阶段将BDD100k数据集中的图像从 $1280 \times 720 \times 3$ 调整为 $640 \times 640 \times 3$,在测试阶段将图片像素从 $1280 \times 720 \times 3$ 调整至 $640 \times 384 \times 3$ 。优化方法采用Adam。由于测试集的标签不是公开的,因而本文在验证集上评估网络。

3.1.2 数据集选择

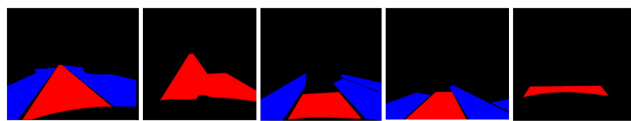
BDD100k是伯克利大学AI实验室(BAIR)发布的目前

最大规模、内容最具多样性且在驾驶场景中具有极大挑战性的公共数据集。BDD100k数据集包含10万段高清视频,每个视频约40s,720p,30fps。数据集中的GT框标签共有10个类别,其中汽车、卡车、公共汽车、火车被选择并合并为一个类别,即车辆;两个分割类也被合并到drivable中。

本文使用BDD100k作为实验研究的基准数据集,根据在BDD100k上训练的多任务网络来准备3个任务的数据集,由此本文将数据集划分为原始图像数据集、道路区域分割集、车道线分割集以及目标检测标注集,训练集有846张,测试及验证集374张,共1220张,其中车道线及车辆目标实例近10000个。YOLOP-L不仅在车道线、可行驶区域以及车辆目标检测具有更好的性能,而且对于复杂的环境背景,例如黑夜、雪地等条件下,亦具有较好的检测准确度。数据集部分图像如图5所示,图5(a)、图5(b)为两组共10张照片,分别对应着训练集中的原始图片与道路可行驶区域图片。数据集共有4880张图片并包含65028个实例,训练集与测试集的比例为2/3,1/3。类似的图,训练集中共有846组。



(a)原始图像数据集



(b)可行驶区域数据集

图5 数据集部分图像

Fig.5 Partial graphs in dataset

3.2 对比实验

本文在训练前使用数据增强来增加图像的可变性,从而使模型在不同的环境中具有鲁棒性。在训练方案中考虑了光度失真和几何失真的问题。本文尝试用不同的算法来训练提出模型。最简单的一种是端到端的训练,进而可以联合学习3项任务。当所有任务确实相关时,这种训练算法是有用的。

此外,还尝试了一些交替优化算法,逐步训练模型。在每个步骤中,模型都可以专注于一个或多个相关的任务,而不考虑那些不相关的任务。即使不是所有的任务都是相关的,本文模型仍然可以通过泛式充分学习每个任务。为了提高模型的性能,采用了一些实用的数据扩充技术和方法。在本节中,将YOLOP-L模型与现有YOLOP、HybridNets模型进行定量与定性分析,以验证YOLOP-L网络在3个检测任务上的性能。

3.2.1 网络参数与运行时间

本文对YOLOP算法进行改进,分为3个解码器,并将FPN与BiFPN融入进原始网络以实现检测的高精度。本节通过与现有网络的参数进行对比实验来验证YOLOP-L网络成本计算的性能,采用原始的YOLOP网络模型作为基准,进行分析。

表2列出了现有网络与YOLOP-L网络的参数和运行时间对比,尽管YOLOP-L具有更大的参数(37.6×10^6),但其网络运行的速度更快,需要的运行时间更少。浮点计算量在一定程度上能估计模型的复杂度,网络模型通过采用深度可分离卷积,计算量显著减少到 14.9×10^9 。这些都得益于所提到的有效的网络结构和重新设计的多类混合损失函数。本文在相同的实验环境下对指标进行了3个任务的所有测试。第二个实验给出了3个任务的结果,包括车辆目标检测、可行驶区域分割和车道线分割。

表2 模型参数和运行速度对比表

Table 2 Comparison of model parameters and operating speed

Network	Params	FLOPs	Speed/fps
YOLOP	7.9×10^6	18.6×10^9	48
HybridNets	12.8×10^6	15.6×10^9	37
YOLOP-L	37.6×10^6	14.9×10^9	32

3.2.2 车辆目标检测

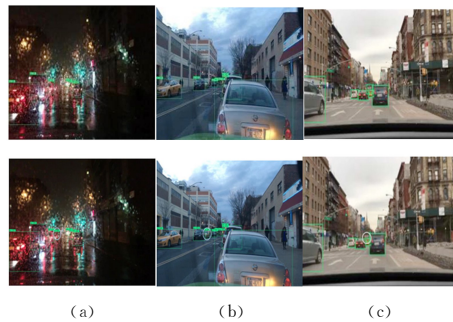
本小节将本文提出的YOLOP-L模型与其他SOTA(State Of The Art)模型进行对比,包括MultiNet,DLT-Net,Faster R-CNN,YOLOV5s,YOLOP,HybridNets。

表3列出了YOLOP-L模型与其余6个模型在车辆目标检测任务上对比的结果。检测主要使用mAP50和Recall作为评估指标,mAP即平均精度,其后数字表示NMS过程中设定IoU阈值的参数。图6为车辆目标检测定性结果。如表3所列,YOLOP-L在BDD100k数据集上的性能优于已有的网络,通过改进YOLOP网络,召回率提升了2.2%并且实现了最优的mAP50,为82.6%。由此表明,模型通过改进融合双向特征金字塔网络使得车辆的检测更为准确。图6(a)为雨天检测结果,图6(b)为阴天检测结果,图6(c)为雾霾天气下的检测结果。与第一行YOLOP网络车辆检测的对比直观验证了本网络的优异性能。

表3 车辆目标检测性能比较

Table 3 Comparison of vehicle target detection performances

Network	mAP50	Recall
MultiNet	60.2	81.3
DLT-Net	68.4	89.4
Faster R-CNN	55.6	77.2
YOLOV5s	77.2	86.8
YOLOP	76.5	89.2
HybridNets	77.3	92.8
YOLOP-L	82.6	91.4



(a) (b) (c)

图6 车辆目标检测任务比较图

Fig. 6 Comparison of vehicle target detection tasks

3.2.3 可行驶区域分割

IoU作为可行驶区域分割任务的评估指标,用于评估不同模型的分割性能。通过与其他5种分割网络进行比较来验证YOLOP-L模型的道路区域分割性能,如表4所列。

表4 可行驶区域分割任务性能比较

Table 4 Performance comparison of segmentation tasks at drivable area

Network	area
	Drivable mIoU
MultiNet	71.6
DLT-Net	71.3
PSPNet	89.6
YOLOP	91.5
HybridNets	90.5
YOLOP-L	92.6

表4列出了本文网络与5个可行驶网络的mIoU,而YOLOP-L实现了92.6%的mIoU,与YOLOP网络相比性能提升了1.1%。在与其他典型算法对比中YOLOP-L评估的结果都较好,这验证了本文构建解码器网络并用focal损失函数替代交叉熵损失函数的性能优势。YOLOP-L较以往的全景驾驶网络模型更灵活多变,整体提高了道路可行驶区域的检测精度。图7表明可行驶区域检测任务在黑夜白天中的性能都有所提升。



(a) (b) (c)

图7 可行驶区域分割任务比较图

Fig. 7 Comparison of segmentation tasks at drivable area

3.2.4 车道线检测

车道线检测是自动驾驶面临的主要挑战之一。由于数据集中车道线有两个注释的特殊性,因此,首先需要对比BDD100k数据集进行预处理,根据两条注释线计算中心线,绘制车道掩码进行训练,YOLOP-L使用准确性和IoU来评估车道线检测的性能,如表5所列。

表5列出了现有的5种车道线检测网络ENet,SCNN,ENet-SAD,YOLOP以及HybridNets与本文网络模型的对比结果。定量结果显示,YOLOP-L以88.2%的准确率和较

高的IoU 优于之前的模型。由于采用 GDL-Focal 多类混合损失函数和融合网络的改进,其准确率提升 2.8%,车道线 IoU 的值较 HybridNets 网络下降 2.45%,但较 YOLOP-L 网络提升 1.95%。

表 5 车道线检测任务性能比较

Table 5 Performance comparison of lane detection tasks

Network	Accuracy	Lane IoU
ENet	34.12	14.64
SCNN	35.79	15.84
ENet-SAD	36.56	16.02
YOLOP	70.50	26.20
HybridNets	85.40	31.60
YOLOP-L	88.20	28.15

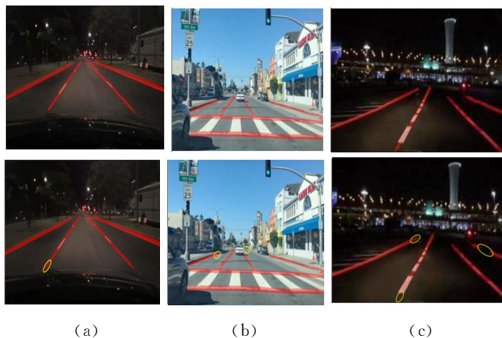


图 8 车道线检测任务比较图

Fig. 8 Comparison of lane line detection tasks

同样,图 8(a)和图 8(b)为两组检测效果对比图,如图 8 所示,所提出的网络模型在各种复杂的天气下都能够很好地工作,且与第一行原 YOLOP 算法相比检测的车道线连续且准确。

3.3 YOLOP-L 驾驶感知任务分析

本节将展示 YOLOP-L 在 3 个任务上的完整直观效果,其定量分析结果如表 6 所列,每个任务在 3 个网络中同一指标下的优秀性能;定性结果如图 9 所示。第一行为 YOLOP 的检测结果。图 9(a)展示了雪地的检测结果,存在检测缺失和可行驶区域分割缺失;图 9(b)、图 9(c)场景为黑夜的检测结果,存在车道线检测偏差和车辆检测缺失的问题;图 9(d)、图 9(e)为白天的检测结果,同样出现了车道线缺失检测和车辆检测缺失的问题,且图 9(e)中较为明显地展示了对于小型车辆检测的效果。第二行 5 个场景展现了 YOLOP-L 模型检测的结果,表明本文模型成功克服了以上驾驶员视角下检测精度不足的问题,在小型车辆目标检测任务上显示出了更好的性能。

表 6 联合检测总体性能比较

Table 6 Comparison of overall performance of joint testing

任务	评估指标	YOLOP	HybridNets	YOLOP-L(OURS)
车辆目标检测	mAP50	76.5	77.3	82.6
可行驶区域分割检测	mIoU	91.5	90.5	92.6
车道线检测	Accuracy	70.5	85.4	88.2

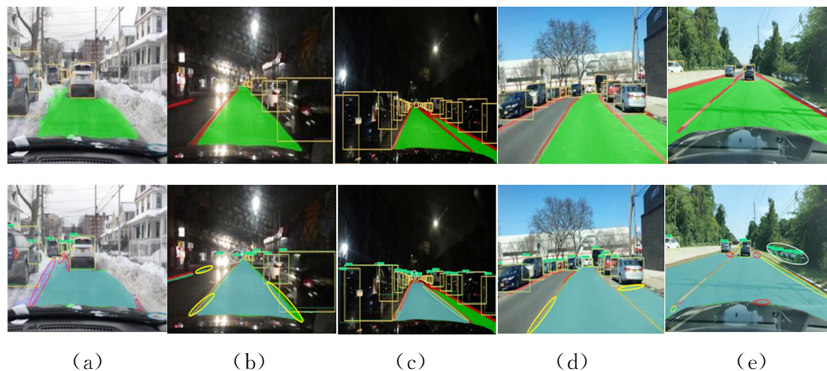


图 9 YOLOP-L 驾驶感知效果图比较

Fig. 9 Comparison of YOLOP-L driving perception effect

结束语 针对目前驾驶员视角下道路场景与背景信息较为复杂的特点以及原先网络的冗余性及交叉熵损失函数对于处理不平衡问题匮乏的缺陷,本文基于以往工作,系统性地研究了多任务的网络架构设计选择,提出了一种高效的多特征融合感知网络 YOLOP-L。模型通过改进双向融合特征网络并建立高效的多类混合损失函数,不仅提升了整体的准确性和性能,而且对于 3 个分类任务也具有更好的检测和分割效果。直接采用未经处理图像的常用方法并不能准确地检测出车辆目标,本文首先对图片进行一系列图像增强技术,主要对自适应提升图片增强的亮度进行预处理。在训练时,SP-Net 模块中将 FPN 与双向特征融合金字塔网络进行结合并在颈部进行 SPP 与 PAN 的网络融合以收集不同阶段生成的特征,通过拼接进行融合以设计更稳定的网络,了解每个级别的重要性,提升了网络对于车辆目标的检测准确性。对比实验表明,所提模型召回率提升了 2.2%;采用 GDL-Focal 多类混合损失函数和融合网络的改进,车道线检测准确率提升

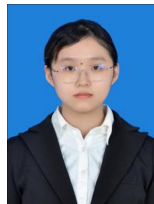
2.8%,车道线 IoU 的值较 HybridNets 网络下降 2.45%;可行驶区域整体检测性能提升 1.1%。最重要的是,与以前的多任务驾驶模型相比,本文模型以更少的 FLOPS 实现了显著的效果。

后续研究的重点仍然是提出一种更加稳定的多特征融合网络,可以执行更多与驾驶员视角下相关的任务。将在可行驶区域检测继续细化,在自动驾驶领域中继续尝试对多个任务(道路全景驾驶过程中会涉及到行人、非机动车、道路标识等)进行检测分类,保证模型满足更多全景驾驶需求。另外,将从网络和解码器角度考虑,如何进行更加准确的交通目标检测以及车道线、可行驶区域分割。

参考文献

- [1] ZOU Q,JIANG H,DAI Q,et al. Robust Lane Detection From Continuous Driving Scenes Using Deep Neural Networks[J]. IEEE Transactions on Vehicular Technology, 2020,69(1):

- 41-54.
- [2] SEO Y W, RAJKUMAR R R. Detection and tracking of boundary of unmarked roads[C]// International Conference on Information Fusion. IEEE, 2014.
- [3] DONG-SI T C, GUO D, YAN C H, et al. Robust extraction of shady roads for vision-based UGV navigation[C]// 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, Acropolis Convention Center, Nice, France. IEEE, 2008.
- [4] KALAKI A S, SAFABAKHSH R. Current and adjacent lanes detection for an autonomous vehicle to facilitate obstacle avoidance using a monocular camera[C]// 2014 Iranian Conference on Intelligent Systems(ICIS). IEEE, 2014:1-6.
- [5] YU L, TAN H, BANSAL M, et al. A Joint Speaker—Listener—Reinforcer Model for Referring Expressions[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:3521-3529.
- [6] WU F, XU Z, YANG Y, et al. An End-to-End Approach to Natural Language Object Retrieval via Context-Aware Deep Reinforcement Learning[J]. arXiv:1703.07579, 2017.
- [7] DUAN K, XIE L, QI H, et al. Corner Proposal Network for Anchor-free, Two-stage Object Detection[C]// European Conference on Computer Vision. 2020.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017:1137-1149.
- [9] ROSS G. Fast R—CNN [C]// 2015 IEEE International Conference on Computer Vision (ICCV). 2015:1440-1448.
- [10] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [C] // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015:1904-1916.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. 2015.
- [12] MA J, SHAO W, YE H, et al. Arbitrary-Oriented Scene Text Detection via Rotation Proposals[C] // IEEE Transactions on Multimedia. 2018:3111-3122.
- [13] 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). 2015:779-788.
- [14] TAN M X, PANG R, LE Q V, et al. EfficientDet: Scalable and Efficient Object Detection[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019:10778-10787.
- [15] PAN X G, et al. Spatial As Deep: Spatial CNN for Traffic Scene Understanding [C] // AAAI Conference on Artificial Intelligence. 2017.
- [16] ZHENG T, FANG H, ZHANG Y, et al. RESA: Recurrent Feature-Shift Aggregator for Lane Detection[C]// AAAI Conference on Artificial Intelligence. 2020.
- [17] LI X, LI J, HU X, et al. Line-CNN: End-to-End Traffic Line Detection With Line Proposal Unit[C]// IEEE Transactions on Intelligent Transportation Systems. 2020:248-258.
- [18] TABELINI L, BERRIEL R, PAIXAO T M, et al. Keep your Eyes on the Lane: Real-time Attention-guided Lane Detection [C] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020:294-302.
- [19] TABELINI L, BERRIEL R, PAIXAO T M, et al. PolyLaneNet: Lane Estimation via Deep Polynomial Regression [C] // 2020 25th International Conference on Pattern Recognition (ICPR). 2020:6150-6156.
- [20] FENG Z Y, GUO S, TAN X, et al. Rethinking Efficient Lane Detection via Curve Modeling[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022:17041-17049.
- [21] DONG-SI T C, GUO D, YAN C H, et al. Robust extraction of shady roads for vision — based UGV navigation [C] // 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Acropolis Convention Center, Nice, France. IEEE, 2008.
- [22] MA B, LAKSHMANAN S, HERO A, et al. Simultaneous detection of lane and pavement boundaries using model-based multi-sensor fusion[J]. IEEE Trans. Intell. Transp. Syst. , 2000:135-147.
- [23] ZHOU L, FANG J, JU Y, et al. Multi-Saliency Detection via Instance Specific Element Homology[C]// 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA). Sydney, NSW, Australia, 2017:1-8.
- [24] XU Z H, LIU Y, GAN L, et al. RINGDet: Road Network Graph Detection by Transformer in Aerial Images[C]// IEEE Transactions on Geoscience and Remote Sensing. 2022.
- [25] WU D S, LIAO M, ZHANG W, et al. YOLOP: You Only Look Once for Panoptic Driving Perception[C]// Machine Intelligence Research. 2021:550-562.
- [26] VU D, NGO B, PHAN H N, et al. HybridNets: End-to-End Perception Network[J]. arXiv:2203.09035, 2022.



LYU Jialu, born in 2001, postgraduate. Her main research interests include object detection and image processing.



ZHOU Li, born in 1987, Ph.D, professor. His main research interests include image processing and scientific visualization.