DOI: 10.13382/j.jemi.B2407565

面向无人驾驶的多任务环境感知算法研究*

宋建辉 刘 鑫 庄 爽 赵亚威 刘晓阳
(沈阳理工大学自动化与电气工程学院 沈阳 110159)

摘 要:在无人驾驶技术中,多任务环境感知算法是确保无人驾驶汽车,在复杂交通环境中安全运行的关键技术之一。针对现有环境感知算法在处理复杂驾驶场景时存在的因天气、光照、遮挡等因素导致的鲁棒性差,出现漏检、分割边界模糊等问题,基于 HybridNets 网络进行改进,提出了一种高性能混合网络 IPHNet,以更准确地完成实时感知任务。该网络使用解码器-编码器结构,采用改进 EfficientNetV2-S 作为主干网络,增强特征提取能力和处理速度。通过重构 BiFPN 来增加不同层级中间信息的特征融合,引入轻量化上采样模块 DySample 减少模型复杂度,保留更多信息。创新性设计了分割模块(SPN),极大地保证了对底层信息提取的完整性与准确性。在 BDD100K 数据集上的实验表明,与基线网络 HybridNets 相比,IPHNet 在车辆检测任务上 mAP 达到 81.4%,提高了 4.1%;车道线分割任务上准确率达到 86.84%,提高了 1.44%,IoU 达到 33.32%,提高了 1.72%;可行驶区域划分任务上 mIoU 提高了 1.8%;FPS 达到 34,验证了 IPHNet 具备一定实时处理能力。

关键词: 无人驾驶;多任务环境感知;车辆检测;车道线检测;可行驶区域划分
中图分类号: TP391.41; TN971.+1 **文献标识码**: A **国家标准学科分类代码**: 520.20

Research on multi-task environment perception algorithm for unmanned driving

Song Jianhui Liu Xin Zhuang Shuang Zhao Yawei Liu Xiaoyang
(School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China)

Abstract: In autonomous driving technology, multi-task environment perception algorithm is one of the key technologies to ensure the safe operation of driverless vehicles in complex traffic environments. In view of the poor robustness of the existing environment perception algorithms in dealing with complex driving scenarios caused by weather, illumination, occlusion and other factors, and the problems of missed detection, and blurred segmentation boundary, an improved high-performance hybrid network IPHNet based on HybridNets network is proposed to more accurately complete real-time perception tasks. This network uses a decoder-encoder structure and adopts an improved EfficientNetV2-S as the backbone network to enhance feature extraction capability and processing speed. By reconstructing BiFPN to increase the feature fusion of intermediate information of different levels, the lightweight up-sampling module DySample is introduced to reduce the complexity of the model and retain more information. The innovative design of the segmentation module (SPN) greatly ensures the integrity and accuracy of the underlying information extraction. Experiments on BDD100K dataset show that compared with the baseline network HybridNets, IPHNet achieves 81.4% mAP on vehicle detection tasks, which is improved by 4.1%. The accuracy of the lane line segmentation task reaches 86.84%, which is increased by 1.44%, and the IoU reaches 33.32%, which is increases by 1.72%. The mIoU of the drivable area division task is increased by 1.8%. The FPS reaches 34, which verified that IPHNet has a certain real-time processing capability.

Keywords: unmanned; multi-task environment awareness; vehicle detection; lane detection; drivable area division

收稿日期: 2024-05-31 Received Date: 2024-05-31
*基金项目:沈阳市中青年科技创新人才支持计划项目(RC210247)、辽宁省属本科高校基本科研业务费专项资金(LJ212410144053)项目资助

0　引　言

随着无人驾驶技术的快速发展,无人驾驶车辆的环境感知能力已成为关键的技术挑战之一。传统的环境感知算法只能完成单一任务如:目标检测、车道线分割和可行驶区域划分等[1]。其中目标检测算法分为两类,基于候选区域的两阶段目标检测算法和基于回归的单阶段目标检测算法[2]。代表性的两阶段算法为 Faster R-CNN[3],采用了 RPN(region proposal network)算法来生成候选区域,避免了传统 R-CNN[4]中的选择性搜索算法和繁琐的特征提取工作。单阶段目标检测应用较广的网络模型有 YOLO[5](you only look once)和 SSD,YOLO 系列算法经过多次改进,已发展到 YOLOv8 版本,将之前 YOLOv5 版本中的 C3 模块换成了 C2f 模块,并配合调整了模块的深度和训练策略。Liu 等[6]提出的 SSD 算法则采用了完全卷积的网络结构,将整张图像作为输入,直接输出边界框的位置和类别。

可行驶区域划分方面,Zhao 等[7]提出的 PSPNet 算法是基于全卷积网络[8](fully convolutional networks, FCN)聚合不同区域的上下文信息,采用了不同大小的池化核对特征图进行操作。在车道线分割算法中,Dong 等[9]提出的 ENet 通过减少解码器中的卷积操作,合理调整网络结构,以获得更好的图像输出效果。Parashar 等[10]提出了 SCNN 算法,在上下文信息和空间信息之间进行平衡,引入空间卷积核和分支卷积核相结合的方式。Hou 等[11]提出的 ENet-SAD 通过在网络自身执行自上而下和分层的注意力蒸馏网络进行表征学习。

随后,一些学者开始尝试整合多个感知任务成一个网络,Teichmann 等[12]提出的 MultiNet 是第一个采用编码器-解码器架构的多任务网络,以减少参数量。Qian 等[13]提出的 DLT-Net 设计了上下文张量,通过融合可驱动区域解码器与其他两个解码器的特征映射实现任务间的互信息共享。随后,Wu 等[14]提出的 YOLOP 成为在车辆检测、车道线分割、可行驶区域分割 3 个任务上第一个具备实时性的多任务检测算法。此后大量基于 YOLOP 的改进算法相继提出,牛国臣等[15]提出了 TDL-YOLO 利用交叉注意力模块得到分割和检测特征图。宋绍京等[16]提出的 MEPNet 设计了 FRFB 模块增大感受野,并在检测网络的头部添加小目标检测层,使得在车辆检测和可行驶区域划分这两个任务上达到了较好的效果。刘云翔等[17]将 YOLOP 中主干网络的部分卷积替换为感受野注意力卷积(RFAConv),在特征金字塔网络中,使用高效跨尺度融合模块替换原有的跨阶段层次模块。Wang 等[18]设计的 A-YOLOM 不同于 YOLOP 的结构,它使用 YOLOv8 的主干网络,并为每个任务分配单独的 Neck 及设计了一个对应的自适应模块,增加模型普适性。Vu 等[19]提出的 HybridNets 使用轻量级 EfficientNet-B3 作为主干网络,并把 YOLOP 中的两个相似的分割任务合并成一个任务进行优化,使得检测精度大幅提升同时计算量也有所下降。武鹏宇等[20]提出一种对 HybridNets 的改进算法,使用 EficientNetV2-S 主干网络,分配 3 个独立的解码器来解决不同难度的问题,并在主干网络与颈部之间加入 A2-Nets[21]双重注意力机制。HybridNets 以及上述这些的网络虽然使得模型精度有所提升,但是面对复杂道路中的关键信息提取仍不充分,使得信息在层层传递过程中出现重要细节和边缘信息的严重丢失,最终导致被遮挡目标和图像边缘目标发生漏检。针对以上所存在的问题,对 HybridNets 网络进行改进,提出了一种高性能混合网络(improvement performance hybridnets, IPHNet),以更加实时准确的完成多任务感知。该研究所做贡献如下:

1)在特征提取时使用改进 EfficientNetV2-S 作为主干网络,该网络在 EfficientNetV2 网络的基础上将 MBconv 中的 SE 模块替换为 CA 模块,提高了 IPHNet 对有用空间坐标信息的准确定位,以及关键信息的提取能力;

2)在颈部网络上,对 BiFPN 进行重构和改进,增加了不同层级中间信息的特征融合,使得不同尺度的信息利用率更高,并结合轻量化上采样模块,在保留尽可能多的信息情况下,减少模型复杂度,提升检测精度;

3)创新性设计了一个分割模块(SPN),该模块通过注意力机制和特征融合技术,对具备高分辨率 P2 特征图中的细节信息进一步处理,以最大程度上保证信息的完全性与准确性。

1　网络模型介绍

1.1　IPHNet 整体网络

对 HybridNets 算法进行改进,提出一个新的网络 IPHNet,使得在面对无人驾驶场景中的复杂交通情况下,能够实时准确地感知周围环境。改进后的网络如图 1 所示。

该网络采用改进的 EfficientNetV2-S 作为主干网络,对特征进行提取后送入颈部网络中。颈部网络在原 BiFPN 基础上,增加了多级跨尺度信息的融合,并引入了轻量化 DySample 的上采样模块,这一设计使网络能获取完整准确的特征信息的同时,保证了轻量化。在分割头部分,将经过 SPN 处理后的 P2 层与经过特征金字塔处理的 P3~P7 层进一步融合,显著提升了分割的精度和效果。在检测头部分,该模块接收来自颈部网络的各尺度特征图,并对每个锚点预测生成回归值,根据预测的边界框偏移量和各类别的概率计算预测框的置信度,以此进

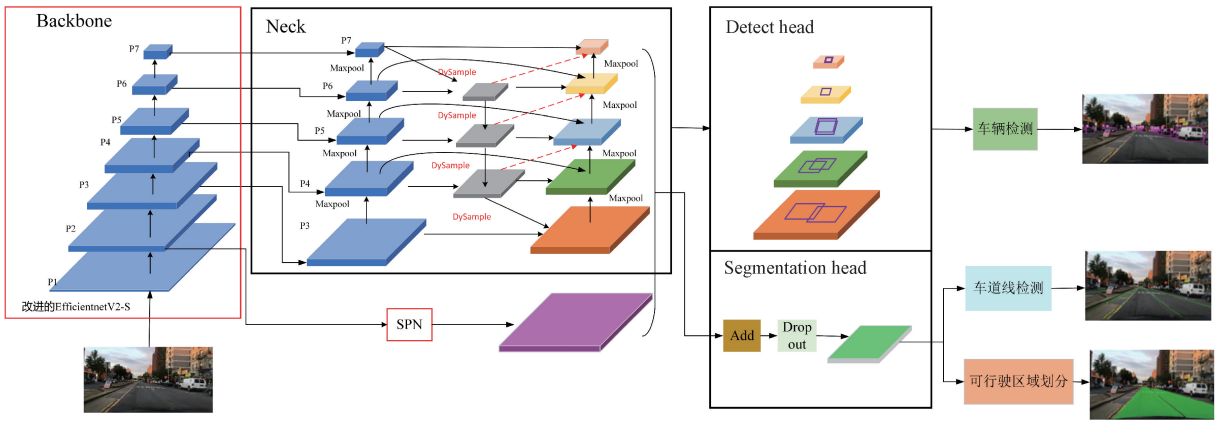


图 1　IPHNet 网络整体结构图

Fig. 1　The overall structure of the IPHNet model

行精准定位。该研究所提出的网络充分利用了改进主干网络的高效特征提取能力和特征金字塔结构的多尺度融合优势以及 SPN 模块对于底层信息的细节处理，大幅提高了识别的准确率和处理速度，进而增强了系统的整体性能和可靠性。

1.2　改进 EfficientNetV2-S 主干网络

原 HybridNets 网络采用 EfficientNet-B3 作为主干网络，其所使用的 MBConv 模块有着相对复杂的结构，使得网络参数量较大，训练速度较慢，而且 MBConv 中的 SE[22]（squeeze-and-excitation）模块只关注了通道间的信息，忽略了空间位置信息。为了解决这些问题，引入了改进的 EfficientNetV2-S[23] 作为新的主干网络。在浅层采用了 Fused-MBConv 模块，该模块将 MBConv 结构主分支中的 1×1 卷积和深度卷积（DWConv k×k）合并为一个 3×3 卷积，并且删除了注意力模块，以此减少了模型的参数量，降低过拟合风险。主干网络如图 2 所示，每层后面的数字指示堆叠的层数。

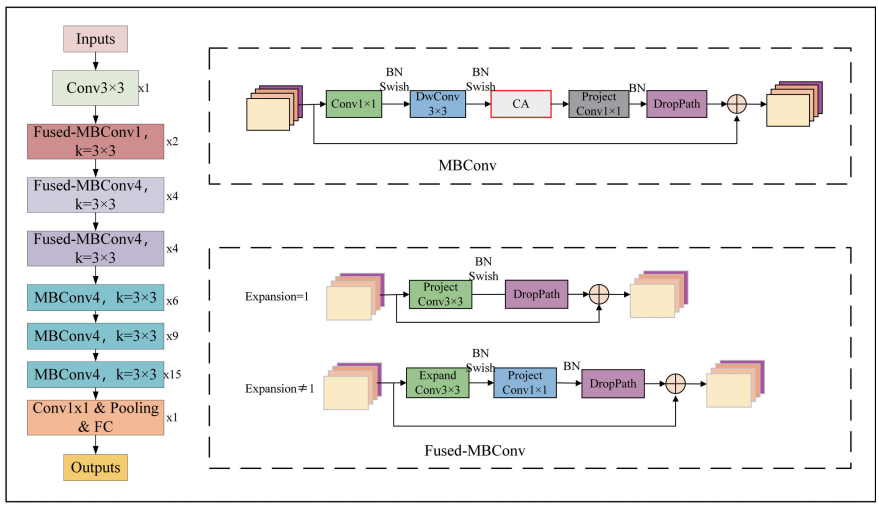


图 2　改进的 EfficientNetV2 主干网络结构图

Fig. 2　Diagram of the improved EfficientNetV2 backbone network

同时，该研究还引入了坐标注意力机制 CA[24]（coordinate attention），替换原有的 SE 模块。CA 通过编码横向和纵向的位置信息到通道注意力中，使网络能够自适应地学习关键信息，从而提高了该网络对有用通道信息的准确定位，而且不会显著增加计算负担，提升了系统对环境变化的适应性和决策的准确性。CA 注意力机制结构如图 3 所示。

具体操作分为坐标信息嵌入和坐标注意力生成 2 个步骤。为了捕获精准位置信息，坐标信息嵌入对尺寸为 $C \times H \times W$ 输入特征图 Input 分别按照 X 方向和 Y 方向进

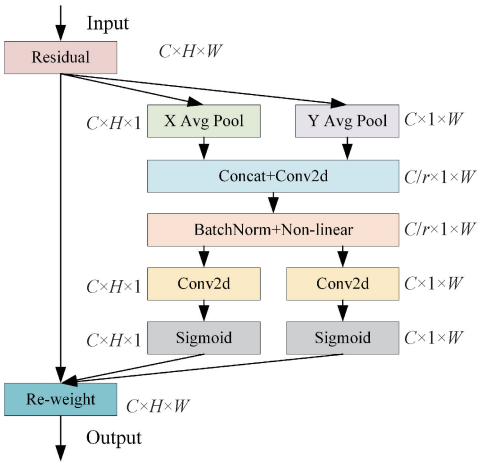


图 3　CA 注意力机制结构图

Fig. 3　Structure diagram of CA attention mechanism

行池化,分别生成尺寸为 $C \times H \times 1$ 和 $C \times 1 \times W$ 的特征图,具体过程如式(1)~(3)所示。

$$z_c = \frac{1}{H \times W} \sum_{W}^{H} \sum_{H}^{W} x_c(i,j) \tag{1}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leqslant j < H} x_c(j,w) \tag{2}$$

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leqslant i < W} x_c(h,i) \tag{3}$$

式中:z_c 表示与第 c 信道相关联的输出,$z_c^h(h)$ 为高度为 h 的第 c 通道的输出,$z_c^w(w)$ 为宽度为 w 的第 c 通道的输出。将生成的 $C \times H \times 1$ 和 $C \times 1 \times W$ 的特征图进行拼接操作,将拼接好的特征图进行卷积操作,随后将数据批量归一化送入 ReLU 激活函数生成中间特征图 f,公式如式(4)所示。

$$f = \delta(F_1([Z^H, Z^W])) \tag{4}$$

式中:F_1 为卷积操作函数,δ 为 ReLU 激活函数,$f \in R^{\frac{c}{r} \times (H+W)}$,表示对空间信息在水平方向和垂直方向进行编码的中间特征映射,然后沿着空间维度将特征图 f 分成两个张量,$f \in R^{\frac{c}{r} \times H}$ 和 $f \in R^{\frac{c}{r} \times W}$,经过通道数转换后送入 Sigmoid 函数得到:

$$g^h = \sigma(F_h(f^h)) \tag{5}$$

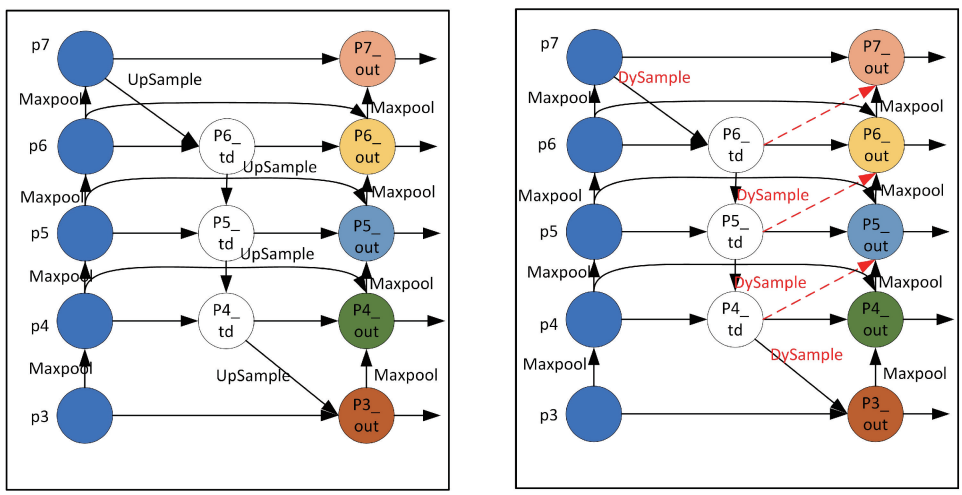
$$g^w = \sigma(F_w(f^w)) \tag{6}$$

式中:g^h 为水平方向权重,g^w 为垂直方向权重,F_h、F_w 为通道数转换函数,σ 为 Sigmoid 激活函数。最后经过对输入 $x_c(i,j)$ 进行加权计算,得到坐标注意力输出:

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{7}$$

1.3　重构并改进 BiFPN

BiFPN[25] 在传统的特征金字塔网络(FPN)的基础上,通过增加自顶向下和自底向上的连接和路径,实现了信息的双向流动。然而,BiFPN 在处理中间层次特征时,未能充分整合这些特征,导致训练过程中关键信息的遗漏,这种遗漏尤其影响那些处于高层与低层之间的重要特征,进而影响整体的检测准确率。此外,为了匹配特征图的尺寸,所需进行的多次上采样会引起信息失真且增加计算成本,削弱了模型在捕捉目标细节上的能力,最终对检测性能产生负面影响。针对这些问题,对 BiFPN 进行了重构和改进。如图 4(a)所示为原 BiFPN 结构图,图 4(b)所示为重构和改进的 BiFPN 结构图,在 BiFPN 网络的基础上,增加了从 P_6^{td} 到 P_7^{out}、P_5^{td} 到 P_6^{out} 以及 P_4^{td} 到 P_5^{out} 的信息传递路径[26],这些改进旨在加强中间信息的有效整合,提高特征的空间分辨率,并优化信息流的连贯性。



(a) 原BiFPN结构图
(a) Original BiFPN structure diagram

(b) 重构并改进的BiFPN结构图
(b) Reconstructed and improved BiFPN structure diagram

图 4　原 BiFPN 结构与重构并改进的 BiFPN 结构对比图

Fig. 4　Comparison between the original BiFPN structure and the reconstructed and improved BiFPN structure

通过这样的结构调整，不仅提高了网络对复杂场景中多尺度特征的处理能力，还增强了对不同层次特征的捕捉，提升了系统的泛化能力。中间信息为：

$$P_6^{td} = Conv\left(\frac{\omega_1 \cdot P_6 + \omega_2 \cdot Resize(P_7)}{\omega_1 + \omega_2 + \varepsilon}\right) \tag{8}$$

$$P_5^{td} = Conv\left(\frac{\omega'_1 \cdot P_5 + \omega'_2 \cdot Resize(P_6)}{\omega'_1 + \omega'_2 + \varepsilon}\right) \tag{9}$$

$$P_4^{td} = Conv\left(\frac{\omega''_1 \cdot P_4 + \omega''_2 \cdot Resize(P_5)}{\omega''_1 + \omega''_2 + \varepsilon}\right) \tag{10}$$

经过进一步信息融合以及下采样得到各层级输出，使得：

$$P_7^{out} = Conv\left(\frac{\omega_1 \cdot P_7 + \omega_2 \cdot P_6^{td} + \omega_3 \cdot Resize(P_6^{out})}{\omega_1 + \omega_2 + \omega_3 + \varepsilon}\right) \tag{11}$$

$$P_6^{out} = Conv\left(\frac{\omega'_1 \cdot P_6 + \omega'_2 \cdot P_6^{td} + \omega'_3 \cdot P_5^{td} + \omega'_4 \cdot Resize(P_5^{out})}{\omega'_1 + \omega'_2 + \omega'_3 + \omega'_4 + \varepsilon}\right) \tag{12}$$

$$P_5^{out} = Conv\left(\frac{\omega''_1 \cdot P_5 + \omega''_2 \cdot P_5^{td} + \omega''_3 \cdot P_4^{td} + \omega''_4 \cdot Resize(P_4^{out})}{\omega''_1 + \omega''_2 + \omega''_3 + \omega''_4 + \varepsilon}\right) \tag{13}$$

式中：P_i^{td} 为各级中间信息，P_j^{out} 为各级输出信息，$Resize$ 是一个上采样或下采样以匹配分辨率的操作，ω 为可学习权重，ε =0.000 1 用来保持数值稳定。

原网络采用最近邻插值法进行上采样，将待插值点周围最邻近的像素值作为插值结果，以扩大特征图尺寸。这种方法仅根据像素位置进行操作，未能充分利用特征图中的内容信息。为了更有效地感知边缘信息并减少计算负担，本网络采用了 DySample 作为上采样算子，这是一个极其轻量化的模块，与 CARAFE[27] 中学习得来的上采样核不同，DySample[28] 从点采样的角度出发，基于生成的上采样位置实现上采样，通过动态计算的偏移量，灵活调整输入特征的采样位置，其上采样核的权重依赖于网格坐标位置。通过使用 DySample，网络不仅能够以极低的计算成本有效扩大感受野，还提高了图像处理的效率和质量，特别是在检测图像中有高速移动的车辆或远处及边缘的车道线等细节时表现出色。DySample 网络结构如图 5 所示。

给定大小为 $C \times H \times W$ 的特征映射 X，经过一个采样点发生器，生成大小为 $2g \times sH \times sW$ 的采样集 S，其中第一维的 $2g$ 表示网格 x 和 y 坐标，网格样本函数使用 S 中的位置，将 X 重新采样到大小为 $C \times sH \times sW$ 的 X' 中。过程为：

$$X' = gridsample(X, S) \tag{14}$$

其中，采样点发生器计算过程为：对于给定上采样比

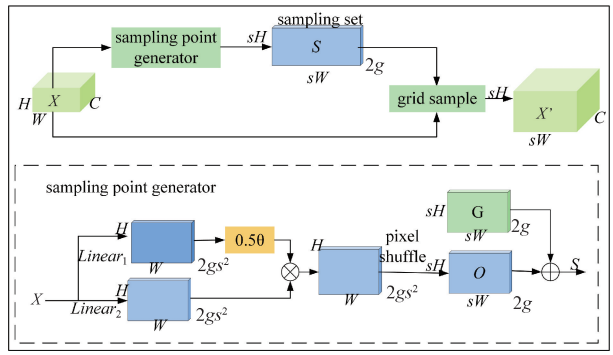


图 5　DySample 网络结构图

Fig. 5　DySample network structure

例因子 s 和尺寸为 $C \times H \times W$ 的特征映射 X，使用输入和输出通道分别为 C 和 $2gs^2$ 的线性层，生成尺寸为 $2gs^2 \times H \times W$ 的偏移量 O，为了增加偏移量的灵活性，进一步通过对输入特征的线性映射，生成“动态范围因子 θ”，来调整原始偏移量的规模和方向，乘以 0.5 既防止过大的空间位移可能导致的图像失真，又保证不会由于重叠影响边界附近的预测。最终偏移量动态操作过程为：

$$O = 0.5\theta(linear_1(X)) \cdot linear_2(X) \tag{15}$$

随后通过像素重排，将其重塑为 $2g \times sH \times sW$，采样集 S 为偏移量 O 与原始采样网格 G 之和。过程为：

$$S = G + O \tag{16}$$

1.4　设计 SPN 模块

P2 层特征图没有参与特征金字塔的融合，但其富含丰富的语义信息，使用普通的特征提取技术无法充分挖掘其中的语义信息。因此，设计了 SPN 模块对 P2 特征图进行处理，用来在繁杂的道路环境中进行更加全面的信息获取。如图 6 所示，本模块先是经过静态填充以扩展特征图的边界信息，随后依次通过 ZLC 模块、特征融合模块 PagFM 和上采样模块，来逐步提升特征的质量并进行细化。其中 ZLC 模块是将卷积与通道注意力机制相结合，注意力机制的具体实现过程为：

$$AttentionOutput = X \cdot Sigmoid(FC(AvgPool(X)) + FC(MaxPool(X))) \tag{17}$$

这样，通过将计算得到的通道注意力权重对原始通道进行加权，增强了网络对于重要特征的敏感性，使得网络能够自适应地聚焦于更有助于提升性能的特征。

在 PagFM[29] 中，I 分支接收来自 ZLC 处理后的准确语义信息，而 P 分支保留原始的细节和边界信息。为了更好地结合二者，首先对 I 分支和 P 分支中对应像素的向量 x、y 分别预处理，再将经过预处理的 y 上采样调整到与 x 相同的空间尺寸，得到 y'。然后 x 与 y 经过 Sigmoid 函数处理后得到 σ。

$$\sigma = Sigmoid(f(x) \cdot y') \tag{18}$$

其中，σ 表示这两个像素属于同一对象的可能性。如果 σ 很高，则更相信 I 分支中的语义特征，否则加入更多的 P 分支信息。最后使用这个相似性来加权融合 x 和 y'，得到最终输出，计算公式为：

$$Out = \sigma \times y' + (1 - \sigma) \times x \tag{19}$$

由此，根据学习到的相似性来动态调整两个分支的融合权重，自适应地决定在每个位置保留多少 I 分支特征的同时引入多少 P 分支的原始特征。

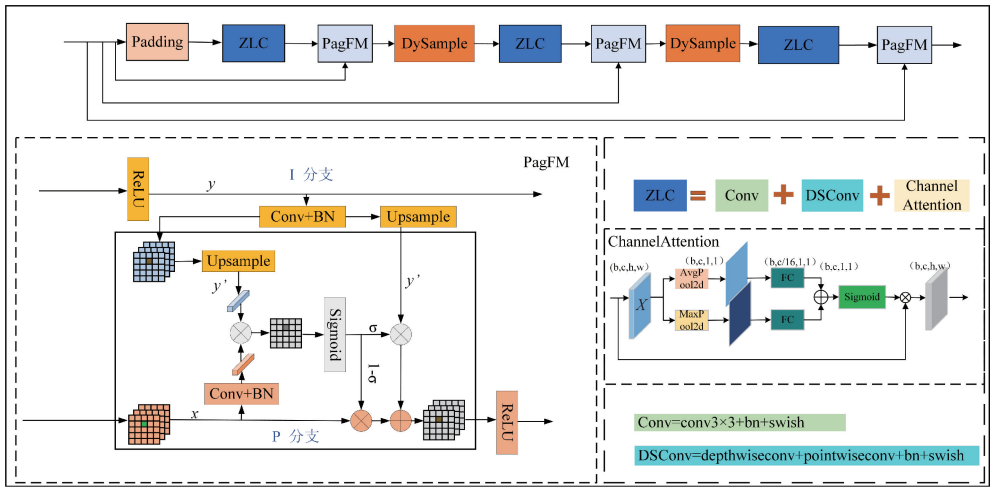


图 6　SPN 模块结构图

Fig. 6　Structure of the SPN module

2　损失函数和实验分析

2.1　损失函数

为了 IPHNet 同时完成车辆检测、车道线分割、可行驶区域分割 3 种任务，该研究使用了一种混合损失。整体网络损失函数如式(20)所示。

$$L_{all} = \alpha L_{det} + \beta L_{seg} \tag{20}$$

其中，L_{det}、L_{seg} 表示目标检测损失和分割损失。α 和 β 是用于平衡检测损失和分割损失的调整参数。检测损失进一步分为 3 个组成部分：

$$L_{det} = \alpha_1 L_{class} + \alpha_2 L_{obj} + \alpha_3 L_{box} \tag{21}$$

α_1、α_1、α_1 是额外的调整参数，其中，L_{class} 是分类损失，L_{obj} 是目标损失，L_{box} 是边界框回归损失。分类损失(L_{class})和目标损失(L_{obj})，这两个损失函数都是采用焦点损失(Focal Loss)实现的，Focal Loss[30] 函数的目的是通过减小损失函数的斜率，专注于难以分类的样本，进而提高分类和目标检测的准确性。边界框回归损失通过平滑的 L1 损失[31](smooth L1 loss)来计算，平滑 L1 损失在预测框和真实框之间取绝对差值，避免了梯度爆炸问题。具体表达式如式(22)所示。

$$smooth\ \ L1(x) = \begin{cases} \delta_1 x^2, x < \delta_2 \\ x - \delta 1, x > \delta_2 \end{cases} \tag{22}$$

其中，x 表示预测框与真实框之间的距离计算公式。

$$x = \delta b_p (|b_x - \hat{b}_x| + |b_y - \hat{b}_y| + |b_w - \hat{b}_w| + |b_h - \hat{b}_h|) \tag{23}$$

b_x、b_y、b_w、b_h 分别表示预测框的坐标和尺寸，$\hat{b}_x$、$\hat{b}_y$、$\hat{b}_w$、$\hat{b}_h$ 则表示真实框的坐标和尺寸，δ 是一个缩放因子，b_p 表示一个正标签已被分配给某个网格单元。为了让回归网络顺利学习，强制设置了一些锚框的大小，使得这些锚框可以被平滑地学习。由此可得：

$$b_p = \begin{cases} 1, & IoU(c_i, b_j) \geqslant 0.5 \quad i = 1, \cdots, \sum_{k=1}^{s} n_k m_k; j = 1, \cdots, p \\ 0, & IoU(c_i, b_j) < 0.5 \quad i = 1, \cdots, \sum_{k=1}^{s} n_k m_k; j = 1, \cdots, p \end{cases} \tag{24}$$

其中，$IoU(c_i, b_j)$ 表示第 i 个锚框 c_i 与真实框 b_j 之间的交并比。如果 IoU 值大于等于 0.5，则认为该锚框是一个正样本(即 $b_p = 1$)，否则被视为负样本(即 $b_p = 0$)。$\sum_{k=1}^{s} n_k m_k$ 表示所有特征图的锚框总数，$n_k m_k$ 是第 k 层特征图的分辨率，p 是每张输入图像中真实边界框的总数。

其次，L_{seg} 是一种多类混合的分割损失函数，主要用于背景、可行驶区域和车道线的多类别分割任务。在语义分割任务中，由于数据分布不平衡导致的小目标分割是一个常见的挑战。为了解决这些挑战，Tversky Loss($L_{Tversky}$)和 FocalLoss(L_{Focal})结合使用，以预测每个像素所属的类别。$L_{Tversky}$ [32] 对类别不平衡问题有良好的

表现,特别是针对那些样本数量较少的类别或小目标,它能够更好地优化这些类别的分割得分。L_{Focal} 旨在最小化像素间的分类错误,并对难分类的标签给予更多关注,减少易分类样本对整体损失的影响,从而提高模型对难分类样本的关注度。$L_{Tversky}$ 定义如式(25)所示。

$$L_{Tversky} = 1 - \frac{TP(c)}{TP(c) + \alpha \cdot FN(c) + (1-\alpha) \cdot FP(c)} \tag{25}$$

其中, $TP(c) = \sum_{n=1}^{N} P_n(c) g_n(c)$,表示类 c 的真正例; $FN(c) = \sum_{n=1}^{N} (1 - Pn(c)) gn(c)$,表示类 c 的假负例; $FP(c) = \sum_{n=1}^{N} Pn(c)(1 - gn(c))$,表示类 c 的假正例。α 是一个平衡参数,用来调整 $FN(c)$ 和 $FP(c)$ 的权重。$P_n(c)$ 是像素 n 属于类 c 的预测概率, $g_n(c)$ 是像素 n 在真实标注中属于类 c 的指示变量。L_{Focal} 的具体公式为:

$$L_{Focal} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{C} g_n(c)(1 - P_n(c))^{\gamma} log(P_n(c)) \tag{26}$$

其中,γ 是一个调节参数,用来控制难分类样本的权重。C 是类别的数量。N 是图像中像素的总数。最终分割总损失为:

$$L_{seg} = L_{Tversky} + \lambda_{LFocal} \tag{27}$$

其中,λ 是一个权重系数,用来平衡这两种损失函数的贡献。

2.2 实验环境和数据集介绍

本项目所用实验环境如表 1 所示。

表 1 实验环境

Table 1 Experimental environment

类别	版本
Operating system	Ubuntu20.04
CPU	12 核/GPU,Xeon(R) Platinum 8255C
GPU	Tesla V100(32 GB)
Pytorch	1.13.1
GPU acceleration CUDA	CUDA11.8
Programming language	Python3.8

使用 Nadam 优化器,训练 200 轮,batch_size 设为 16,Mosaic 默认 1.0,dropout 默认 0.2。采用伯克利大学制作的 BDD100K[33] 开源驾驶数据集,涵盖不同时间、不同天气条件包括晴天、阴天和雨天以及不同的驾驶场景。

2.3 实验与结果分析

该研究对多个环境感知任务的网络进行了评估,包括车辆检测、车道线分割和可行驶区域划分,检测任务中采用平均精度均值 mAP、召回率 Recall 作为评估指标,使用 mIoU 评价可行驶区域划分的实验结果,准确率 Accuracy 和 Lane IoU 作为车道线评价指标。

根据表 2 所示的车辆检测结果,该网络在 mAP 和召回率 Recall 这两个关键性能指标上表现良好,仅次于最佳表现的 MEPNet。具体来说,IPHNet 网络性能指标要明显好于单任务网络;由于使用改进的 EfficientNetV2 和重构并改进 BiFPN,使得 IPHNet 与基线网络 HybridNets 相比,在 mAP 上有 4.1%的增长,在 Recall 上也有 0.3%的提高;与文献[20]相比,在车辆检测任务中的性能指标均优于文献[20]。虽然在车辆检测的对比结果中,IPHNet 与 MEPNet 相比略显不足,但是与基线网络相比在车辆检测任务中实现了显著的性能提升。

表 2 车辆检测结果对比

Table 2 Comparison of vehicle inspection results

网络	任务类型	*mAP/%*	*Recall/%*	*FPS*
Faster R-CNN[3]	单任务	64.9	77.2	5.3
YOLOv8s	单任务	81.3	91.9	110
MultiNet[12]	多任务	60.2	81.3	8.6
DLT-Net[13]	多任务	68.4	89.4	9.3
YOLOP[14]	多任务	76.5	89.2	41.0
TDL-YOLO[15]	多任务	78.0	88.6	36.5
MEPNet[16]	多任务	**83.3**	**95.5**	–
文献[17]	多任务	78.9	90.7	31.7
A-YOLOM(s)[18]	多任务	81.1	86.9	96.2
HybridNets[19]	多任务	77.3	92.8	29
文献[20]	多任务	79.8	91.8	45
IPHNet	多任务	81.4	93.1	34

表 3 为不同网络在可行驶区域划分任务中的实验对比结果,采用 mIoU 指标对各网络的分割效果进行了评估。由表 3 可以看出,MEPNet 和 IPHNet 在可行驶区域分割任务上显示了高准确率,尽管 MEPNet 在可行驶区域划分任务中效果最佳,但 IPHNet 网络以 92.3% 的 mIoU 紧随其后,相较于基线网络 HybridNets,提升了 1.8%,与文献[20]相比 mIoU 提高了 0.5%,这一结果证明了 IPHNet 在可行驶区域分割方面的优势。

表 3 可行驶区域划分结果对比

Table 3 Comparison of drivable area division results

网络	任务类型	*mIoU/%*	*FPS*
PSPNet[7]	单任务	89.6	11.1
MultiNet[12]	多任务	71.6	8.6
DLT-Net[13]	多任务	71.3	9.3
YOLOP[14]	多任务	91.5	41.0
TDL-YOLO[15]	多任务	91.4	36.5
MEPNet[16]	多任务	**92.5**	–
文献[17]	多任务	92.2	31.7
A-YOLOM(s)[18]	多任务	91.0	96.2
HybridNets[19]	多任务	90.5	29
文献[20]	多任务	91.8	42
IPHNet	多任务	92.3	34

表 4 是不同网络在车道线分割任务上的性能比较。通过表中对比结果可以看出,单任务网络在准确率和 IoU

指标上的表现相对较低,这主要是因为它们只处理单一任务而缺乏在实际场景中不同感知任务间的信息共享。多任务网络 MEPNet 虽然在车辆检测以及可行驶区域划分任务中表现出较好的性能,但是在车道线分割任务中,远不如 IPHNet 网络的 86.84%准确率和 33.32%的 IoU,这会导致 MEPNet 在路况复杂的地区,不能够准确的识别车道线,使得无人驾驶车辆做出错误的判断,造成无法想象的后果。与多任务网络文献[20]相比,IPHNet 的准确率和 IoU 分别提高了 0.57%和 0.77%。综合前两个任务的评价指标,IPHNet 在处理多任务时保持高性能的同时更具平衡性,更适合应用于无人驾驶场景下的环境感知任务。

表 4　车道线分割结果对比

Table 4　Comparison of lane line segmentation results

网络	任务类型	*Accuracy*/%	*Lane IoU*/%	*FPS*
ENet[9]	单任务	34.12	14.64	100.0
SCNN[10]	单任务	35.79	15.84	19.8
ENet-SAD[11]	单任务	36.56	16.02	50.6
YOLOP[14]	多任务	70.50	26.20	41.0
TDL-YOLO[15]	多任务	72.30	26.50	36.5
MEPNet[16]	多任务	76.50	27.20	-
文献[17]	多任务	72.70	26.60	31.7
A-YOLOM(s)[18]	多任务	84.90	28.8	96.2
HybridNets[19]	多任务	85.40	31.60	29
文献[20]	多任务	86.21	32.55	43
IPHNet	多任务	**86.84**	**33.32**	34

2.4　消融实验与实时性能比较

为了进一步验证 IPHNet 网络中各部分模块的有效性,在 BDD100 K 数据集上进行了一系列的消融实验。这些实验都是在相同的实验条件下进行的,以基线网络 HybridNets 为基准,逐一加入 EfficientNetV2-S、改进的 EfficientNetV2-S、重构 BiFPN、重构并改进 BiFPN 和 SPN 模块。实验结果如表 5 所示。

由表 5 实验结果可知,从第 2、3 组实验看出:与在基线网络中引入 EfficientNetV2-S 相比,加入改进后的 EfficientNetV2-S,各项感知任务的评价指标均有所提升,表明了改进的 EfficientNetV2-S 能够加强对关键信息的提取能力。比较 4、5 组实验,在基线网络中逐步加入重构 BiFPN、重构并改进 BiFPN 的实验结果表明,重构并改进 BiFPN 进一步提升不同语义级别的特征融合,可以有效改善车辆检测性能。第 6 组实验中只加入 SPN 模块,该模块用于分割任务中,增强对具有高分辨率底层信息的利用,使得可行驶区域的 mIoU 指标提升 0.6%,车道线分割指标 Accuracy 和 Lane IoU 分别提升 0.21%和 0.27%。第 7 组实验,将改进后的 EfficientNetV2-S 与重构并改进 BiFPN 相结合,使得整体指标大幅增长,尤其是车辆检测的两个指标,与基线网络相比,分别提升了 4.3%和 0.3%,突出了这两个模块对于整体网络性能的提升以及泛化能力。第 8 组实验也就是本研究的 IPHNet 网络,在第 7 组实验的基础上加入了 SPN 模块,使得两大分割任务的指标再次提升,但是车辆检测任务中的 mAP50 却有所下降,这是因为随着模型复杂度的增加,所需的计算资源也在增加,使得最终权衡指标时,倾向于分割任务,导致了 mAP50 指标的小幅下降。

表 5　消融实验结果对比

Table 5　Comparison of ablation experimental results　　(%)

组别	EfficientNetV2-S	改进的 EfficientNetV2-S	重构 BiFPN	重构并改进 BiFPN	SPN	*mAP*50	*Recall*	*mIoU*	*Accuracy*	*Lane IoU*
1						77.3	92.8	90.5	85.40	31.60
2	√					78.5	92.8	90.8	85.55	31.89
3		√				79.2	92.9	91.2	85.63	32.08
4			√			78.2	92.8	90.9	85.46	31.86
5				√		78.5	92.9	91.1	85.60	31.92
6					√	77.3	92.8	91.1	85.61	31.87
7		√		√		**81.6**	**93.1**	91.7	86.74	32.62
8		√		√	√	81.4	**93.1**	**92.3**	**86.84**	**33.32**

实时性能参数比较如表 6 所示,表中展示了两个代表性网络与 IPHNet 网络在参数量(Params)、计算量(FLOPs)和每秒传输帧数(FPS)上的对比数据。通过与 HybridNets 相比较,IPHNet 虽然参数量和计算量都有提升,但是由于使用轻量化的主干网络和上采样模块,并配以有效的网络设计和训练策略,使得 FPS 升高即实时性能更好,说明了网络设计的合理性。在与 YOLOP 的参数比较中,IPHNet 虽然实时性能略差,但是仍然具有实时处理能力,同时,通过表 2、3 和 4 的对比实验可以看出,IPHNet 在检测和分割任务上的精度都明显高于 YOLOP,综合性能更强。

表 6　实时性能参数比较

Table 6　Comparison of real-time performance of the model

网络	*Params*/M	*FLOPs*/B	*FPS*
YOLOP[14]	7.9	18.6	41
HybridNets	12.83	15.6	29
IPHNet	24.67	27.3	34

2.5　可视化结果分析

为了更直观的验证 IPHNet 网络模型的有效性，在 BDD100K 数据集上对 YOLOP、HybridNets 以及 IPHNet 进行了可视化比较，其实验结果可视化如图 7~9 所示。

图 7 所示为同样不同网络在白天时段的检测结果，图 7(a)为原图，图 7(b)列出了 YOLOP 的 3 种场景，第 1 种场景中存在将不可行驶区域误检测为可行驶区域，第 2 种场景中出现了车道误检以及可行驶区域误检的情况，第 3 种场景中出现了远处车辆漏检和可行驶区域缺失；图 7(c)是基线网络 HybridNets 的检测结果，第 1 个场景中存在可行驶区域缺失的问题，第 2 种场景中车道线分割不清晰，第 3 种场景中丢失了一定的可行驶区域；图 7(d)是 IPHNet 的检测结果，对于 3 个场景均能良好的完成任务，说明了 IPHNet 网络具有良好的性能。

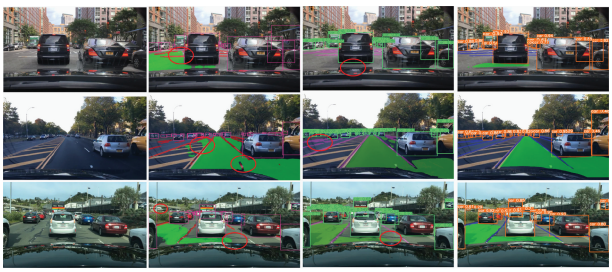


图 7　白天时段可视化效果图

Fig. 7　Visualization of daytime period

图 8 所示为雨雪天气下的检测结果，图 8(a)为原图，图 8(b)中 YOLOP 的检测结果出现较大的误差，在第 1 种场景中出现了将车前机盖错误的识别为可行驶区域，第 2 种场景中错误的将车辆前进方向左侧的积雪识别成可行驶区域，第 3 种场景下则是将左侧对向车道识别为了可行驶区域以及右侧车道识别不完全；在图 8(c) HybridNets 中的第 1 种场景中可行驶区域形状不完整，第 2 种场景下存在车辆漏检的问题，在第 3 种场景中出现了远处物体误检为车辆以及右侧车道线漏检问题；而图 8(d) IPHNet 第 2 种场景中能够完整的检测出所有车辆，同时也没有出现车道线和可行驶区域漏检、误检现象。在雨雪天气中 IPHNet 整体效果明显要优于其他两个网络。

图 9 所示为同在夜间时段时的可视化效果，图 9(a)为原图，图 9(b)的 YOLOP 网络的检测结果表明，在第 1、2 种场景中存在一定的可行驶区域缺失问题，第 3 个场景中出现了大量车辆漏检问题；图 9(c)的 HybridNets 网络中第 1 种场景中出现了远处车辆漏检的现象，在第 3 种场景中存在车道线丢失问题；在图 9(d)的 IPHNet 网络的检测结果中，虽然在第 1 种场景中也存在一小部分可行区域缺失问题，但是明显比前两个网络的检测效果好很多。从第 2、3 场景中可以看出 IPHNet 网络在夜间光线不好的情况下，仍能够表现出良好的检测与分割效果。

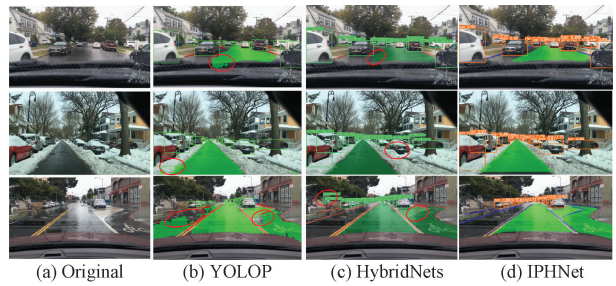


图 8　雨雪天气的效果图

Fig. 8　Effect of rain and snow weather

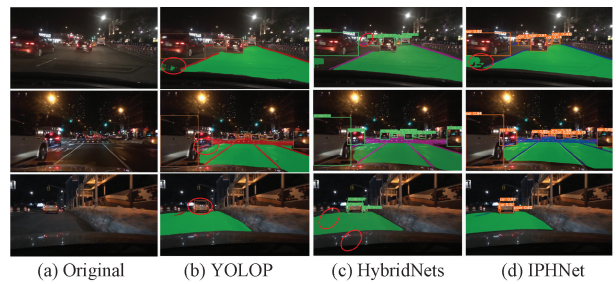


图 9　夜间时段可视化效果图

Fig. 9　Visualization rendering of the nighttime period

3　结　论

本研究基于 HybridNets 进行改进，提出了一种高性能的网络模型——IPHNet。核心思路在于通过高效的特征提取与创新性的模块设计，全面提升模型在复杂场景中的表现力。IPHNet 采用结合 CA 注意力机制的 EfficientNetV2-S 作为主干网络，同时对 BiFPN 进行重构和加入轻量化上采样模块，使得 IPHNet 不仅提升了特征提取效率，还显著提高了处理复杂驾驶场景的鲁棒性。此外，加入创新性设计的 SPN 分割模块，进一步提高了从高分辨率图像中提取细节和准确信息的能力。在 BDD100 K 数据集上的实验结果表明，IPHNet 在车道线分割、车辆检测以及可行驶区域分割等关键任务上均表现出卓越的性能，且具有较高的实时性。在后续的研究工作中，尽量保持检测精度的同时进一步网络轻量化，使其在资源有限的设备中也能正常运行。

参考文献

[1]　冯明驰，卜川夏，萧红. 面向 AR-HUD 的多任务卷积神经网络研究 [J]. 仪器仪表学报，2021，42(3)：41-250.

FENG M CH，BU CH X，XIAO H. Research on multi-task

convolutional neural network facing to AR-HUD [J]. Chinese Journal of Scientific Instrument, 2021, 42 (3): 241-250.

[2] 梁继然,陈壮,董国军,等. 结合注意力机制和密集连接网络的车辆检测方法[J]. 电子测量与仪器学报, 2022,36(3): 210-216.
LIANG J R, CHEN ZH, DONG G J, et al. Vehicle detection method combining attention mechanism and dense connection network [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36 (3): 210-216.

[3] REN SH 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, 2016, 39(6): 1137-1149.

[4] 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, 2014: 580-587.

[5] REDMON J, DIVVALA S, GIRSHICK R,et al. You only look once: Unified, real-time object detection [C]. Computer Vision & Pattern Recognition. IEEE, 2016.

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[7] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:2881-2890.

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

[9] DONG CH X. Image semantic segmentation using improved enet network [J]. Journal of Information Processing Systems, 2021, 17(5): 892-904.

[10] 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.

[11] HOU Y N, MA ZH, LIU CH X, et al. Learning lightweight lane detection cnns by self attention distillation [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1013-1021.

[12] TEICHMANN M, WEBER M, ZOELLNER M, et al. Multinet: Real-time joint semantic reasoning for autonomous driving[C]. 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018: 1013-1020.

[13] 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.

[14] 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.

[15] 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]. 北京航空航天大学学报, 2024, 50 (5): 1491-1499.
NIU G CH, WANG X N. A multi-task traffic scene detection model based on cross-attention[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50 (5): 1491-1499.

[16] 宋绍京, 陆婷婷, 孙翔, 等. 面向自动驾驶的多任务环境感知算法[J]. 电子测量技术, 2023,46(24): 157-163.
SONG SH J, LU T T, SUN X, et al. Multi-task environment perception algorithm for autonomous driving [J]. Electronic Measurement Technology, 2023, 46(24): 157-163.

[17] 刘云翔, 马海力, 朱建林, 等. 基于感受野注意力卷积的自动驾驶多任务感知算法 [J]. 计算机工程与应用, 2024, 60 (20): 133-141.
LIU Y X, MA H L, ZHU J L, et al. Autonomous driving multi-task perception algorithm based on receptive-field attention convolution [J]. Computer Engineering and Applications, 2024, 60(20): 133-141.

[18] WANG J, WU Q M, ZHANG N. You only look at once for real-time and generic multi-task[J]. ArXiv preprint arXiv:2310.01641, 2023.

[19] VU D, NGO B, PHAN H. Hybridnets: End-to-end perception network [J]. ArXiv preprint arXiv: 2203.09035, 2022.

[20] 武鹏宇,张远辉,刘康. 基于改进 HybridNets 的多任务驾驶感知方法 [J]. 激光杂志, 2024, 45(10): 80-85.
WU P Y, ZHANG Y H, LIU K, et al. Multi-task driving perception method based on improved HybridNets [J]. Laser Journal, 2024, 45 (10): 80-85.

[21] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.

[22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition,2018: 7132-7141.

[23] TAN M, LE Q. Efficientnetv2: Smaller models and faster training [C]. International conference on machine learning. PMLR, 2021: 10096-10106.

[24] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[25] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.

[26] 张润梅,贾振楠,李佳祥,等. 基于多感受野特征增强的改进 EfficientDet 遥感目标检测算法[J]. 电光与控制,2024,31(7):53-60.
ZHANG Y H, JIA ZH N, LI J X, et al. Improved EfficientDet remote sensing target detection algorithm based onmulti-sensory field feature enhancement [J]. Electronics Optics & Control, 2024,31(7):53-60.

[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.

[28] LIU W, LU H, FU H, et al. Learning to upsample by learning to sample[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 6027-6037.

[29] XU J, XIONG Z, BHATTACHARYYA S P. PIDNet: A real-time semantic segmentation network inspired by PID controllers[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 19529-19539.

[30] LIN T Y,GOYAL P, GIRSHICK R,et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.

[31] GIRSHICK R. Fast R-CNN [C]. Proceedings of the lEEE International Conference on Computer Vision (lCCV). Boston, USA: IEEE, 2015: 1440-1448.

[32] SALEHI S S M, ERDOGMUS D, GHOLIPOUR A. Tversky loss function for image segmentation using 3D fully convolutional deep networks[C]. International Workshop on Machine Learning in Medical Imaging. Cham: Springer International Publishing, 2017: 379-387.

[33] 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,2020: 2636-2645.

作者简介

宋建辉(通信作者),2004 年于哈尔滨工业大学获得学士学位,2006 年于哈尔滨工业大学获得硕士学位,2010 年于哈尔滨工业大学获得博士学位,现为沈阳理工大学教授,主要研究方向为多传感器信息融合、智能检测技术、目标识别与跟踪。

E-mail: hitsong@126.com

Song Jianhui (Corresponding author) received her B. Sc. degree from Harbin Institute of Technology in 2004, M. Sc. degree from Harbin Institute of Technology in 2006, and Ph. D. degree from Harbin Institute of Technology in 2010, respectively. Now she is a professor in Shenyang Ligong University. Her main research interests include multi-sensor information fusion, intelligent detection technology, target recognition and tracking.

刘鑫,2022 年于沈阳理工大学获得学士学位,当前于沈阳理工大学在读硕士研究生,主要研究方向为智能机器人技术及应用。

E-mail:2846151918@qq.com

Liu Xin received his B. Sc. degree from Shenyang Ligong University in 2022. Now he is a master's student in Shenyang Ligong University. His main research interests is intelligent robot technology and application.