驾驶区域分割与车道线检测多任务网络方法研究

李 政[1],辜丽川[2,3]
1. 安徽农业大学 信息与人工智能学院,合肥 230036
2. 智慧农业技术与装备安徽省重点实验室,合肥 230036
3. 安徽省农情信息感知与智能计算工程研究中心,合肥 230036

摘 要:目的 交通驾驶感知系统是自动驾驶技术的关键组成部分之一,该系统包含多个任务并且需要保证高实时性和高可靠性,针对现有方法目标检测效果差,部分遮挡情况下目标区域语义分割不精确等问题,提出一种端到端的多任务网络模型 MTNet,用于完成交通驾驶感知系统中可驾驶区域分割和车道线检测两项关键任务。**方法** 在颈部网络设计中改进特征金字塔 FPN(Feature Pyramid Network)模块以获得更高效的检测性能;在网络结构中引入 ELAN(Efficient Layer Aggregation Networks)结构和重参数化方法对原始卷积进行升级;根据不同任务特点设计了更精准的多任务损失函数,同时提出了更有效的训练策略。**结果** 通过在 BDD100k (Berkeley DeepDrive)数据集上进行大量实验,表明可驾驶区域分割任务的平均交并比(Mean Intersection over Union,mIoU)达到 95.6 %,车道线检测任务的准确率(Accuracy)达到 89.3 %。**结论** 实验结果表明:MTNet 展现出对复杂背景干扰的鲁棒性,提升了目标检测分割的精确度;无论是在白天还是夜晚,复杂还是简单场景中,都具有良好的检测分割效果。

关键词:多任务网络;车道线检测;可驾驶区域分割;端到端训练
中图分类号:TP391.4 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2026.0002.008

Research on Multi-task Network for Drivable Area Segmentation and Lane Line Detection

LI Zheng[1], GU Lichuan[2,3]
1. School of Information and Artificial Intelligence, Anhui Agricultural University, Hefei 230036, China
2. Key Laboratory of Intelligent Agricultural Technology and Equipment of Anhui Province, Hefei 230036, China
3. Anhui Agricultural Situation Information Perception and Intelligent Computing Engineering Research Center, Hefei 230036, China

Abstract: Objective The traffic driving perception system is a critical component of autonomous driving technology. It involves multiple tasks and requires high real-time performance and reliability. Existing methods suffer from issues such as poor detection performance for small objects and imprecise semantic segmentation of target regions under partial occlusion. To address these problems, an end-to-end multi-task network model, called MTNet, is proposed to accomplish two critical tasks in the traffic driving perception system: drivable area segmentation and lane line detection. **Methods** MTNet improved the neck network design by enhancing the feature pyramid network (FPN) module, enabling more efficient detection performance. The network architecture incorporated the efficient layer aggregation network (ELAN) structure and reparameterization techniques to upgrade the original convolutions. Moreover, precise multi-task loss functions were designed based on the characteristics of each task, and an effective training strategy was designed. **Results** Extensive

收稿日期:2024-03-10 **修回日期:**2024-05-12 **文章编号:**1672-058X(2026)02-0061-08
基金项目:国家自然科学基金项目(62306008)资助;安徽省高校协同创新计划项目(GXXT-2022-055)资助.
作者简介:李政(1998—),男,硕士研究生,从事计算机视觉研究.
通信作者:辜丽川(1974—),男,教授,博士,从事人工智能与大数据研究. Email:glc@ahau.edu.cn.
引用格式:李政,辜丽川. 驾驶区域分割与车道线检测多任务网络方法研究[J]. 重庆工商大学学报(自然科学版),2026,43(2):61-68.
LI Zheng, GU Lichuan. Research on multi-task network for drivable area segmentation and lane line detection[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(2): 61-68.

experiments conducted on the BDD100k (Berkeley DeepDrive) dataset demonstrated the effectiveness of MTNet. The drivable area segmentation task achieved an average intersection over union (IoU) of 95.6 %, while the lane line detection task achieved an accuracy of 89.3 %. **Conclusion** The experimental results show that MTNet exhibits robustness against complex background interference and improves the accuracy of object detection and segmentation. The model achieves good detection and segmentation performance in various scenarios, whether during the day or at night, and in both complex and simple environments.

Keywords: multi-task network; lane line detection; drivable area segmentation; end-to-end training

交通驾驶感知系统是自动驾驶技术的关键组成部分之一,该系统包含多个任务并且需要保证高实时性和高可靠性。在该系统的两项重要任务中,车道线检测任务可以获取道路上每条车道线的走向和形状,可驾驶区域分割任务可以获取安全的驾驶区域,两项任务为自动驾驶车辆的行车安全和路线规划提供了重要信息。

全卷积网络 FCN(Fully Convolutional Network)[1]是语义分割领域的重要里程碑,它将传统的全连接层替换为卷积层,实现了对图像进行像素级别的分类,从而实现了语义级别的图像分割,但是对像素逐个计算卷积导致效率低下。随着该领域研究的不断深入,越来越多的高性能模型被设计出来。DeepLab 系列模型[2-3],使用 Atrous 卷积(稀疏卷积)来提取图像特征,并使用 FCN 进行图像分割。因为使用 Atrous 卷积由于图像边界效应,很难获得图像的大范围特征,导致分割结果不够精细。UNet[4]设计了 Encoder-Decoder 结构,Encoder 部分为多层卷积池化,不断扩大感受野;Decoder 部分使用上采样以恢复图片的原尺寸得到分割结果。这样的网络结构冗余很大,导致网络训练很慢。UNet 独特的解编码和跳跃连接结构,使得研究者对这种 U 型网络结构有着持续的研究[5]。PSPNet(Pyramid Scene Parsing Network)[6]采用空间金字塔池化,实现了多层次的语义特征融合。该方法对输入图像进行多尺度特征提取,将这些特征图连接在一起,完成了多层次的语义特征融合。但该方法没有充分考虑像素之间的关系,缺乏空间一致性。

在车道线检测任务中,Pan 等[7]提出切片 CNN 结构,引入额外的通道维度,以捕捉图像在空间维度上的特征关系,从而更好地捕捉图像的结构信息。这种特殊类型的卷积对于依赖空间关系的任务,具有重要的意义。但是该模型参数复杂度较大,且当节点较多时容易过拟合。Zheng 等[8]延续了 Pan 的思路,对 SCNN(Spatial Convolutional Neural Network)的切片 CNN(Convolutional Neural Network)结构进行了改进,加入切片间不同步幅大小信息的传递,同时解耦相邻层之间时序上的依赖,增加并行处理能力,大大提高了处理速度。但是该模型在处理复杂场景,如道路交叉口、复杂道路标记或车辆遮挡等情况下可能导致模型的性能下降或无法准确地检测车道线。Hou 等[9]使用语义分割和知识蒸馏思路来解决车道线检测问题,提出一种自注意力蒸馏方法,添加 SAD(Self-Attention Distillation)模块,让浅层特征来学习高层特征的表达,从而加强网络的整体特征表达能力,同时保持了模型的轻量级设计。该方法注重实时性和效率,这可能会与分割精度之间存在权衡,尤其是在需要精细分割的场景中,无法捕捉到细小目标或边界。

上述一系列关于可驾驶区域分割和车道线检测的工作,仍然存在目标检测效果差,部分遮挡情况下目标区域语义分割不精确等问题。除此之外,在实时交通驾驶感知系统中,为每个单独任务运行单独的网络模型通常是不可行的,通过不同网络模型执行不同的任务会因为延迟而产生结果不同步的问题,并且依次执行多个任务可能会花费更长的时间。在这种情况下,由于多任务网络共享特征提取的主干网络,可以实现多项任务间的信息共享,因此它在节省计算成本的同时可以提高每项任务的性能。

针对上述问题,提出一种多任务交通驾驶感知网络 MTNet,用于完成交通驾驶感知系统中的可驾驶区域分割和车道线检测两项关键任务。网络结构受 YOLO 系列的启发,设计了更为有效的多任务网络结构,同时设计了更精准的多任务损失函数以及更有效的训练策略,用于解决和提升目标区域检测分割不精确的问题。在 BDD100k[13]数据集上,可驾驶区域分割任务的 mIoU 达到 95.6 %,车道线检测任务的准确率达到 89.3 %,同时在 NVIDIA 3090 上运行每秒的帧数(FPS)可以达到 62,能够满足实时性的要求。

1 方法的提出

1.1 多任务网络框架

提出一种多任务网络框架模型 MTNet。它由一个编码器和两个解码器构成,其网络结构如图 1 所示。编码器包括骨干网络和颈部网络,其主要用于从输入图像中提取融合生成不同层次和不同语义级别的特征供后续解码器使用,解码器用于解决相应的任务。两个解码器共享编码器提取的特征,可以实现多项任务间的信息共享。这样减少了重复计算量,缩短了多个任务的推理时间,并使网络能够轻松地进行端到端训练。

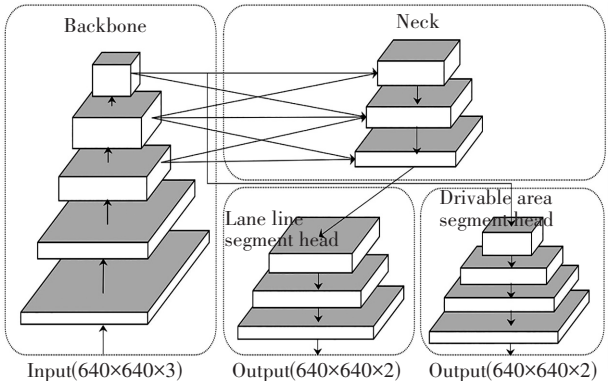


图 1　MTNet 网络结构

Fig. 1　MTNet network structure

1.1.1　骨干网络

骨干网络用于对输入图像进行特征提取,特征提取的好坏直接影响到后续多个任务的性能。本文采用 ELAN[10] 结构对主干进行构建,ELAN 结构利用群卷积增加特征的数量,并将不同组的特征组合起来,以得到不同的特征图。它通过控制最短和最长的梯度路径,使网络能够学习到更多的特征,且具有更强的鲁棒性。

骨干网络结构如图 2 所示。骨干网络接收 640×640×3 的图像作为输入,首先经过 4 个基本模块,每个模块主要由 Conv2d、BN (Batch normalization)[14]、MetaAconC[15] (CBM)组成。Conv2d 为卷积层,BN 层对特征图进行批量归一化,MetaAconC 是自适应激活函数,其表现出色,它训练了多组可学习参数来对每个通道进行处理,并且可以通过自适应函数来控制是否激活网络神经元,它同样可以关注和捕获空间信息来增加感受野,与卷积捕获的空间信息相互促进,以达到更好的效果。之后特征图变为 160×160×128 的大小,紧接着后面连接的是 4 个 ELAN 模块,并且在 ELAN 模块前面加入 MP 结构用于降维,其上方字母 C 表示通道数。

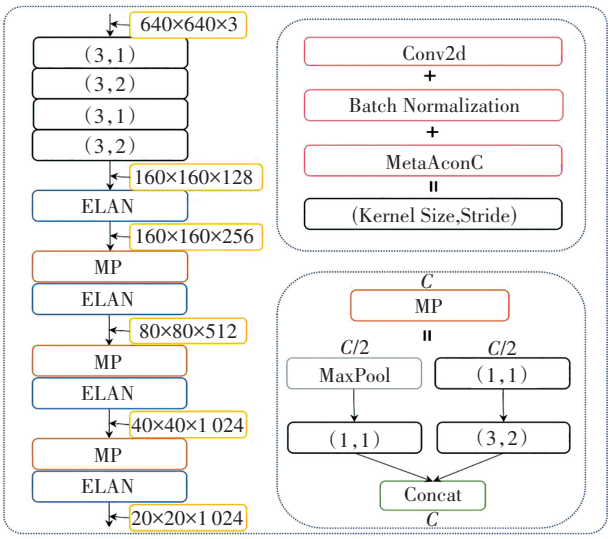


图 2　主干网络结构

Fig. 2　Backbone network structure

1.1.2　颈部网络

特征金字塔网络 FPN 旨在聚合从骨干网络提取的不同分辨率特征。传统 FPN 通过自顶向下的路径来融合多尺度特征。为了解决 FPN 单向信息流的局限性,使用 GFPN(Generalized Feature Pyramid Network)[16] 模块,并且使用 ELAN 的连接对原始卷积进行升级,在不增加额外计算负担的情况下,实现更高的精度。其结构如图 3 所示,颜色越红的节点表示的特征图越小。

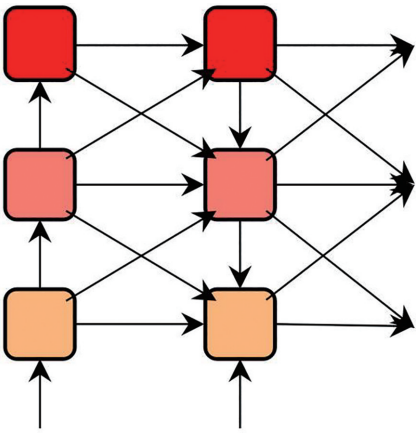


图 3　GFPN 网络结构

Fig. 3　GFPN network structure

1.1.3　解码器结构

MTNet 总共有两个任务,对应着两个解码器,其两个任务解码器的网络结构如图 4 所示。图 4(a)为可驾驶区域分割解码器,图 4(b)为车道线检测解码器。图中括号内的数字分别代表卷积核大小(Kernel Size)和步长(Stride)。

在可驾驶区域分割任务中,数据集中的场景非常多样且复杂,场景中不仅有单向车道,还有双向多车道。网络输出的可驾驶区域一定不能出现反向车道区域,这也是其他现有模型最容易出现错误的地方。为了使网络可以区分双向车道,在训练过程中必须要学习到一些区分方向的定位标志,如前方车辆的行驶方向、车道线等,这需要网络拥有强大的特征提取能力。GFPN 结构可以很好解决该问题,因为它可以融合不同尺度不同层次信息并且包含语义信息和空间位置信息。

为了使网络分割得更加精准,需要更多的空间位置信息、细节信息。因此,把可驾驶区域分割任务的网络分支连接在骨干网络的最后一层,这样可以提升网络分割的精确度。可驾驶区域分割任务解码器结构如图 4(a)所示。

骨干网络的末端连接,其特征图大小为 $W/32\times H/32\times 1\ 024$。为了将图像还原到原始尺寸,在此解码器中总共应用 5 个最近插值上采样(Upsample: Nearest),将输出的特征图恢复为 $W\times H\times 2$ 大小,$W\times H$ 表

示原图大小,2 表示输出的每个像素对于目标区域与背景的概率值。

车道线检测任务难度大,因为在 BDD100K 数据集中车道线的变化多样且复杂。遮挡部分没有标注,大量车道线是虚线等问题导致车道线形状非常零散,并且还存在光照影响和语义模糊等问题。对于区域较为零散且较小的目标,特征中应该囊括更多的背景,因为小目标的检测受背景影响较大,所以解码器需要高级特征和低级特征的融合来增强检测能力。为了获取融合后的特征,将该任务解码器分支连接到颈部网络 GFPN 层的末端,结构如图 4(b)所示。

GFPN 层的末端连接,其特征图大小为 $W/8\times H/8\times 128$。与可驾驶区域分割任务不同的是,在解码器上采样过程中使用了 3 次反卷积(ConvTranspose2d),将特征图恢复到原始图片尺寸 $W\times H\times 2$ 的大小。反卷积上采样相对于其他上采样方法,有更多的可学习参数,在增加少量时间成本的情况下,可以进一步提高车道线检测任务的准确性。

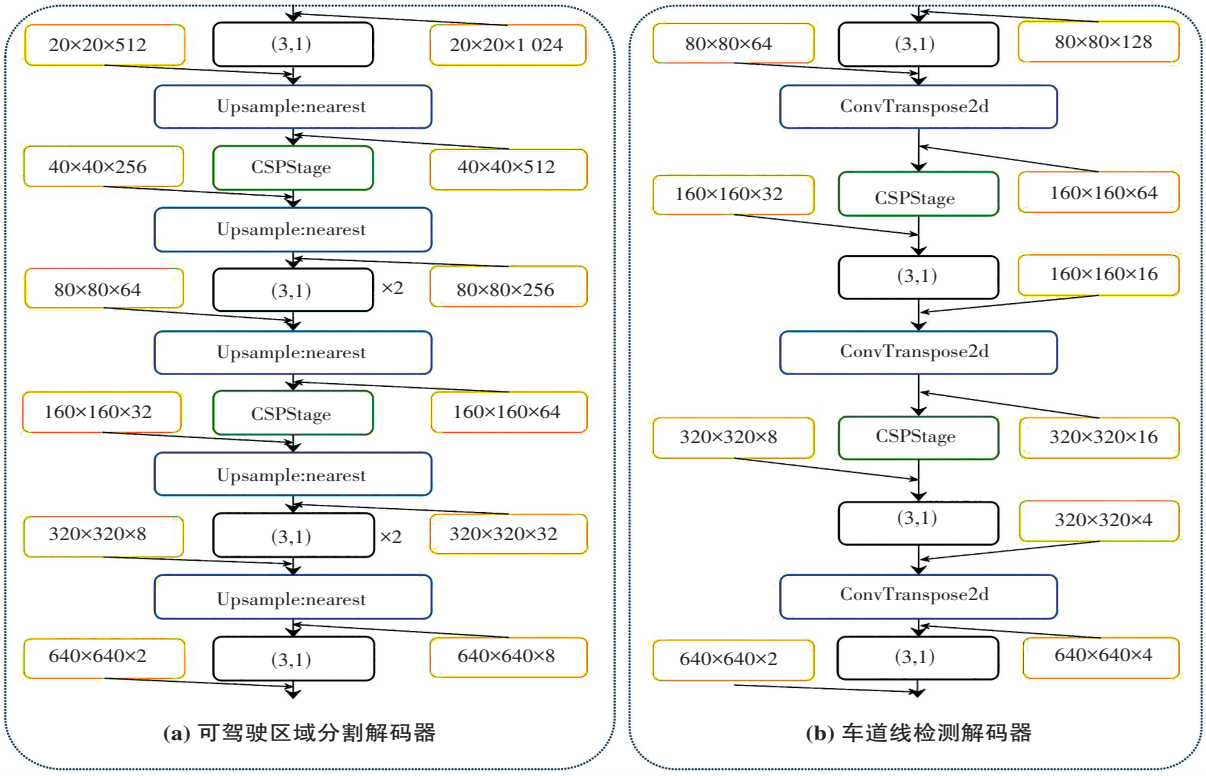


图 4 解码器结构

Fig. 4 Decoder structure

1.2 损失函数设计

本文提出的多任务网络中有两个解码器,所以多任务损失函数包含两个部分,分别是可驾驶区域分割任务损失 l_{da}、车道线检测任务损失 l_{ll}。l_{all} 是所有损失的加权和,如式(1)、式(2)所示。其中的 φ_1、φ_2、α_1、α_2 是用于平衡总损失的参数。

$$l_{all}=\varphi_1 l_{da}+\varphi_2 l_{ll} \tag{1}$$

$$l_{all}=\varphi_1 l_{BCE}+\varphi_2(\alpha_1 l_{focal}+\alpha_2 l_{dice}) \tag{2}$$

对于可驾驶区域分割任务损失 l_{da},如式(3),使用的是交叉熵损失,目的是最大限度地减少网络输出结果和真实结果的分类误差。

$$l_{da}=\frac{1}{N}\sum_{i=1}^{N}(y_i\log(\sigma(p_i)))+\frac{1}{N}\sum_{i=1}^{N}((1-y_i)\log(1-\sigma(p_i))) \tag{3}$$

其中,$\sigma(x)=1/(1+e^{-x})$,N 表示样本大小,$y_i\in\{0,1\}$ 是二元标签,$p_i\in[0,1]$ 是每个样本预测的概率。

车道线检测任务损失 l_{ll},如式(4)—式(6)所示,使用的是 Focal Loss[17] 和 Dice Loss[18] 的加权和。因为对于车道线检测等难以分类的任务,使用 Focal Loss 减少了简单示例的损失贡献,并加强了对纠正错误分类示例的比例,可以有效引导模型关注错误示例,从而提高检测精度。Focal Loss 通过为容易错误分类的示例分配更多的权重并对简单示例降低权重来处理不平衡问题,其在样本不平衡的情况下特别有效。对于整幅图像来说,车道线细长且零散并且所占面积很小,大多数像素通常都是背景的情况下,Focal Loss 较为合适。Dice Loss 也可以缓解样本中前景背景不平衡带来的消极影响。Dice Loss 训练更关注对前景区域的挖掘,但会存在损失饱和问题,因此单独使用 Dice Loss 往往并

不能取得较好的结果。由 Dice Loss 和 Focal Loss 组成的混合损失可以更好地提高车道线检测能力。

$$l_{ll}=\beta_1 l_{focal}+\beta_2 l_{dice} \tag{4}$$

$$l_{focal}=\begin{cases}-\alpha(1-p_t)^{\gamma}\log(p_t) & y=1\\ -(1-\alpha)p_t^{\gamma}\log(1-p_t) & y=0\end{cases} \tag{5}$$

其中,α 是用于调整正类和负类样本的权重,γ 是用于修改焦点损失的注意力参数,y 是样本的标签,p_t 是概率。

$$l_{dice}=1-\frac{2\sum_{i=1}^{N}y_i\hat{y}_i}{\sum_{i=1}^{N}y_i+\sum_{i=1}^{N}\hat{y}_i} \tag{6}$$

其中,y_i 和 $\hat{y}_i$ 分别表示像素 i 的标签值和预测值,N 是像素的数量。

2　仿真实验与结果分析

2.1　实验数据集

本节使用 BDD100K 数据集用于训练和验证模型。BDD100K 是一个在驾驶场景中具有挑战性的公共数据集,该数据集包含 10 万张以驾驶员为视角的透视图像,大小为 1 280×720×3,被广泛用作自动驾驶计算机视觉研究的评估基准。BDD100K 数据集有 3 个部分,分别是 70 000 图像的训练集、10 000 图像的验证集和 20 000 图像的测试集。由于 20 000 测试集并未公开数据标签,所以选用 10 000 验证集来测试模型。将数据集两个区域类别“direct”,“alternative”合并成一个“drivable”类别,并且将训练集中的车道线掩码图扩展到 8 个像素宽,同时保持验证集车道线为 2 个像素宽。

2.2　实验细节

实验过程使用 Adam 优化器来训练模型,使用 Warm-up[19] 和“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 0.001,Batch size 设置为 16,将 Momentum 和 Weight decay 参数设置为 0.937 和 0.000 5,训练总轮数设定为 150 轮。训练一轮是指将训练数据集中的所有样本都输入到模型中,计算损失函数并更新模型参数的一次迭代过程。在网络训练过程中使用指数移动平均值(EMA)和分组权重衰减,并设置 Warm-restart,目的是引导模型更快更好地收敛。为了获得更强的鲁棒性还在训练过程中使用了 Mosaic 和 Mixup 数据增强策略[20],本文采用在前 50 轮进行交替开启和关闭 Mosaic 和 Mixup 数据增强策略,50 轮过后关闭该数据增强策略。交替开启和关闭的意思是按照轮数的奇偶性开启和关闭,奇数如 1、3、5 等轮次开启,偶数 2、4、6 等轮次关闭。

这么做的原因主要有两个。一是该数据增强策略在可驾驶区域分割任务中可以提高约 0.5%~1.5%的性能,但是在车道线检测任务中会大幅降低 5%~15%的性能,并且由于可驾驶区域分割任务收敛速度较快,在前期使用 Mosaic 和 Mixup 可以提高可驾驶区域分割任务的性能,训练后期关闭该策略可以最大限度地防止对车道线检测任务的影响。二是 Mosaic 和 Mixup 数据增强策略在训练过程中非常耗时,开启耗时大约是不开启的 2~3 倍。如果开启 Mosaic 和 Mixup 数据增强策略,可能会大幅增加训练的时间成本。

在训练和测试阶段,将 BDD100K 数据集中的图像大小从 1 280×720×3 调整为 640×640×3。除了上面所提到的 Mosaic 和 Mixup 数据增强策略,还使用了其他基本的数据增强技术,如旋转,缩放,平移,水平翻转,剪切和调整图像的亮度、色度、对比度、饱和度等来处理几何失真和光度失真以增强模型的鲁棒性。

2.3　评价指标

可驾驶区域分割任务和车道线检测任务是语义分割任务,所以使用平均交并比 f_{mIoU}、精确度 $f_{Accuracy}$ 作为评价指标。f_{mIoU} 能够精确地衡量预测结果与真实标签的重叠程度,计算方式如下:

$$f_{mIoU}=\frac{1}{n+1}\sum_{i=0}^{n}\frac{p_{ii}}{\sum_{j=0}^{n}p_{ij}+\sum_{j=0}^{n}p_{ji}-p_{ii}} \tag{8}$$

其中,设定 i 表示真实值,j 表示预测值,p_{ij} 表示将 i 预测为 j 的概率,n 是测试集中所有样本的数目。

$f_{Accuracy}$ 适用于二分类语义分割问题,能够衡量模型对于前景和背景的分类准确率,计算方式如下:

$$f_{Accuracy}=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}} \tag{9}$$

其中,N_{TP} 表示被预测正确的正样本数,N_{TN} 表示被预测正确的负样本数,N_{FP} 表示被预测错误的正样本数,N_{FN} 表示被预测错误的负样本数。

2.4　实验结果分析

在本节中对多任务网络模型进行端到端的训练,然后将其与所有两项任务中的其他现有代表性模型进行详细的实验对比。

2.4.1　可驾驶区域分割任务结果

在可驾驶区域分割任务中使用 f_{mIoU} 来评估不同模型的分割能力,表 1 展示了不同网络模型的分割结果。MTNet 模型达到了目前最优的结果,超越了 SOTA 模型 YOLOPv2[21] 93.2 %的结果,达到 95.6 %,比 YOLOPv2 提高了 2.4 %。根据此任务的特点,设计了单独的解码器分支。可驾驶区域所占面积相对较大且较为规整,并且具有明显的边界。为了保证网络可以捕获更多的位置和细节信息,选择将该分支连接到主干网络的最后一层,为此可以获得更多浅层特征,使可驾驶区域分割更加精确。

表 1 可驾驶区域分割结果

Table 1 Drivable area segmentation results /%

网络模型	$f_{Accuracy}$	f_{mIoU}
MultiNet[22]	—	71.6
DLT-Net[23]	—	71.3
PSPNet[6]	—	89.6
YOLOP[24]	—	91.5
HybridNets[25]	—	90.5
YOLOPv2[21]	97.9	93.2
MTNet(our)	**99.1 (+1.2)**	**95.6 (+2.4)**

2.4.2 车道线检测任务结果

车道线检测是自动驾驶技术中面临的主要挑战之一。在实验中遵守 YOLOP 所提出的数据处理方式,预处理 BDD100K 数据集。将双车道线中心作为新的单车道线,绘制一个 8 像素宽的车道线掩码图进行训练,而测试集的车道线宽度为 2 像素,这样的做法导致了所有网络 f_{IoU} 的结果偏低。使用像素准确度 $f_{Accuracy}$ 和车道线的 f_{IoU} 作为评估指标,实验结果如表 2 所示。

表 2 车道线检测结果

Table 2 Lane line detection results /%

网络模型	$f_{Accuracy}$	f_{IoU}
ENet[9]	34.12	14.64
SCNN	35.79	15.84
ENet-SAD	36.56	16.02
YOLOP	70.50	26.20
HybirdNets	85.40	**31.60**
YOLOPv2	87.31	27.25
MTNet(our)	**88.10 (+0.79)**	29.8 (-1.8)

本章所提出的是端到端的多任务网络模型,无须进行结果后处理操作。但为了进一步提高车道线的任务性能,尝试对该任务进行结果后处理操作。使用传统图像处理方式进一步完善模型输出的结果。使用 OpenCV 进行图像形态学处理,尝试使用但不限于如 Erode、Dilate、Thinning、Open-Close、TopHat、BlackHat 等算法操作,对使用前和使用后的结果进行对比实验,结果如表 3 所示。f_{time} 表示形态学处理额外消耗的时间。通过对车道线任务输出后的数据进行形态学处理,进一步提高了 $f_{Accuracy}$,但是也会导致 f_{IoU} 的降低,并且随之也会增加额外的处理时间。

表 3 原始网络和形态学处理后的结果

Table 3 The results of the original network and morphological processing

方 法	$f_{Accuracy}$/%	f_{IoU}/%	f_{time}/s
YOLOP	70.50	26.20	—
HybirdNets	85.40	**31.60**	—
YOLOPv2	87.31	27.25	—
MTNet-mop(our)	**89.30 (+1.99)**	27.1 (-4.5)	0.122

2.5 消融实验

为了验证本文多任务网络的有效性,分别对网络进行端到端的多任务训练与 3 个任务单独训练进行结果对比实验。f_{speed} 所指的运行时间只包含模型推理时间,不包括 NMS 等后续处理时间。最终实验结果如表 4 所示。

表 4 多任务与单任务消融实验

Table 4 Multi-task and single-task ablation experiment

Training method	Dirvable $f_{Accuracy}$/%	Dirvable f_{mIoU}/%	Line $f_{Accuracy}$/%	Lane f_{IoU}/%	f_{speed}/ms
Da-Seg(only)	99.1	95.6	—	—	10.3
Line-Seg(only)	—	—	94.3	26.4	10.8
Multi-task	98.8	95.9	90.1	27.9	13.9

多任务训练与两个任务单独训练的实验结果表明:虽然单个任务推理时间要略小于多任务,但是两个单任务网络消耗的总时间要远大于多任务。这说明多任务网络可以大大缩短多个任务的推理时间,这是由于两个任务共享编码器数据,减少了重复的计算。除此之外,多任务网络的各个任务性能与单个任务的推理性能不相上下,这说明本文提出的多任务网络模型在上述两个任务的联合训练中是非常有效的。

2.6 可视化结果分析

图 5、图 6 分别展示了 MTNet 和现有模型检测方法在 BDD100K 数据集上的两个任务推理结果的可视化图像对比。为了测试模型的鲁棒性,分别选择白天和夜晚共 6 个不同的场景。图中红色区域表示模型检测出的车道线,绿色区域表示模型检测出的车辆可驾驶区域。

图 5 为白天 3 个不同场景下的推理检测结果对比。YOLOP 模型在第一个场景中出现车道线检测不连续的问题;可驾驶区域分割也不是很精确,同样的问题在第二、三场景中也十分明显(有问题的区域使用灰白色椭圆标出)。HybridNets 模型在 3 个场景中也出现了可驾驶区域分割错误以及不精确的问题,同时车道线检测也出现不连续现象。可以非常直观地看出,MTNet 在白天的 3 个不同场景中表现最为出色,都较为准确地检测出了正确的结果。

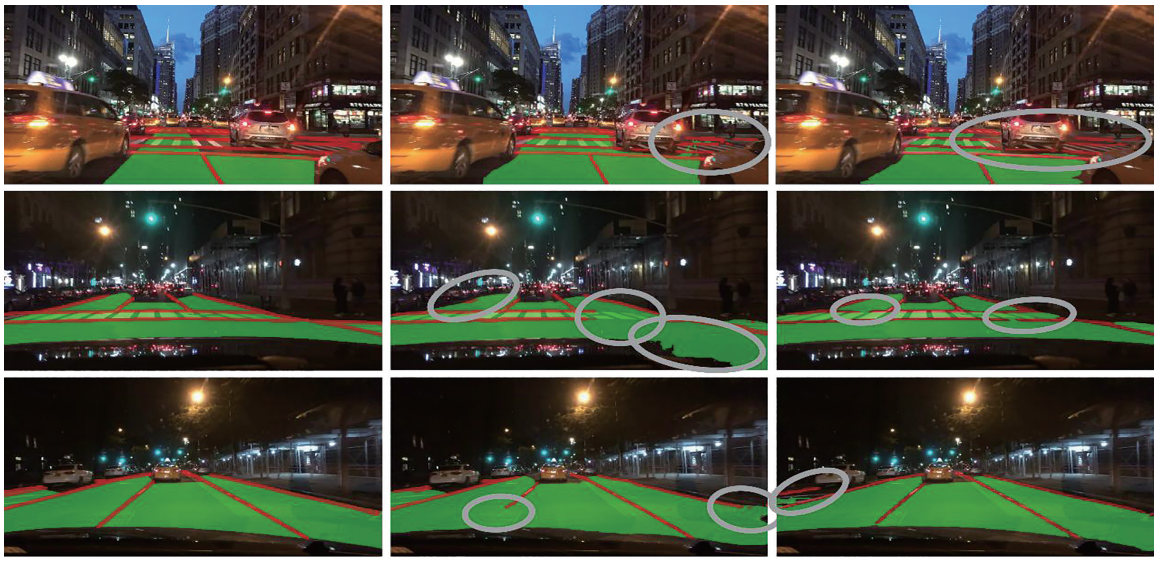
图 6 为夜晚 3 个不同场景下不同模型的推理检测

结果对比。夜晚光线较暗且光线复杂,检测难度大。YOLOP 和 HybridNets 模型在不同场景中都不同程度地出现了检测错误,检测不精确等问题。MTNet 则表现最佳,车道线及可驾驶区域检测分割规整精确。

MTNct　　UOLOP　　HybridNets

图 5　BDD100K 数据集白天测试结果

Fig. 5　Test results on the BDD100K dataset(day-time)



MTNct　　UOLOP　　HybridNets

图 6　BDD100K 数据集夜晚测试结果

Fig. 6　Test results on the BDD100K dataset(night-time)

综上所述,通过对白天夜晚 6 个不同难度场景进行可视化对比测试,MTNet 在所有场景中都表现良好,检测分割准确率高,鲁棒性较强。

3　结论与展望

提出一种端到端的多任务网络模型 MTNet,它可以同时处理自动驾驶过程中的可驾驶区域分割、车道线检测两个重要的视觉感知任务。在颈部网络设计中改进了 FPN 结构,解决了单向信息流的局限性;在网络中引入 ELAN 结构和重参数化方法对原始卷积进行升级;此外,还设计了新的多任务损失函数以及更为有效的训练策略。通过大量实验测试验证:MTNet 展现出对复杂背景干扰的鲁棒性,提升了目标检测分割的精确度;无论是白天还是夜晚,复杂还是简单的场景都具有良好的检测分割效果,并且在准确性和速度方面都达到了先进水平。通过消融实验验证了本文提出的多任务网络模型在这两项任务联合训练中的有效性,并且可以轻松实现端到端的训练。未来工作中,可以考虑先对所要分割的目标进行检测定位,然后再进行像素

粒度的分割,这样可以减少时间开销,加快模型推理速度,提高检测精度。

参考文献(References):

[1] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3431-3440.

[2] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[3] WEBER M, WANG H, QIAO S, et al. DeepLab2: A TensorFlow library for deep labeling[EB/OL]. 2021: arXiv: 2106.09748. https://arxiv.org/abs/2106.09748.

[4] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer, 2015: 234-241.

[5] 陈亮, 高文根, 张晨, 等. 新型架构下的密集网络在肺部影像的分割研究[J]. 重庆工商大学学报(自然科学版), 2023, 40(6): 53-60.
CHEN Liang, GAO Wen-gen, ZHANG Chen, et al. Segmentation of lung images with dense network models under a new architecture[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2023, 40(6): 53-60.

[6] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6230-6239.

[7] PAN X, SHI J, LUO P, et al. Spatial as deep: Spatial CNN for traffic scene understanding[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 4536-4545.

[8] ZHENG T, FANG H, ZHANG Y, et al. RESA: Recurrent feature-shift aggregator for lane detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 3547-3554.

[9] HOU Y, MA Z, LIU C, et al. Learning lightweight lane detection CNNs by self attention distillation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2020: 1013-1021.

[10] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

[11] 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. Piscataway: IEEE Press, 2016: 779-788.

[12] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6517-6525.

[13] 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. Piscataway: IEEE Press, 2020: 2633-2642.

[14] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[EB/OL]. 2015: arXiv: 1502.03167. https://arxiv.org/abs/1502.03167.

[15] MA N, ZHANG X, LIU M, et al. Activate or not: learning customized activation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 8028-8038.

[16] JIANG Y, TAN Z, WANG J, et al. GiraffeDet: a heavy-neck paradigm for object detection[EB/OL]. 2022: arXiv: 2202.04256. https://arxiv.org/abs/2202.04256.

[17] 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. Piscataway: IEEE Press, 2017: 2999-3007.

[18] MILLETARI F, NAVAB N, AHMADI S A. V-net: fully convolutional neural networks for volumetric medical image segmentation [C]//Proceedings of the Fourth International Conference on 3D Vision. Piscataway: IEEE Press, 2016: 565-571.

[19] LOSHCHILOV I, HUTTER F. SGDR: stochastic gradient descent with warm restarts[EB/OL]. 2016: arXiv: 1608.03983. https://arxiv.org/abs/1608.03983.

[20] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. 2020: arXiv: 2004.10934. https://arxiv.org/abs/2004.10934.

[21] HAN C, ZHAO Q, ZHANG S, et al. YOLOPv2: better, faster, stronger for panoptic driving perception[EB/OL]. 2022: arXiv: 2208.11434. https://arxiv.org/abs/2208.11434.

[22] TEICHMANN M, WEBER M, ZÖLLNER M, et al. MultiNet: real-time joint semantic reasoning for autonomous driving[C]//Proceedings of the IEEE Intelligent Vehicles Symposium. Piscataway: IEEE Press, 2018: 1013-1020.

[23] 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, 2020, 21(11): 4670-4679.

[24] 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.

[25] VU D, NGO B, PHAN H. HybridNets: end-to-end perception network [EB/OL]. 2022: arXiv: 2203.09035. https://arxiv.org/abs/2203.09035

责任编辑:李翠薇